

Understanding Flash Crash Contagion and Systemic Risk: A Calibrated Agent-Based Approach

James Paulin

Hertford College
University of Oxford



Department of Computer Science

Supervisors: Anisoara Calinescu and Michael Wooldridge

Examiners: Michael Wellman and Edith Elkind

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2019

To my daughters,
Martha, Ivy, and Nellie

Abstract

The global financial system is a sociotechnological complex network, in which millions of economic agents interact over timescales ranging from months to milliseconds. The decade since the near-collapse of this system has been characterised by the meteoric rise of computerised algorithmic trading. This transition has resulted in markets that are vulnerable to new forms of systemic risk, as exemplified by the Flash Crash of May 2010, and related events. The failure of extant models to predict, or mitigate, either the Global Financial Crisis, or the Flash Crash, has led financial regulators to increasingly engage with academia in search of new approaches through which to understand such systems. Agent-based modelling, rooted in complex systems theory, has emerged as a compelling method. However, there remains a pressing regulatory need to understand how systemic (macro-level) network effects are influenced by (micro-level) algorithmic trading.

To address this, we developed and assessed a novel combination of macro- and micro-level agent-based models of financial markets, demonstrating the benefits and applicability of the agent-based methodology. We constructed quantitatively validated network representations of all liquid asset portfolios held by investors in US equities, and then simulated the exchange of assets using quantitatively calibrated and validated market mechanisms, incorporating key regulatory interventions such as price limits and circuit breakers. Our microstructure calibration dataset spans a range of assets and dates that far exceeds previous approaches. By subjecting the validated model to counterfactual scenarios, such as price shocks, alternative network topologies, and alternative agent behaviours, we presented the first in-depth analysis of systemic risk over intraday timescales. We assessed the effectiveness of current, and proposed, regulations at mitigating this new form of high-frequency financial distress.

We find that the actions of distressed market participants facilitate new propagation channels for intraday systemic risk, and flash crash contagion, across the financial network. We also find that systemic risk has a complex non-linear dependence on network topology. We further identify stable and unstable network and agent-behaviour scenarios. The scope and complexity of the model allows the first quantitative analysis of the systemic effect of price limits and circuit breakers. Our insights include the fact that current regulation is inadequate in certain circumstances, and that its effectiveness strongly depends on fund liquidation behaviour. We perform the first ever quantification of the effectiveness of a novel policy proposed by a major market participant, and find it could potentially reduce systemic risk.

Based on our work and results, we argue that a hybrid micro-macro approach offers a powerful new lens, through which regulators can assess novel threats to financial stability in the modern, computerised, marketplace.

Acknowledgements

This thesis would simply not exist without the exceptionally generous financial support of OxFORD Asset Management. In particular, my thanks go to OxFORD founder partner Dr. André Stern, who has encouraged and challenged me in equal measure for the almost fifteen years since he first introduced me to the financial markets and the world of algorithmic trading. We’ve shared many exciting milliseconds, and I hope we will share many more. I am also deeply grateful to OxFORD partner Dr. Tom Newcomb, who’s support, pragmatism, and good humour, regardless of the circumstances, made the journey so much more enjoyable. I am indebted to many esteemed OxFORD colleagues, past and present, for their enthusiasm and feedback throughout my DPhil, especially Drs. David Lamper, Steve Kurlander, David Turner, Chris Townsend, and Andy Spencer.

I needn’t have been daunted about returning to academia several years into my professional career, because in Dr. Ani Calinescu and Prof. Mike Wooldridge I had a supervisory dream team. Both Ani and Mike showed me exceptional generosity, both intellectually, and practically. Our Friday meetings were a weekly fixture that I will sorely miss, and the wisdom Ani and Mike shared extended far beyond the bounds of this thesis — I will carry their lessons with me always. Whether it was Ani’s red-penned encouragement to “help the reader”, or Mike’s good-natured resistance to my ongoing campaign to bury him in graphs, our interactions were always a true joy, and a source of strength and inspiration. I have benefited immeasurably from their knowledge, experience, and compassion, and I have no doubt that we will find ways to continue to work together in the years to come.

I am also grateful to many others from within the academic community. Firstly, I would like to thank Profs. Edith Elkind and Paul Goldberg for their generous feedback at my transfer and confirmation examinations. My thanks also go to Prof. Sarit Kraus, and Anat Hashavit, for a rewarding collaboration. For their thought-provoking conversation and feedback I am grateful to Profs. Doyne Farmer, and Sam Staton. Profs. Rahul Savani and Renaud Lambiotte, and Drs. Martin Gould and Emmet Spier kindly offered practical support. This thesis greatly benefited from many interesting conversations with Rafa Baptista (*in memoriam*), Maarten Scholl, Alissa Kleinnijenhuis, Thom Wetzer, and participants at WEHIA 2019. I’m also hugely grateful for the support and enthusiasm of Drs. Paul Harrenstein and Julian Gutierrez, and Drs. to-be Peter Marinov, Jiarui Gan, and Dongge Han.

Finally, I owe an enormous debt of gratitude to my wife Aimée, our daughters, and both of our families. Thank you, all, for sharing the highs and lows with me, for asking so many interesting questions (and having the graciousness to indulge my sometimes-lengthy replies), and, most of all, for your unwavering love and support.

Contents

List of Figures	iii
List of Tables	v
List of Equations	vi
Glossary of Financial Terms	vii
Abbreviations	xi
Mathematical Notation	xiii
1 Introduction	1
1.1 The Micro-Macro Feedback Loop	4
1.2 Research Questions	6
1.3 Methodology	7
1.4 Key Contributions	8
1.5 Structure of the Thesis	10
1.6 Publications	10
2 Agent-Based Modelling for Complex Financial Systems	11
2.1 The Complex Nature of Financial Systems	12
2.2 Agent-Based Modelling	15
2.3 Systemic Risk	18
2.4 Macro-level Phenomena	25
2.5 Micro-level Phenomena	31
2.6 Calibration and Validation	39
2.7 Regulatory Policy and Interventions	48
2.8 The Need for a New Micro-Macro Modelling Approach	52
2.9 Summary of Previous Research	53
3 A New Approach to Simulating Systemic Risk and Flash Crash Contagion	55
3.1 A Note on Epistemology in this Thesis	56
3.2 Micro-Macro Approach	57
3.3 Agent-Based Model	61
3.4 Systemic Risk Metrics	70

3.5	Simulation Protocol Summary	73
3.6	Mathematical Notation	75
3.7	Empirical Validation	76
3.8	Trading Behaviour and Asset Price Shocks	79
3.9	Leverage, Capital, and Margin Tolerance	82
3.10	Bipartite Network Topology	88
3.11	Conclusions	98
4	Calibration and Validation	101
4.1	ABM Calibration and Validation	102
4.2	Big Data Approach	112
4.3	Model Output Moments and Stylised Facts	125
4.4	Initial Transient Analysis	129
4.5	Global Sensitivity Analysis	131
4.6	Calibration Methodology	136
4.7	Longitudinal Calibration and Validation Results and Discussion	152
4.8	Out-of-Sample Validation Results	157
4.9	Analysis of Calibrated Parameters	160
4.10	Flash Crashes in the Calibrated Model	165
4.11	Conclusions	170
5	Real-World-Scale Policy Evaluation	175
5.1	Real-World-Scale Agent-Based Methodology	177
5.2	Real-World-Scale Intraday Systemic Risk	186
5.3	Regulatory Policy: Price Limits and Circuit Breakers	195
5.4	Simulated Regulatory Interventions	202
5.5	<i>In Silico</i> Policy Evaluation	219
5.6	Conclusions	231
6	Conclusions	235
6.1	Summary	235
6.2	Key Contributions	237
6.3	Future Work	240
A	Trader Agent Details	243
B	Bipartite Network Generation	247
C	Global Sensitivity Analysis Supplement	249
D	Out of Sample Validation Supplement	255
E	Policy and Interventions Supplement	257
	Bibliography	264

List of Figures

2.1	Margin spiral	27
2.2	Schematic representation of the limit order book	32
2.3	Micro-macro modelling summary	53
3.1	Relationships between entities in the model	58
3.2	Generative bipartite network topology	66
3.3	Single asset model validation	78
3.4	Single asset flash crash occurrence and onset	81
3.5	Probability of contagion vs leverage, capital, and hysteresis	84
3.6	Systemic risk univariate sensitivity	86
3.7	Probability and extent of contagion	90
3.8	Flash crash propagation speed	93
3.9	Probability of contagion for non-uniform allocations	97
4.1	Summary of datasets relevant to our model	113
4.2	Comparative scope of microstructure ABM empirical datasets.	118
4.3	Block diagram of our calibration and validation methodology.	125
4.4	Initial transients in the simulation	130
4.5	Sensitivity of moments to ABM parameters	134
4.6	Comparison of calibration methods	149
4.7	Longitudinal calibration results	154
4.8	Out-of-sample validation	158
4.9	Calibrated parameter distributions	161
4.10	Probability of flash crash occurrence in the calibrated ABM	167
5.1	13-F fund–asset network degree correlation	178
5.2	Full-scale model instantiation diagram	184
5.3	Probability of contagion dependence on fund leverage and shocked as- set centrality	188
5.4	Probability of contagion dependence on shocked asset centrality	191
5.5	Limit-Up Limit-Down State Machine	196
5.6	Price Limit and Circuit Breaker State Machine	200
5.7	Probability and extent of contagion dependence on fund behaviour and ω_{mcap}	206
5.8	Fund distress cascade quenching and market-wide circuit-breaker acti- vation dependence on fund behaviour and ω_{mcap}	212

5.9	Pre market-wide circuit breaker contagion extent dependence on fund behaviour and ω_{mcap}	216
5.10	BlackRock trigger fraction	221
5.11	Systemic risk under LULD-MWCB-BKRRK	223
5.12	Pre-circuit-breaker extent of contagion under LULD-MWCB-BKRRK	226
C.1	Sensitivity analysis convergence	250
C.2	Sensitivity analysis correlation	251
C.3	Sensitivity analysis Monte Carlo replicates (constant budget)	252
C.4	Sensitivity analysis Monte Carlo replicates (increasing budget)	253
D.1	Probability of flash crash occurrence in the calibrated ABM	256
E.1	Extent of contagion dependence on fund leverage and shocked asset centrality	258
E.2	Probability of distress dependence on fund leverage and shocked asset centrality	259
E.3	Flash crash extent dependence on fund leverage and shocked asset centrality	260
E.4	Probability of fund distress dependence on fund behaviour and ω_{mcap}	261
E.5	Pre market-wide circuit breaker contagion extent dependence on fund behaviour and ω_{mcap}	262
E.6	Shocked asset degree distribution	263

List of Tables

2.1	Price impact function survey	29
2.2	Stylised facts survey	42
3.1	Bank balance sheet model	61
3.2	Fund balance sheet model	63
3.3	Trader agent parameters	67
3.4	Parameters in the theoretical model	75
3.5	Other notation used in this chapter.	76
3.6	Sampled locations in (λ^0, τ_c, C^0) space	89
3.7	Fund–asset network connected components	91
4.1	ABM dataset scale comparison	119
4.2	Market microstructure dataset properties	121
4.3	Model parameters requiring calibration	122
4.4	Summary of fixed model parameters	124
4.5	Summary of validation and calibration approach for the model	124
4.6	Hyperparameters of the random search algorithm	139
4.7	Hyperparameters of the Nelder-Mead simplex algorithm	141
4.8	Hyperparameters of the genetic algorithm	142
4.9	Hyperparameters of the Markov chain Monte Carlo algorithm	145
5.1	Comparison of network features between models	181

List of Equations

Equation 3.1	56
Equation 3.2	62
Equation 3.3	62
Equation 3.4	62
Equation 3.5	62
Equation 3.6	70
Equation 3.7	70
Equation 3.8	71
Equation 3.9	71
Equation 3.10	71
Equation 3.11	72
Equation 3.12	73
Equation 3.13	73
Equation 4.1	103
Equation 4.2	104
Equation 4.3	104
Equation 4.8	106
Equation 4.9	106
Equation 4.10	107
Equation 4.11	107
Equation 4.12	108
Equation 4.13	108
Equation 4.14	108
Equation 4.15	109
Equation 4.16	109
Equation 4.17	109
Equation 4.18	110
Equation 4.19	114
Equation 4.20	114
Equation 4.21	133
Equation 4.22	133
Equation 4.23	138
Equation 4.24	143
Equation 4.25	148
Equation 5.1	187
Equation 5.2	187
Equation 5.3	201
Equation 5.4	201
Equation 5.5	201
Equation 5.6	213
Equation 5.7	214
Equation 5.8	214

Glossary of Financial Terms

- aggressive order** A spread-crossing order that is priced to match against an order resting on the limit order book, *see* [spread](#)
- algorithmic trading** Delegation of fine-grained trading decisions (e.g., limit order entry) to an automated system
- arbitrage** Exploitation of structural mis-pricing opportunities between assets, markets or other entities
- ask** A price at which an investor is willing to sell an asset, *see* [limit order book](#)
- asset** A transferable financial entity, possessing (at least) a value
- asset class** A group of related assets, such as equities, futures, or bonds
- auction** A price-based order matching mechanism used by trading venues
- bail-out** Offering extra capital to a failing institution such that it can avoid defaulting on its loan obligations
- basis point** One-hundredth of one percent. Also written bps, bpts.
- best ask** The current lowest price at which any investor is willing to sell an asset
- best bid** The current highest price at which any investor is willing to buy an asset
- bid ask spread** The price difference between the best bid and the best ask
- bid** A price at which an investor is willing to buy an asset, *see* [limit order book](#)
- broker** A financial intermediary providing market access and associated services including execution algorithms
- capital** Money used for investing
- central bank** Government institution that may bail-out failing institutions by acting as lender-of-last-resort
- circuit breaker** A mechanism intended to protect markets against excess volatility by stopping trading, *see* [volatility](#)
- close** The end of the trading day, typically accompanied by an auction to determine the close price, *see* [uncross](#)
- continuous trading** A price-time priority double auction, the usual operational mode of exchanges
- contract** Unit of a derivative asset, *see* [derivative](#)
- cover** Closing a short position by buying back sold shares, *see* [short](#)
- crash** A sudden and significant fall in the value of one or more assets
- dark pool** A trading venue that does not publish a data feed of transactions or available liquidity
- default** When an investor cannot service a debt contract due to insufficient funds
- deleverage** Reduce leverage by reducing absolute position sizes, *see* [unwind](#)
- derivative** A contract, the value of which is derived from an underlying asset

diversification Spreading risk by holding investments in multiple securities

equities Shares of public companies, *see* [asset class](#)

execution algorithm A trading algorithm designed to balance immediacy with price impact rather than generate profit on its own, *see* [price impact](#)

execute Transact in the marketplace, *see* [fill](#)

exchange A financial institution that facilitates trading, *see* [trading venue](#)

fill A report from a trading venue confirming that an order has transacted in the market and a trade occurred

fire sale Distressed selling of assets at discounted prices to cover default liabilities

fixed income An asset class that pays a specific rate of interest (e.g., bonds)

flash crash Events such as those of 6th May 2010, where US equity markets crashed and partially recovered in a matter of minutes

future Asset class of derivative contracts with fulfilment requirements e.g., physical delivery of goods, *see* [derivative](#)

fund An asset management fund, such as a mutual fund, or a hedge fund, or the investment division of a bank

hedge Using one investment to offset risk associated with a second

illiquid An asset with low liquidity such that transactions in the asset become difficult

index The weighted average value of a set of assets, e.g., the S&P 500

instrument Tradable asset, *see* [asset](#)

intraday Pertaining to events occurring during the period while markets are open for business, *see* [trading day](#)

latency arbitrage A high-frequency trading strategy that exploits communications latency to exploit asset mis-pricings between trading venues

leverage Borrowing capital for the purchase of assets, e.g., entering into a mortgage contract with a lender

limit order book The current set of passive orders representing assets availability for sale (offers) and purchase (bids) at specific prices

limit order An instruction to buy or sell a certain number of units of an asset at a specific price

liquidate Completely unwind a position, *see* [unwind](#)

liquidity The value of some quantity of an asset expressed as the product of price and quantity, expressing the ease with which the asset can be traded

lit venue A trading venue that publishes data feeds of transactions and/or available liquidity, *see* [dark pool](#)

long Holding a positive position in a security, *see* [short](#)

margin call The action of a lender requiring further capital from an investor to cover trading losses

margin Capital held by a lender as collateral for leveraged investment, *see* [leverage](#)

market cap Market capitalisation. The total value of a company's equity (the product of price and shares outstanding)

market maker A market participant that maintains passive limit orders on both bid and ask sides of the limit order book attempting to profit from the spread

market A for-profit entity that facilitates the exchange of assets between participants

market order An instruction to immediately buy or sell a certain number of units

of an asset at the prevailing best market price

meta-order A large order that is broken into smaller orders, *see* [execution algorithm](#)

offer The current lowest price at which any investor is willing to sell an asset, *see* [ask](#)

open Start or resume trading an asset, *see* [uncross](#)

option A financial derivative contract which provides a right to buy or sell an asset at pre-agreed price at a future time

order An instruction to buy or sell a specific quantity of an asset, *see* [asset](#)

order book The current set of orders in the market, *see* [limit order book](#)

passive order A buy (sell) order that is priced below (above) the prevailing best bid (ask) and will rest on the limit order book, thus providing liquidity

portfolio The set of positions owned by an investor

position An investment in an asset, can be long or short

price impact The market response to trading activity such that buying (selling) activity moves prices up (down) as a function of demand

return Asset time-wise price difference, often expressed as a percentage

regulator Government authority with responsibility for orderly market function

security Tradable asset, *see* [asset](#)

share Tradable unit of an asset, typically used to refer to an equity

shares outstanding The total number of equity shares issued by a company

short Equivalent to holding a negative position in an asset, *see* [long](#)

short selling Borrowing a security and selling it, hoping the price will fall so that the security can be repurchased at a profit and returned, *see* [short](#)

spread Price difference between the current best bid and ask, *see* [limit order book](#)

stub quote Limit orders placed very far from prevailing prices with commensurately small probability of execution during normal market conditions

tick size Minimum price increment between bids and offers on the limit order book

trading day The hours during which an exchange operates. E.g., 9:30–16:00 for US equities, *see* [open](#) & [close](#)

trading strategy A systematic investment or set of investments intended to generate profit by exploiting perceived asset mis-pricing

trading venue A facility for matching buyers with sellers of assets

uncross Resolve an auction by optimally matching orders on the limit order book

unwind Reduce the size of an asset position towards zero by selling or buying to cover as appropriate, *see* [cover](#)

volatility Variation in price movements typically defined as the standard deviation of returns over some horizon, *see* [return](#)

volume Quantity of an asset, expressed as a number of shares or contracts

Abbreviations

13-F	Form 13-F US regulatory portfolio ownership disclosure
ABM	Agent-based model
CDA	Continuous double auction
DGP	Data generating process
DoE	Design of experiments
DSGE	Dynamic stochastic general equilibrium
EUR	Euro currency asset class
FSM	Finite state machine
FTSE	Financial Times stock exchange, the index of the largest 100 UK equities
FX	Foreign exchange asset class
GFC	Global Financial Crisis (of 2007–2009)
GMV	Gross Market Value, the total value of a portfolio
HFT	High-frequency trading
HPC	High-performance computing
LOB	Limit order book
LULD	Limit-Up-Limit-Down rule, a regulatory circuit breaker policy
mcap	Market capitalisation
mDGP	Model data generating process
MWCB	Market-wide circuit breaker
NYSE	New York stock exchange
rwDGP	Real-world data generating process
SEC	Securities and Exchange commission, the US regulator
SEM	Standard error in the mean
SIFI	Systemically important financial institution
S&P 500	Standard and Poor's index of the largest 500 US equities
USD	United States Dollars currency asset class

Mathematical Notation

This table lists the key mathematical notation used in this thesis. For brevity, it does not include symbols that are only used in intermediate steps in a derivation. Where a symbol has more than one definition, the meaning is given by the usage context.

$[0, n)$	A real number range e.g., $\{x \in \mathbb{R} \mid 0 \leq x < n\}$
$(0, \dots, n)$	An integer sequence e.g., $\{x \in \mathbb{Z} \mid 0 \leq x \leq n\}$
t	Simulation time-step
δt	Simulation time-step interval
T	Simulation total duration
T^*	Simulation shock time
$\mathcal{S}$	Set of experimental scenarios
M	Number of Monte Carlo replicates
n_f	Number of funds
n_a	Number of assets
$\mathbf{A}$	Fund-asset initial positions (\$USD)
$\mathbf{S}^t$	Fund-asset positions (shares)
π	Fund GMV (\$USD)
$\mathbf{v}, \mathbf{v}', \mathbf{v}''$	Asset market caps (\$USD) (estimate, true, fixed)
p^t	Asset price
$\mathcal{A}$	Set of assets (meaning given by context)
C^t	Fund capital (\$USD)
λ^t	Fund leverage
τ_c	Fund leverage tolerance
ρ	Fund diversification
β	Fund crowding parameter
α	Fund allocation uniformity
k_{fund}	Fund degree
η	Distressed seller volume fraction
δ	Distressed seller timescale (seconds)
γ	Default cascade indicator threshold
$\mathcal{D}$	Set of trials in which default cascades observed
$P_{contagion}$	Probability of contagion
$\Omega_{contagion}$	Extent of contagion
$\mathbb{I}_{cascade}$	Default cascade indicator
$\mathbb{I}_{default}$	Single fund default indicator
$f_{default}$	Fraction of funds in default

$\mathbb{I}_{FC}$	Flash crash indicator
s	Flash crash speed
ζ	Flash crash speed measurement percentile
P_{quench}	Probability of quenched cascade
$\mathbf{g}(\cdot)$	ABM input-output transformation function
$\mathbf{Z}_0$	ABM initial conditions
$\mathbf{X}_t$	ABM intermediate state
$\mathbf{y}_t$	ABM intermediate output
$\mathbf{Y}$	ABM composite output
$\mathbf{Y}_R$	Empirical (real-world) output
$\boldsymbol{\mu}(\cdot)$	Output moment function
$\boldsymbol{\mu}$	Model output moments
$\boldsymbol{\mu}_S$	Synthetic output moments
$\boldsymbol{\mu}_R$	Empirical output moments
$\mathcal{P}_\mu$	Empirical moment distribution
(μ_-, μ_+)	Empirical moment confidence interval
$\Theta \subset \mathbb{R}^D$	Parameter space of dimension D
$\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}$	Parameter vector (raw, calibrated, pseudo-real)
$\mathcal{L}(\cdot)$	Loss function
$\hat{\mathcal{L}}(\cdot)$	Estimated loss function
$\tilde{\mathcal{L}}(\cdot)$	Pseudo-real loss function
$\mathbf{W}$	Loss function weight matrix
ψ	Hybrid loss function hyperparameter
$\mathcal{P}_\mathcal{L}$	Empirical loss distribution
$\mathcal{L}^{\text{MCR}}$	Moment coverage ratio (MCR)
$\mathcal{L}^{\text{jMCR}}$	Joint moment coverage ratio (jMCR)
τ'^μ	jMCR validation threshold
b	Market cap band index
N	Number of parameter vector samples
I	Number of parameter vector instantiations
S_{1k}, S_{Tk}	Sensitivity indices (first order, total order)
$\mathcal{Q}$	Quarter (as in portion of calendar year)
ω_{mcap}	Shocked asset market cap percentile
ω_{degree}	Shocked asset degree percentile
$\mathcal{I}$	Set of assets in S&P 500 index
$\mathbf{w}$	S&P 500 index weights
$I_{\text{SP500}}^t, R_{\text{SP500}}^t$	S&P 500 index value and return
$\mathcal{F}$	Set of fund behavioural policies
$\mathcal{U}$	Unconditional sale fund behaviour
$\mathcal{R}$	Randomised sale fund behaviour
$\mathcal{H}$	High-mcap sale preference fund behaviour
$\mathcal{L}$	Low-mcap sale preference fund behaviour
ϕ	Conditional sale fund behaviour fraction
b	BlackRock policy parameter (%)

*Flash crashes are to markets as rogue waves
are to sailors. @jkpaulin*

— Grady Booch, twitter, January 2019

1

Introduction

DURING the *Flash Crash* of 6th May 2010, over a trillion dollars were wiped off the value of US companies in an event that has been largely attributed to the rise of *algorithmic trading* (CFTC and SEC, 2010). Between 14:30 and 14:45, the Dow Jones Industrial Average recorded the largest single day point loss in its entire history, only for prices to revert to near-pre-crash levels just fifteen minutes later. Regulators were left scrambling to explain how the markets had been left vulnerable to this new, algorithmically-driven, form of high-speed systemic failure (CFTC and SEC, 2010). These unprecedented events triggered far-reaching consequences, and the causes and implications are still being actively investigated (Menkveld and Yueshen, 2019). However, although the circumstances giving rise to the Flash Crash have been intensely studied since 2010, the role played by algorithmic traders in propagating *systemic risk* (cascading failure) through financial networks remains poorly understood.

In the context of networks representing the overlap of asset portfolios held by investment funds (hereafter *funds*), systemic risk arises as a result of *price impact* (Brunnermeier and Pedersen, 2009), i.e., the market’s response to trading activity such that excess buying causes prices to rise, and excess selling causes prices to fall. Despite this, over-simplified systemic risk models typically assume non-crisis linear price impact functions (Benoit et al., 2017) that ignore the microstructural features of order placement and trader behaviour. Systemic events, such as the Flash Crash of 6th May 2010, demonstrate that during crises (conditions of abnormal perturbation

when system dynamics are far from equilibrium), price impact can deviate strongly from linearity (CFTC and SEC, 2010). Analyses of flash crashes have found that a market’s ability to function as an orderly price discovery mechanism is highly dependent on the micro-level behaviour of market participants, in particular those involved in algorithmic trading (CFTC and SEC, 2010; Menkveld, 2016).

To date, little progress has been made in understanding the mechanisms by which the micro- and macro- levels of financial markets interact with, and modulate, each other, during unfolding crises. But the rapid uptake over the past decade of algorithmic trading (and its behavioural subset, *high frequency trading* (HFT), Menkveld, 2016) means that market dynamics today are radically different from those that were prevalent during the Global Financial Crisis (GFC) of 2007–2009. Market liquidity provision played a key role during the GFC, as deeply interconnected financial institutions competed for a dwindling liquidity supply, in interbank, and other, global markets. The present dominance of HFTs means that liquidity provision has become increasingly capricious (Van Kervel and Menkveld, 2019), and prone to evaporating precisely when it is needed most — during a crisis. Regulators and policy makers have thus been left struggling to keep pace with the realities of the modern computerised marketplace (Bookstaber, 2012; Haldane and Turrell, 2019). There is growing concern that the next global financial crisis will be triggered by a trading algorithm causing a price dislocation such as a flash crash, exacerbated by undependable liquidity provision, forcing market participants to sell into falling prices, and thus escalating into a full-blown systemic crisis (Wigglesworth and McLannahan, 2018).

In response to the failure of leading models available to regulators during the GFC, the discipline of *agent-based modelling* (ABM) has flourished (LeBaron and Tesfatsion, 2008; Farmer and Foley, 2009; Gatti et al., 2011; Fagiolo and Roventini, 2017)¹. Through their failure to predict or mitigate the GFC, the incumbent *dynamic stochastic general equilibrium* (DSGE) models were shown to be too simplistic and dogmatic, failing to account for the amplifying effects of the complex network of interconnections between financial institutions. ABM, by contrast, offers a natural approach to representing the heterogeneous properties of strongly interacting market participants (Fagiolo and Roventini, 2017; Bouchaud, 2019). By relaxing assumptions of perfectly rational behaviour, agent-based models preclude themselves from the kind of closed-form analysis that makes DSGE models so attractive to economists. Due to their ability to capture individual heterogeneous behaviour, ABMs are able to more

¹We use “ABM” to refer to agent-based modelling as a discipline, and also to refer to a given agent-based model, with the interpretation given by the context.

faithfully represent the bounded rationality and imperfect reasoning of real-world market participants — what ABMs lack in analytical tractability, they make up for in model expressiveness and flexibility.

The complex interactions of heterogeneous agents give rise to *emergent* systemic properties. This feature, which again reflects that of real-world systems (Hayek, 1945), leads to substantial difficulty in comparing ABMs to the real-world (Gilli and Winker, 2003; Gatti et al., 2018). The issues of model *calibration* and *validation* are central to the ABM methodology, and the discipline is in the process of advancing from a position of *qualitative* consistency with real-world observables, to one of *quantitative* consistency (Axtell and Epstein, 1994; Fagiolo et al., 2019). The momentum behind this significant progress is being driven by enhanced access to fine-grained datasets (for example, financial “Big Data”), growth in computational capability, and the use of interdisciplinary methods at the interface of computer science, statistics, and machine learning (e.g., Lamperti et al., 2018). However, market complexity is also being fuelled by the same drivers (Lo, 2017), and so it is necessary for the ABM community to also advance in terms of a sound methodology available to support model calibration and validation. A major goal of both academics and practitioners is to develop an understanding of the techniques that are appropriate for large-scale ABM calibration and validation, and to arrive at a unified set of terminology and methodological processes, such that it is possible for ABM researchers in different fields to benefit from each other’s innovations and results (Fagiolo et al., 2019).

A methodology for quantitatively calibrating and validating ABMs will enable their wide-scale deployment as digital laboratories for counterfactual, *in silico* regulatory policy evaluation. By definition, financial crises are tail events with few realised examples available for study. Previous empirical work had to use natural experiments to infer the effect of regulatory policy changes (for example, investigating market properties before, and after, structural changes are introduced, Jain et al., 2016). In contrast, ABM offers a compelling alternative approach where the experimenter is free to construct unlimited crisis scenarios at will, and to use observations of resulting system dynamics to build a statistical understanding of the effect of policy amendments.

In this thesis, we propose, implement, and assess a realistic microstructure model of price impact during crisis periods, embedded within a multiple fund, multiple asset environment characteristic of models from the macroscopic systemic risk literature. The model allows us to explore the propagation of financial distress associated with inherently *intraday* (during trading hours) market phenomena such as flash crashes.

Events taking place over timescales measured in minutes, seconds, or shorter are simply ignored in the overwhelming majority of macroscopic financial models. However, the impact of such events can, as illustrated by the 6th May 2010 Flash Crash, be measured in trillions of dollars, and serve to significantly undermine the orderly functioning of global financial markets.

With this fundamental *micro-macro* model in hand, we proceed to calibrate and validate the model using an extensive set of high-frequency market data. The methods we developed and assessed allowed us to instantiate the first real-world-scale micro-macro ABM of the financial system, which we deployed to quantify systemic risk and flash crash contagion through the real-world financial system. Using this model, we assessed relevant existing and candidate regulatory policies, aimed at mitigating intraday systemic risk.

The results and analyses that we present in this thesis are of direct relevance to regulators and policy makers as they seek to understand and quantify threats to financial stability, including the impact of network topology on the spread of flash crashes over intraday timescales, and the effectiveness of regulatory policy at damping contagious distress under alternative fund behavioural assumptions. Our results are also of wider relevance to the ABM community, as we demonstrate how to combine microscopic with macroscopic modelling assumptions, agent types, agent behaviours, and data, all within the overarching paradigm of complex networked systems. As we will show in the chapters that follow, the field of computer science is extremely well-placed to contribute to the financial ABM agenda. Although this thesis is interdisciplinary, it is primarily written with a computer science audience in mind. To that end, we have included a glossary that defines the specialist financial terms relevant to this thesis. Conversely, we have endeavoured to handle the computer-science-related content of this thesis in such a way that our message remains relevant and accessible to regulators, researchers, and practitioners in the financial markets.

1.1 The Micro-Macro Feedback Loop

Existing macroscopic systemic risk models incorporate over-simplified assumptions in their modelling of price impact. The model we develop in this thesis addresses this weakness. In reality, asset prices are formed due to the action of individual traders placing limit orders on stock exchanges, and this is true whether the trader is a global bank seeking to reduce position sizes over several days, or a high-frequency trader targeting fleeting profit opportunities lasting only fractions of a second. However,

assets do not exist in isolation. Rather, they are held within the portfolios of funds. Funds are thus connected to each other by virtue of the assets they hold in common. Similarly, assets are connected to each other by virtue of being held in the same portfolio by a fund (Allen and Babus, 2008). Following established methodology from the macroscopic systemic risk literature (Huang et al., 2013; Caccioli et al., 2014; Cont and Schaanning, 2017), we represent the *fund-asset network* as a bipartite graph where edges only exist between fund nodes and asset nodes, and represent asset investments by funds. We extend previous works by controlling network topology with a *preferential attachment* method (Barabási and Albert, 1999) which allows us to tune the degree of portfolio overlap, and hence the connectivity, between funds. We proceed to build on this foundation by directly representing the empirical fund-asset network using publicly available regulatory filing data (Guo et al., 2016).

When funds borrow from a bank or brokerage to finance an asset portfolio, they are required to maintain leverage (the ratio of portfolio value to collateral) below a pre-agreed limit as the value of the assets in their portfolio varies (Brunnermeier and Pedersen, 2009). If the portfolio value suffers a loss due to price fluctuation, the fund may be required to either deposit additional capital, or rapidly liquidate assets (*deleverage*) in order to satisfy the leverage constraint. This scenario is known as a *margin call* and can be enacted by the lending bank on a daily or intraday basis (Gai et al., 2011; Brunnermeier and Pedersen, 2009). Loss to the lender bank accrues when the liquidation value of a client fund’s portfolio is less than the value of the outstanding loans for which the portfolio acts as collateral. In this circumstance, it is impossible for the fund to repay its debt, and the fund *defaults*. As Brunnermeier and Pedersen (2009) describe, a problematic amplification mechanism known as a *margin spiral* can occur if a fund is forced to sell its assets due to losses: the action of selling further depresses prices, increasing losses and requiring further rounds of *distressed selling*. Systemic risk arises when different funds have investments in the same assets. Here, distressed selling by a fund causing prices to fall via price impact will cause losses at all other funds with investments in the distressed assets. These other funds may then be required to reduce the size of their investments in order to manage their leverage, and so the distress propagates in what is known as a *fire sale* (Shleifer and Vishny, 2011).

Previous studies have investigated the extent to which contagion between assets during the 2010 Flash Crash can be attributed to cross-market arbitrage activity (Menkveld and Yueshen, 2019). Leverage management, by contrast, represents

a potential contagion channel for intraday distress that is capable of directly connecting and coordinating large institutional investors and portfolio managers in a high-frequency systemic deleveraging spiral. The continued dominance of algorithmic trading in modern markets suggests that understanding the nature of crisis liquidity provision is of central importance in the development of regulations to mitigate potential future financial crises.

In the present work, we simulate crisis price impact due to algorithmic trading with a *limit order book* (LOB) agent-based model, calibrated to exhibit extreme transient price impact phenomena such as flash crashes. The micro-model is derived from Paddrik et al. (2012). We combine this micro-level ABM with a macro-level network model of overlapping leveraged asset portfolios held by funds, derived from works including Caccioli et al. (2014) and Huang et al. (2013). The result is a novel synthesis, combining existing models from both the market microstructure, and the macro-level systemic risk, literatures. We thus model the feedback between market institutions and individual traders that Kobayashi and Hashimoto (2011, p.370) refer to as the “micro-macro loop”. We are aware of few related studies that attempt such a synthesis (Gerig, 2015; Torii et al., 2015; Xiong et al., 2015). However, if regulators are to truly understand the threats to the financial system posed by the global web of interconnected financial institutions, deploying myriad interacting trading algorithms, operating at millisecond latencies, with impacts felt over strictly-intraday timescales, we argue that such a synthesis is not just appropriate, but essential.

1.2 Research Questions

The central question that underpins the research in this thesis is:

What role do algorithmic trading systems have in the contagion of distress through a network of financial institutions, connected via shared-asset investments?

We approach this question through a set of three closely-coupled phases, each incorporating its own set of detailed research questions. In the first phase, our research questions concern the effects on systemic risk (1) of algorithmic trader behaviour, (2) of fund-asset network topology including portfolio *crowding*, which is a measure of the similarity of the asset portfolios held by different funds, and fund *diversification* which relates to the number of assets in which funds invest, and (3) of the use of

leverage by investors. We also question how these factors affect the speed at which flash crashes propagate between assets in the financial system.

In the second phase, we seek to understand how best to calibrate and validate hybrid micro-macro models of the financial system. Our research questions concern the selection of appropriate methods to make use of financial Big Data for large-scale microstructure calibration, over a range of assets and dates that has not previously been demonstrated in the literature. Our research questions regard the relative effectiveness of alternative calibration methods, and the generality of the resulting calibrated and validated model.

In the final phase, our research questions concern the effectiveness of real-world intraday systemic risk regulatory policies including *price limits* and *circuit breakers* on market function (Sifat and Mohamad, 2019). We wish to determine the relative effectiveness of competing novel policies and current regulatory practice, in terms of both the probability and extent of systemic collapse. We also wish to determine the extent to which policy effectiveness varies between competing, plausible, market participant behavioural scenarios.

1.3 Methodology

In order to address our research questions, we designed, developed, and evaluated a hybrid micro-macro ABM that allowed us to characterise the effect on systemic risk of leverage management policies and fund–asset network topology. In addition to considering systemic stability (characterised by the default of funds on their margin loans), the use of a limit order book ABM allows us to investigate the *speed* at which financial distress propagates between assets over short intraday timescales characteristic of flash crashes. We investigate systemic risk (1) due to portfolio crowding, and (2) due to portfolio diversification. We further investigate the effects on systemic risk of the amount of leverage and capital deployed by funds, and of the tolerance with which funds manage their leverage under fluctuating asset price conditions.

To extend the validity and applicability of the model, we subjected it to quantitative calibration and validation, using the *method of simulated moments* (Winker et al., 2007) to quantify the model’s ability to reproduce statistical regularities of financial data (so-called *stylised facts*, Cont, 2001). We make use of *tick-level* US equity market transaction data spanning 2009–2018, and encompassing the full range of liquid assets traded on the major US exchanges. The scale and fidelity of our data far exceed previously reported micro-level datasets used for ABM calibration, and

due to this scale require careful preparation and handling. We defined and assessed procedures to fully exploit this dataset. We furthermore made use of regulatory filing data to specify the macro-level ABM components. We defined a set of loss functions that allow us to quantify the model’s ability to produce output that is consistent with reality, and that allow the model to be calibrated using gradient-based methods, for which we performed a comparative evaluation. We additionally assessed the potential for *global sensitivity analysis* (Saltelli et al., 2010) to determine whether model parameters meaningfully contribute to model output.

We addressed our final set of research questions, regarding regulatory policy evaluation, by representing in the model *all* liquid investments in US equities for a particular recent epoch. We further augmented the model by implementing both current, and candidate, intraday regulatory policies based on price limits and circuit breakers. Such policies (including the current US market regulation) operate by reference to the level of a market *index*, by definition a function of a large number of assets. In order to understand the microstructural circumstances leading to circuit breaker triggering, and to assess the effect on systemic risk of such regulatory measures, a hybrid micro-macro model again offers a compelling modelling framework. By simulating targeted shocks to the network, we are able to establish the relative importance of asset size and connectivity, and to assess the extent to which observations from smaller scale, synthetic fund–asset networks hold in the full-scale model, where certain assumptions regarding fund and asset homogeneity are relaxed. We performed counterfactual scenario experiments comparing existing policy with different specifications of a novel policy, proposed by a major market participant (BlackRock, 2015).

1.4 Key Contributions

The key contributions in this thesis are:

1. We develop a novel hybrid micro-macro ABM, forming an early bridge between (macro-level) systemic risk models and (micro-level) limit order book microstructure models.
2. We use this model to assess the extent to which leverage management practices by funds constitute a sufficient contagion channel for flash crash propagation and *default cascades* (contagious fund failures).

3. We investigate the dependence of intraday systemic risk and flash crash contagion on the topology of the fund–asset network, and identify non-monotonic relationships, with stable and unstable network configurations.
4. We extend and enhance the ABM by subjecting it to rigorous quantitative calibration and validation, comparing the effectiveness of a range of calibration methods using an extensive real-world microstructure dataset spanning the full asset size spectrum, for the period 2009-Q1 – 2018-Q4.
5. We present the first quantitative analysis of the impact on systemic risk of intraday regulatory policies designed to mitigate excess market volatility, including price-limits (per-asset), and market-wide circuit breakers (with activation as a function of members of the S&P 500 index). We examine and assess the effect of agent behaviour on intraday systemic risk propagation, and find surprising differences from the results of related macro-only systemic risk models.

Our results contribute to the growing literature on the use of ABMs to investigate the complex, emergent dynamics of financial systems (LeBaron, 2000; Tesfatsion, 2002; Farmer and Foley, 2009; Haldane and May, 2011; Fagiolo and Roventini, 2017; Haldane and Turrell, 2018). ABMs are able to reproduce empirically observed features of financial systems that cannot easily be captured within mainstream DSGE approaches (Fagiolo and Roventini, 2017). The complex interactions of heterogeneous, boundedly-rational, autonomous agents give rise to emergent phenomena that cannot be predicted from an understanding of agent behaviour alone. However, this expressiveness comes at a cost. ABMs typically feature a large number of parameters, making model calibration difficult and costly in terms of computational resources. It is, furthermore, necessary to perform repeated stochastic *Monte Carlo* simulations of ABMs in order to comprehensively sample observed system dynamics (Franke and Westerhoff, 2012). However, the Global Financial Crisis of 2007–2009 demonstrated the inadequacy of mainstream DSGE models in analysing the systemic risk of the financial system (Fagiolo and Roventini, 2017). ABMs offer a compelling alternative, and advances in their methodology, and in computer power and data availability, will further enhance their effectiveness. Regulators are in great need of tools and methodology to prevent the next financial crisis from occurring. By freeing practitioners from the shackles of historical data series, counterfactual scenario analysis via agent-based modelling promises to offer a powerful and compelling framework for the development of novel policy.

1.5 Structure of the Thesis

In Chapter 2, we survey existing financial simulation, calibration, validation and regulatory policy analysis literature, and we establish the need for a new class of micro-macro financial market model. In Chapter 3, we design, develop, and evaluate such a hybrid micro-macro model, and significantly extend the state-of-the-art in multi-asset microstructure simulation scale. In Chapter 4, we subject the model to quantitative calibration and validation, and present results comparing the effectiveness of competing calibration approaches, as well as discussing the model’s generality. In Chapter 5, we instantiate full real-world-scale micro-macro ABMs that incorporate the current US regulatory standard price limit and circuit breaker policies. We measure the effectiveness of these policies under alternative fund behavioural assumptions, and compare the effectiveness of a novel policy proposal. Chapter 6 concludes.

1.6 Publications

- Material from Chapter 2 has been published as [Paulin et al. \(2018\)](#). “Agent-based modeling for complex financial systems.” *IEEE Intelligent Systems*, 33(2):74–82.
- Material from Chapter 3 has been published as [Paulin et al. \(2019\)](#). “Understanding flash crash contagion and systemic risk: A micro-macro agent-based approach.” *Journal of Economic Dynamics and Control*, 100:200–229.
- Material from Chapter 4 is being prepared for submission as a journal paper.
- Chapter 5 is planned for preparation as a journal paper.

“Although the economic cost has been dire, the 2008 crisis may have been a blessing for getting theoretical economics out of the rut it had dug for itself.”

Bouchaud (2019, p.27)

2

Agent-Based Modelling for Complex Financial Systems

IN THIS CHAPTER, we review the application of computer-based techniques to the modelling and simulation of financial systems. We particularly focus on the notion of financial stability, or *systemic risk*, as opposed to financial prediction. We review the systemic risk literature from both macroscopic and microscopic perspectives, including systemic risk metrics.

We begin by briefly surveying the nature of our current global financial systems in Section 2.1, which allows us to consider the suitability of two competing classes of models. We contrast the efficacy and explanatory power of *dynamic stochastic general equilibrium* (DSGE) models with alternative models rooted in complex systems theory. We draw conclusions relating to the need for combined macro- and micro-level models of systemic risk. In Section 2.2 we review the discipline of agent-based modelling (ABM) and assess its suitability for modelling complex financial systems. In Section 2.3 we consider systemic risk with reference to three prominent case studies. We discuss how systemic risk can be measured in macro-level and micro-level systems. In Section 2.4 we look in detail at the macro-level systemic risk modelling literature. We consider channels for financial distress propagation through networks of various kinds, and identify key shortcomings regarding the treatment of price impact in macro-level models. In Section 2.5 we consider micro-level market phenomena, and, in particular, market microstructure modelling. We discuss how price impact

arises at the micro-level, and survey attempts to model extreme price impact events such as flash crashes. We find that microstructure models rarely feature a substantial number of assets, and so are unable to model systemic phenomena such as the propagation of flash crashes through networks of interconnected assets. In Section 2.6 we consider the state of the art of ABM calibration and validation. We discuss the challenges inherent in comparing ABMs to the real world, and identify the key approaches proposed to address them. We offer suggestions for studies that could further enhance the quantitative sophistication of ABM calibration and validation. In Section 2.7 we consider the role of ABMs as digital laboratories for the *in silico* evaluation of regulatory policies. We consider the needs expressed by regulators, and the potential for ABM to meet those needs, with a survey of both macro- and micro-level ABM policy investigations. We find that hybrid micro-macro policy examples are both lacking in the literature, and are necessary, if regulators are to fully understand the likely consequences of changes to market regulation before such interventions are made reality. We summarise our arguments regarding the need for a new micro-macro modelling paradigm in Section 2.8.

2.1 The Complex Nature of Financial Systems

It is well-established that modern financial systems are complex. They are frequently ascribed properties such as:

- emergence and self-organisation (Bouchaud and Challet, 2017)
- non-equilibrium dynamics and chaos (Johnson et al., 2003)
- interconnectedness across space and scales (Allen and Babus, 2008)
- heterogeneous, irrational and emotional participants (Lo, 2004)

The problem of reproducing the observational characteristics of financial systems in tractable models has a heritage measurable in decades. For example, Mandelbrot (1963) discusses the surprising distribution of asset price movements. Price changes are seen to be *leptokurtic*, or *fat-tailed*, meaning the probability of a large price change (usually a price fall, or *crash*) is greater than that which might be expected if price changes were normally distributed as they are assumed to be in the classical *Efficient Markets Hypothesis* (EMH) (Tsang and Martinez-Jaramillo, 2004).

The fat-tailed distribution of price changes remains a feature of modern financial markets (Johnson et al., 2003; Cont, 2001). The persistence of this, and other related phenomena that we discuss below, is perhaps surprising given the radical changes

markets have undergone since the introduction of electronic trading systems in the 1980s (Buchanan, 2015) and due to the ever increasing level of global interconnection (Krugman, 1991). During this time, DSGE models came to dominate economics, primarily as a result of their attractive analytical properties (Fagiolo and Roventini, 2017). However, during the 2007–2009 financial crisis and resulting Great Recession, DSGE models failed to either predict the unprecedented financial distress or to offer any solutions to the ongoing crisis (Fagiolo and Roventini, 2017).

The failure of DSGE models meant that economists and governments were increasingly keen to consider the alternative framework offered by complex systems theory (Farmer and Foley, 2009; Haldane and May, 2011; Haldane and Turrell, 2019). Instead of the closed-form analytic style of DSGE models characterised by what Colander (2005, p.930) describes as “the ‘holy trinity’ assumptions of rationality, greed and equilibrium”, complex systems theory offers models characterised by emergence, bounded rationality and heterogeneity. We discuss these aspects in more detail next.

2.1.1 Emergence

Hayek (1945) considers the emergent properties of financial systems and the interconnectedness of the participants therein: (p.256)

“The whole acts as one market, not because any of its members survey the whole field, but because their limited individual fields of vision sufficiently overlap so that through many intermediaries the relevant information is communicated to all.”

Hayek (1945)’s notion of a system “acting as one” as a result of the interactions of systemic components (generally referred to as *agents*) is what is meant by *emergence*. However, emergence implies more than simple aggregation. As an example, in Schelling (1971)’s model of social segregation, simple micro-level behavioural rules implemented as a cellular automaton (Langton, 1986) give rise to surprising macro-level properties, seemingly unpredictable from an understanding of the micro-level alone. This notion that behaviour at macro-levels can be radically different from any observable micro-level behaviour is what characterises emergence. The sequence of prices of financial assets resulting from the complex interactions of traders placing orders to buy and sell shares provides an obvious example from financial systems (Smith, 1962). The domino-effect failure of banks in the 2007–2009 crisis provides another (Haldane and May, 2011).

Models of agent interaction structures have developed from non-spatial differential equations modelling disease epidemics (Kermack and McKendrick, 1927), through cellular automata (Schelling, 1971; Langton, 1986) and more recently to incorporate graph-theoretic concepts (Newman, 2003) giving rise to network theory (Namatame and Chen, 2016). All of these approaches give rise to endogenous dynamics and produce emergent phenomena.

2.1.2 Bounded Rationality

The analytical tractability of DSGE models rests on an assumption of perfect utility-maximising rationality and perfect information availability on the part of the entities in the model (Fagiolo and Roventini, 2017). It is this property that admits of the “General Equilibrium” solution to DSGE models and similarly underpins the Efficient Markets Hypothesis. The rationality assumption, in particular, is criticised in the behavioural finance literature. Kahneman and Tversky (1979) introduce the notion of *cognitive bias* by identifying various circumstances under which human behaviour is predictably irrational. Lo (2004) seeks to explain the seemingly irrational behaviour of traders in terms of an emotional response, and similarly Bouchaud and Challet (2017) argue that agent herding, rather than utility optimisation, drives market prices. The boundedly-rational and locally-optimising agents of complex systems theory (Tsang and Martinez-Jaramillo, 2004) could therefore be argued to more realistically represent participants in real financial markets, compared with DSGE models.

2.1.3 Heterogeneity

Global financial systems are extremely heterogeneous (Farmer and Foley, 2009; Battiston et al., 2016b). Banks, insurers, hedge funds, manufacturers, stock exchanges, pension funds, high-net-worth individuals and households are just some of the agents in the global economy. Each of these participants has their own set of constraints, behavioural preferences and modes of interaction. Indeed, the interaction of multiple market participants, in particular banks and the issuers of mortgage related financial products, has been deeply implicated in the 2007–2009 crisis (Crotty, 2009; Haldane and May, 2011). The *representative agent* approach of DSGE models, where agents of a given class are taken to be identical and thus representable by a single individual is, by definition, incapable of modelling this kind of interaction. As Fagiolo and Roventini (2017, p.3, original emphasis) argue: “economists should consider the *economy as a complex evolving system*, i.e., as an ecology populated by heterogeneous

agents whose far-from-equilibrium interactions continuously change the structure of the system.”

2.1.4 Summary of Financial Market Complexity

The dominant models in use by economists and governments, DSGE models, rest on assumptions that are deeply at odds with observed properties of financial markets and those that participate in them. We have also seen arguments that DSGE models failed during the 2007–2009 crisis (we explore this in further detail in Section 2.3.1). In contrast to the simplifying assumptions required to produce analytically tractable models of financial systems, we have seen that the tenets of complex systems theory are aligned with the heterogeneous, emergent reality of financial markets.

However, important questions remain. Given the more realistic assumptions of complex systems theory, are the results from counterfactual experiments an improvement over those of DSGE models? How do complex systems models generate results? And what are the drawbacks of using this alternative class of models? We turn to these issues in the next section.

2.2 Agent-Based Modelling

The heterogeneous, interacting, decision-making entities featured by models in the complex systems theory literature are known as *agents* (Wooldridge and Jennings, 1995), and they are utilised in what are known as *agent-based models* (ABMs). However, not all entities in ABMs qualify as agents. We adopt the definition given by Wooldridge (2001, p.21):

“An *agent* is a computer system that is *situated* in some *environment*, and that is capable of *autonomous action* in this environment in order to meet its delegated objectives.”

The rich, emergent dynamics of ABMs means that it is quite possible that simulations never converge to stable equilibria. For this reason, the method of generating results from ABMs is quite different to the approach demanded in DSGE modelling. ABMs are not typically solved. Rather, repeated stochastic instantiations of simulations are conducted resulting in distributional outputs. This approach is known as *Monte-Carlo* modelling. The quality of the results are established by comparing a model’s ability to reproduce statistical regularities that match empirical observations (Fagiolo et al., 2007; Chen et al., 2012). A financial example is the fat-tailed distribution (excess

kurtosis) of price movements, mentioned in Section 2.1. These regularities are known as *stylised facts* and are discussed in detail in Section 2.6.

2.2.1 Agent-Based Computational Economics

The emergence of *agent-based computational economics* (ACE) as a research discipline (Tesfatsion, 2002) was facilitated by the availability of both powerful computers and high-quality data from financial institutions. As LeBaron and Tesfatsion (2008, p.249) state: “Limited but representative samples of real-world micro data provide an important check on the empirical plausibility of simulated micro-level distributions.” Since the influential survey by LeBaron (2000), financial ABMs have proliferated (LeBaron, 2016). A persistent difficulty concerns the modelling of detailed agent behaviour (Wellman, 2016). Rather than attempt to endow agents with highly sophisticated artificial intelligence, Gode and Sunder (1993) demonstrated that markets populated by simple agents with *no intelligence whatsoever* were capable of exhibiting very high, almost optimal, efficiency. This computational confirmation of earlier results from experiments with human participants by Smith (1962), was much celebrated as an example of the capability of the new modelling paradigm. Cliff et al. (1997) demonstrated (and rectified) a significant shortcoming of the model, and the *zero-intelligence* (ZI) approach became established as a useful method by which to avoid introducing bias into results. ZI agents, interacting through realistic price formation mechanisms (such as auctions, Wellman, 2011), have been repeatedly shown to accurately reproduce stylised facts of financial markets (Chen et al., 2012; Gould et al., 2013). We discuss these results in further detail below.

2.2.2 Agent-Based Models as Laboratories

In addition to providing insight into the underlying dynamical properties of financial systems, financial ABMs also provide an effective *in silico* test environment for experimenting with new policy interventions (Aymanns et al., 2018). Examples include testing the effectiveness of the Basel accords in comparison to a Pigovian tax (Poledna and Thurner, 2016) and testing various restrictions on the activity of traders (Jacob Leal and Napoletano, 2019). The ability to sample distributional parametrisations of candidate policies has obvious benefits in comparison with the alternative option of simply deploying regulation directly into the marketplace, including not least among them cost. The academic ABM community has vocally promoted this use of their models, e.g., Farmer and Foley (2009); Battiston et al. (2016b). Central

banks and government regulators have, in turn, strengthened ties with academia e.g., [Haldane and May \(2011\)](#); [Bookstaber et al. \(2018\)](#); [Haldane \(2016\)](#); [Haldane and Turrell \(2019\)](#).

2.2.3 Challenges of Agent-Based Modelling

Despite significant potential benefits of ABM, the effective utilisation of the methodology is not without challenge. Arguably chief among these is the difficulty of calibrating and validating ABMs against real data. The very expressiveness and freedom offered by agent-based modelling can leave models at risk of over-fitting due to the typically-significant number of parameters involved. Questions about whether and how to calibrate models (i.e., setting input parameter values based on empirical data) and how to validate (i.e., measuring success at reproducing a set of stylised facts) are central to the success of ABM as a discipline, and are discussed in detail in Section 2.6. However, consensus has been slow to emerge about how to undertake these operations ([Fagiolo et al., 2007](#); [Chen et al., 2012](#); [Fagiolo et al., 2019](#)).

The inherent variety and flexibility of ABM implementation approaches means that results can be difficult to generalise e.g., to inform policy, and difficult to compare between models. The lack of a standard approach to calibration and validation, as well as a lack of standards regarding other aspects of agent-based modelling such as how to describe models, whilst appearing sub-optimal, is arguably characteristic of a young science that is still in its early stages of development ([Llacay and Pfeffer, 2018](#)).

Despite these difficulties, regulators are becoming increasingly aware of the need to collect richer data from market participants, and to foster relations with academia, in order to move beyond the demonstrated inadequacies of classical economics and embrace the possibilities afforded by complex systems theory. The Chief Economist at the Bank of England offers a contemporary perspective ([Haldane, 2016](#), p.11):

“To be clear, ABMs are no panacea for the modelling ills of economics. In discussing them here, the implication is not that they should replace DSGE models, lock, stock and barrel. Rather, their value comes from providing a different, complementary, lens through which to make sense of our dappled economic and financial world, a lens which other disciplines have found useful when understanding their worlds or when devising policies to improve it.”

Although the use of ABMs for the prediction of real financial markets and price series is, in principle, possible ([Chen et al., 2012](#)), ABMs have already proven their ability

in the area of financial stability. It is to this specific area of application that we now turn.

2.3 Systemic Risk

Systemic risk is defined by [Schwarcz \(2008, p.204\)](#) as “the risk that (i) an economic shock such as market or institutional failure triggers (through a panic or otherwise) either (X) the failure of a chain of markets or institutions or (Y) a chain of significant losses to financial institutions, (ii) resulting in increases in the cost of capital or decreases in its availability, often evidenced by substantial financial-market price volatility”.

Since the 2007–2009 financial crisis, the systemic risk literature has blossomed ([Čihák, 2007](#); [Gai and Kapadia, 2010](#); [Haldane and May, 2011](#); [Battiston et al., 2012a](#); [Geanakoplos et al., 2012](#); [Di Gangi et al., 2018](#); [Greenwood et al., 2015](#); [Benoit et al., 2017](#); [Poledna and Thurner, 2016](#); [Serri et al., 2017](#); [Duarte and Eisenbach, 2018](#); [Zhou and Li, 2019](#)). Many alternative definitions of systemic risk have been proposed (as surveyed by, for example, [Bisias et al. \(2012\)](#)). However, the definitions tend to be united by the key role of *interconnectedness*. Indeed, the notion that the interactions of markets and institutions can cause chained failures to spread to a significant fraction of the financial system surely justifies the consideration of systemic risk as an emergent property of financial markets. As discussed in [Section 2.1](#), the tools of complex systems theory are appropriate for representing such systems.

In this section, we present three case studies, each of which illustrates a different way in which systemic risk arises in financial systems:

1. The 2007–2009 financial crisis.
2. The August 2007 “Quant Quake”.
3. The 6th May 2010 “Flash Crash”.

For each of these case studies, we identify the role of networks in propagating systemic risk. We further discuss metrics that have been developed to characterise systemic risk in these scenarios. Discussion of modelling approaches aimed at reproducing the phenomena exhibited by these case studies is deferred to [Sections 2.4 and 2.5](#).

2.3.1 The Financial Crisis of 2007–2009

In 2008, major banks were believed to be “too big to fail”. The subsequent collapse of Lehman Brothers sent shockwaves through the financial system, ushering

in a new *macroprudential* regulatory era. In contrast to the previous approach of *microprudential* regulation, focussed on assuring the stability of banks in isolation, macroprudential regulation acknowledges the need to protect the stability of the entire interconnected system (Haldane and May, 2011). The mitigation of systemic risk became a primary concern for regulators. The failure of established DSGE models to account for the effect of interconnectedness on systemic risk led regulators to seek alternative approaches: “Looking at financial risk through a network lens indicates a fundamentally different rationale for prudential regulation” (Haldane and May, 2011, p.354).

As Haldane and May (2011) discuss, the distress or failure of a bank can cause distress at other banks from which the failing bank has borrowed money. For example, although Lehman Brothers looked independently robust (Cont, 2017), it was vulnerable to distress caused by the failure of its counterparties. If a bank is unable to pay back its debt on time it is said to *default*. Creditor banks will lose some or all of the money lent in this scenario, which in turn may cause the creditor to find it difficult to repay its own debts, and so the distress propagates. Thus the *interbank network* of loans can act as a *contagion channel* for systemic risk. In Section 2.4 we present a detailed discussion of academic research into the modelling of systemic risk arising in interbank networks.

2.3.2 The Quant Quake of August 2007

In addition to the direct financial distress contagion discussed in the preceding section, systemic risk can also arise through indirect interactions. In the case of the August 2007 “Quant Quake”¹, distress propagated through the network of hedge funds connected indirectly by their overlapping investments. Hedge funds are known for their use of *leverage* (Ang et al., 2011), which is the practice of borrowing money to finance investments in financial assets (a familiar example of leverage is the residential mortgage). The assets act as collateral for the loans, and can be seized by the lender if the borrower is unable to repay its debt when required to do so. Leverage, defined as the ratio of assets to capital reserves, is required to be maintained below some pre-agreed ratio. Further discussion of the implications of this practice is given in Section 2.4.2.

¹The subset of hedge funds that specialise in the quantitative analysis of statistical market anomalies are known as *Quant* funds (Khandani and Lo, 2007). The crisis of August 2007 is also known as the “Quant Meltdown”, but a comparison of search engine results implies that the event is more widely referred to as the “Quant Quake”.

Khandani and Lo (2007) argue that in August 2007, distress at a single hedge fund caused it to reduce its leverage. The fund sold out of positive (*long*) investments, and bought back negative (*short*) investments. This act caused the price of the financial assets involved to move via the mechanism of *price impact* which is discussed in detail in Section 2.5.3. Thus, other hedge funds with similar investments saw prices move in loss-making directions among both their long and short investments. As discussed by Khandani and Lo (2007) and Cont and Wagalath (2013), these losses led other hedge funds to also reduce the size of their investments to control their leverage, and thus the financial contagion propagated between funds. Khandani and Lo (2007) and Cont and Wagalath (2013) argue that overlap at the trading strategy (or *risk factor*, Fama and French, 1993) level describes the method of contagion more clearly than consideration of individual investments in single financial assets. The secretive and complex nature of the hedge fund industry makes it unlikely that the full picture of August 2007 will ever emerge (Khandani and Lo, 2007), leaving room for alternative analyses and explanations, e.g., Cont and Wagalath (2013)’s emphasis on mutual funds in contrast to Khandani and Lo (2007)’s consideration of leverage.

2.3.3 The Flash Crash of 6th May 2010

Compared to the previous two case studies, the Flash Crash of 6th May 2010 is distinguished by its extremely short timescale. In the space of just 20 minutes, equity prices on US markets experienced their largest one-day fall in history, followed by an unprecedented rebound almost back to pre-crash levels (CFTC and SEC, 2010). Analyses of the events of 6th May 2010 have highlighted the importance of liquidity provision via automated order placement activity (*algorithmic trading*) and in particular *high-frequency trading* (HFT) (CFTC and SEC, 2010; Easley et al., 2012). Algorithmic trading, facilitated by the availability of direct electronic access to stock markets and real-time data feeds from those markets (Buchanan, 2015), removes humans entirely from the fine-grained buying and selling behaviour taking place on millisecond timescales in modern financial markets. Johnson et al. (2013) introduce the notion of a “robot phase transition” — a timescale below which financial transactions are entirely between automated systems. Interaction speed is ultimately limited by the laws of physics, leading HFT practitioners to pay large sums to *co-locate* their trading systems in the same physical space as the computers running stock exchanges in order to save precious milliseconds (Buchanan, 2015).

CFTC and SEC (2010) and Nanex (2010) report that the turmoil on 6th May 2010 was triggered by a significant (but not excessive) series of sell orders by a single firm

(Waddell and Reed) in a single financial instrument (the Standard and Poor’s 500 Equity Index e-Mini Future, one of the most actively traded financial instruments in the world). It has been reported that HFT firms absorbed the sell orders until their investment limits were reached, forcing them to also commence selling. In this second round, HFTs rapidly sold to each other in what has been described as a “hot potato” effect. Other exacerbating effects were also present, including the decision by other market participants to avoid trading due to the highly unusual patterns of activity. The removal of this usual flow of orders caused prices to destabilise further. Crucially, what began in a single financial asset ended up engulfing hundreds of equities as the distress spread. Confidence in the market was further shaken when regulators attempted a *post hoc* damage limitation exercise, by cancelling trades that had occurred during the most extreme phase of the crash (CFTC and SEC, 2010). The backlash over this practice led to the establishment of consistent trade cancellation policies in US markets².

A Flash Crash on the scale of 6th May 2010 has yet to recur, however *mini flash crashes* have been reported on a regular basis (Johnson et al., 2013; Calcagnile et al., 2018). These are both smaller in terms of the magnitude of price movements, and also in terms of the number of financial assets affected. The mechanisms at work, however, are argued to be similar (Calcagnile et al., 2018, p.14):

“Why has the synchronization among different assets increased through the recent years? We hypothesize that a major role is played by the dramatic increase of algorithmic trading. Thanks to the technological innovation, faster information processing is responsible for the more rapid propagation for large price movements through different assets.”

Several ABMs have been proposed that reproduce some of the phenomena observed on 6th May 2010 and we discuss these in detail in Section 2.5.4.

In addition to the mini flash crashes that are frequently observed in equity markets, significant flash events have been observed across a wide range of asset classes. Although the precise details differ between flash crashes, common features include the activity of algorithmic trading systems, often during periods of relative market inactivity (such as outside of regular business hours, in the case of the Swiss currency de-peg incident of January 2015 (Breedon et al., 2018), or the Sterling incident of October 2016 (Jackson et al., 2017)). Other selected flash events include a major

²<https://www.sec.gov/rules/sro/bats/2010/34-62886.pdf>, accessed 2019-10-23.

October 2014 US treasury bond flash rally (bonds move inversely to equities by construction) which remains only partially explained (Levine et al., 2017), as well as the February 2018 volatility index (VIX) flash crash³ which triggered collapse of the so-called inverse VIX *Exchange Traded Fund* (ETF). This incident again occurred during a quiet market period, when liquidity supply was much lower than during the regular trading day. Similar circumstances surround two significant foreign exchange flash crashes in 2019⁴.

Despite a steadily improving understanding of the flash crash phenomenon, it would seem that the markets remain vulnerable to this form of algorithmically driven systemic risk.

2.3.4 Measuring Systemic Risk

As illustrated by the above case-studies, systemic risk can take a variety of forms depending on the nature of the entities involved, and their interactions. It is therefore appropriate to consider the question of how systemic risk should be quantified in these different scenarios. Measurement of systemic risk offers two key benefits, both of which can inform regulatory policy:

1. Measurement facilitates possible early-warning systems based on longitudinal observation of metrics.
2. The marginal contribution to systemic risk by different institutions allows interventions such as systemic risk taxes to be deployed.

Network Metrics

By representing the network of loans made between banks as a weighted adjacency matrix, systemic risk is characterised by the largest eigenvalues of this matrix (Brede and Sinha, 2005; Caccioli et al., 2014; Acemoglu et al., 2015b; Bardoscia et al., 2017). Battiston et al. (2012b) introduce the *DebtRank* algorithm which uses notions of network eigenvector centrality to determine the key institutions that cause systemic risk. Acemoglu et al. (2015b) take a similar approach with their *harmonic distance* which relates to susceptibility of banks to the distress of others. *Value-at-risk* (VaR) captures the distribution of expected losses (it is given as a confidence interval). Adrian and Brunnermeier (2016) extend the concept to networked systems by introducing *CoVaR*, the VaR conditional on the failure of an institution. VaR has been

³<https://cnb.cx/2FPsMgh>, accessed 2019-10-22.

⁴<https://reut.rs/2N2184L>, accessed 2019-10-22.

criticised for failing to adequately quantify *tail risk*, i.e., extreme losses in the tails of the distribution. As we saw in Section 2.1, financial systems often exhibit fat-tailed distributions and so the Gaussian assumption underlying VaR leaves it open to criticism. This argument leads Acharya et al. (2017) to introduce the *systemic expected shortfall* (SES), which is based on the total loss of all banks in the system conditional on the failure of one bank. Acharya et al. (2017) also introduce the *marginal expected shortfall* (MES) which attributes SES among banks according to the losses attributable to each individual institution. SES and MES are argued to more accurately reflect tail risk. Hué et al. (2019) use *Granger causality* to assess the systemic importance of a set of 90 large banks by considering the propagation of market price dislocations between banks. Duarte and Eisenbach (2018) introduce an *aggregate vulnerability index* to assess systemic risk due to indirect bank connections via overlapping portfolios and fire sales. Other practitioners have made connections between network entropy and financial instability (e.g., Billio et al., 2016; Pele and Mazurencu-Marinescu-Pele, 2019). Braverman and Minca (2018) introduce further systemic risk metrics in the context of overlapping asset portfolios, this model has the unusual feature of endogenous price impact arising from the activity of a simplified market-maker agent.

It is clear from the range of systemic risk metrics that have arisen that the appropriate choice of metric may differ between model classes, and as a function of the researcher's objectives. For example, is the goal to assess the scale of the shock to the system in terms of the probability or extent of financial contagion (Caccioli et al., 2014), or to evaluate individual entity contributions to systemic risk (Battiston et al., 2012b)? Despite significant progress, a decade of alternative systemic risk metric proposals suggests the matter remains unresolved (Cont and Schaanning, 2019).

Crowding Metrics

When systemic risk is propagated via overlapping investments in either assets, or in risk factors, it is necessary to quantify the degree of overlap or *crowding*. An early metric of this kind is the Herfindahl index (Rosenbluth, 1955) which measures the degree of diversity of *market share* in a given market. Where the market is defined by two banks that may invest in two assets, Beale et al. (2011) demonstrate that the optimal strategy for a single bank (full diversification, i.e., equal investment in either asset) corresponds with maximal crowding, which is the worst outcome as far as a regulator responsible for systemic risk of the entire system is concerned. Beale et al. (2011) dub this the “Regulator’s Dilemma”, and it can be seen that this

reflects the “Prisoner’s Dilemma” where the dominant strategy equilibrium results in suboptimal social welfare. Bruno et al. (2015) and Sias et al. (2016) propose metrics to compare the degree of overlap of the set of investments (the *portfolios*) held by two institutions. In both works, overlap is defined in terms of cosine distance. Cahan and Luo (2013) introduce a similar metric for crowding of trading strategies based on pair-wise correlations between sets of assets.

Flash Crash Metrics

Metrics for the detection of flash crashes (or indeed mini flash crashes) are varied. Johnson et al. (2013, p.3) propose the following:

“For a large price drop to qualify as an extreme event (i.e., black swan crash) the stock price had to tick down at least ten times before ticking up and the price change had to exceed 0.8%. ... In order to explore timescales which go beyond typical human reaction times, we focus on black swans with durations less than 1500 milliseconds”

Vuorenmaa and Wang (2014) emphasise the characteristic price rebound effect observed in flash crashes, and define a flash crash as a price fall of 10% within a short time interval, followed by a rebound of at least 50% of the size of the loss. Jacob Leal et al. (2016) define a flash crash as a price fall of 5% with significant reversion within 30 minutes. Based on empirical data from the Tokyo stock exchange following the introduction of a platform supporting HFT, Jain et al. (2016) define a high-frequency *tail event* as the 5th percentile of negative returns over a one-minute time horizon. Aldrich et al. (2016) claim to have calibrated their metric against the real Flash Crash, defining it as a one-minute price movement of 5% or 200 spreads (the spread is the difference between the most competitive buy and sell prices for the financial asset — see Section 2.5.1). In contrast, Paddrik et al. (2012) look at the minimum price reached by assets in simulation, and do not consider the time dimension. Oriol and Veryzhenko (2019, p.1080) raise the important issue of measurement granularity: “we compute 10-second absolute midquote returns. Brogaard et al. (2018) explain this frequency as a compromise between two important issues: (1) an analysis of the extreme price movements that emerge from the brief liquidity dislocation requires a short sampling interval (2) too short sampling frequency may split an extreme price movement into small price variations and the total crash or jump will be not captured”. The variety of these metrics shows that there is no single accepted quantitative definition of a flash crash. The selection of an appropriate metric is considered on a case-by-case basis,

building on the minimal consensus position that a valid flash crash metric should incorporate both the speed, and the magnitude, of price movements.

2.4 Macro-level Phenomena

In Sections 2.1 and 2.2 we argued that the methods of complex systems theory, especially agent-based modelling, offered a compelling alternative approach by which to study financial systems. We reviewed arguments as to why mainstream DSGE models had failed during the 2007–2009 financial crisis, and in Section 2.3 we surveyed how the 2007–2009, and other, crises gave rise to different forms of systemic risk. In this section we unite these themes by reviewing the state of the art of systemic risk modelling. We characterise interbank networks of the kind introduced in the previous section as *macroscopic*, because, as we will show, they ignore important features of the *microscopic* details of financial asset transactions (such as those implicated in the 6th May 2010 Flash Crash).

Thorough surveys of macro-level phenomena and systemic risk have been produced by Glasserman and Young (2016), Benoit et al. (2017), Aymanns et al. (2018), and Caccioli et al. (2018). A full survey of this burgeoning literature is beyond the scope of the present chapter. However, there is consistency in the systemic risk contagion channels identified by these surveys, and in the sections that follow we consider in detail the four key channels:

1. Direct contagion due to the failure of debtors.
2. Direct contagion due to reluctance of creditors to continue to provide loans (known as *liquidity hoarding*).
3. Indirect contagion due to overlapping portfolios.
4. Indirect contagion due to overlapping trading strategies.

2.4.1 Interbank Networks

Seminal works by Allen and Gale (2000) and Eisenberg and Noe (2001) demonstrated how *balance sheet* (asset and liability) connections between banks could form sufficient network infrastructure for financial contagion to take place. These models predicted that the failure of one bank could cause knock-on effects leading to the failure of other banks in the system. Indeed, this phenomenon was subsequently observed during the 2007–2009 crisis. Energised, the interbank modelling literature has grown rapidly in the years since the crisis.

Gai and Kapadia (2010) introduced the notion that interbank networks are “robust, yet fragile”. That is, the authors found, using simulation, that more densely connected interbank networks were more robust to shocks (less likely for defaults to occur), but that, if defaults *did* occur, the extent of distress would be larger. The proposed explanation is that having diversified counterparties allows the risk due to small shocks to be shared out, making defaults less likely. However, if defaults occur, higher diversification means that it is more likely that any given bank will be affected. Acemoglu et al. (2015b) and Zhou and Li (2019) present similar arguments to support their findings that more complete networks are more stable under low shock regimes, but less stable under high shock regimes. The importance of network topology applies both to direct interbank networks, but also, as we shall see, to networks of asset portfolios owned by banks.

Models such as that reported by Arinaminpathy et al. (2012) consider alternative interbank contagion channels. In addition to the direct loss contagion described above, it is also possible for banks, when facing financial distress, to decline to continue to provide loans to other banks. In this scenario banks are said to *hoard liquidity* and the knock-on effect to other banks is similar to the direct loss contagion. If banks have their loans recalled, they may struggle to meet their own obligations, and subsequently default.

However, Arinaminpathy et al. (2012) do not restrict themselves to these two channels. The third channel, shared asset indirect contagion, is of central importance to this and many other prominent models e.g., Nier et al. (2007), and the remainder of this section considers such models. Battiston et al. (2016a) find that the shared-asset channel may in fact have the highest impact on systemic risk, and Arinaminpathy et al. (2012) note that “in combination, these channels mutually reinforce each other in a non-linear way”.

2.4.2 Leverage and Margin Calls

In order to appreciate the role played by overlapping investments in propagating systemic risk, it is necessary to understand how investments are financed. As mentioned in Section 2.3.2 above, hedge funds, and indeed banks, make use of loans to support their investments. The ratio of the value of the investments to the value of the underlying capital is known as leverage. Leverage is offered by the brokerage business of investment banks (Ang et al., 2011) and must be maintained within agreed limits. If a leveraged fund suffers losses due to price movements of assets within its portfolio, its leverage increases. The fund is required to return leverage to agreed

limits and, in order to enforce this behaviour, the creditor bank may issue a *margin call* (Brunnermeier and Pedersen, 2009). The fund must then either provide more cash to the bank, or reduce the size of its investments (*deleverage*). As Brunnermeier and Pedersen (2009) describe, and as represented in Figure 2.1, a problematic feedback effect known as a *margin spiral* can occur if a fund is forced to sell its assets due to losses. The action of selling further depresses prices, worsening losses, and so requiring further rounds of *distressed selling*.

Systemic risk arises when different funds have investments in the same assets. Here, distressed selling by a fund causing prices to fall via price impact will cause losses at all other funds with investments in the distressed assets. These other funds may then be required to reduce the size of their investments in order to manage their leverage and so the distress propagates (Shleifer and Vishny, 2011). We review models of this phenomenon, known as a *fire sale*, in the next section.

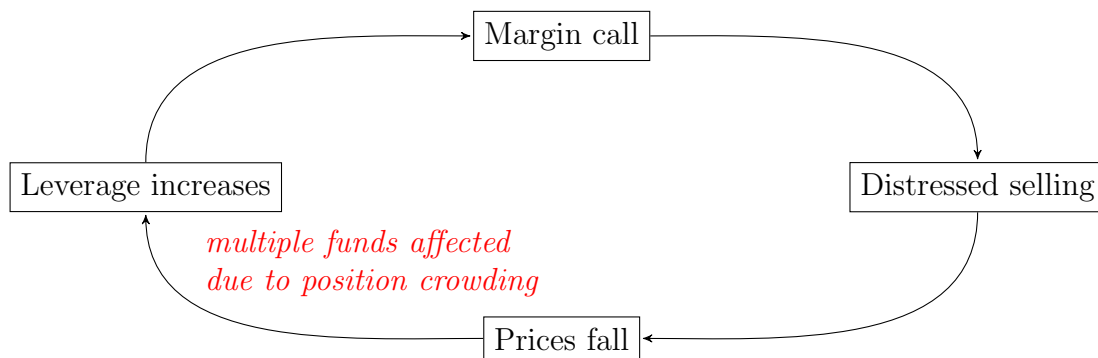


Figure 2.1: Margin spiral, adapted from Brunnermeier and Pedersen (2009). This figure demonstrates the connection between fund losses and price movements, as funds seek to reduce leverage to contractually-agreed limits. It also indicates the point at which distress at one fund propagates to others via shared asset holdings (*crowding*).

2.4.3 Shared Asset Networks and Fire Sales

Models of interbank networks often consider, in addition to direct loans made between banks, investments made in a pool of assets. Nier et al. (2007), Adrian and Shin (2010) and Shleifer and Vishny (2011) established the importance of the *shared asset* contagion channel for systemic risk that results from the distressed fire-selling described in the previous section. Battiston et al. (2016a) report that disregarding systemic risk due to fire sales may significantly underestimate overall systemic risk. Huang et al. (2013) and Caccioli et al. (2014) investigate the effect of the topology of

the network defined by overlapping asset portfolios. These works report phase transitions, whereby below critical values of leverage the system is generally stable, and above which instability is extremely likely. Interestingly, [Caccioli et al. \(2014\)](#) find that, at high levels of asset diversification, default cascades are less likely to occur, but are more severe when they do occur. This is a further example of the “robust yet fragile” behaviour introduced by [Gai and Kapadia \(2010\)](#). [Caccioli et al. \(2015\)](#) initialise the interbank loan network using real data (we discuss the use of real data in ABMs in detail in Section 2.6), however the network of bank investments in assets is not constrained by real data, and a single representative asset is used in the model. [Ma et al. \(2019\)](#) develop a two-contagion-channel model, which is qualitatively validated using Chinese banking sector data. They find, as with previous studies, that the shared asset contagion channel is highly relevant to the propagation of systemic risk, dominating interbank contagion in cases where an initial shock is delivered to an asset, as opposed to a bank.

[Capponi and Larsson \(2015\)](#) take a different approach and explore the effect on systemic risk when asset owners strategically select a subset of their investments to reduce in size. They find that it is both individually and systemically, optimal for banks to preferentially reduce investments in more *liquid* assets, i.e., those which exhibit the least price impact in their model. We discuss liquidity in detail in Section 2.5.3. A similar conclusion is reached by [Duarte and Eisenbach \(2018\)](#) and [Cont and Schaanning \(2019\)](#).

[Cont and Schaanning \(2017\)](#) present a model of multiple institutions investing in multiple assets, and characterise systemic risk in relation to the level of portfolio overlap. Using empirical data on the European banking system, the authors argue for the importance of accounting for shared asset contagion in models of systemic risk. This paper is illustrative of a conspicuous trend in macro-level interbank and shared asset models — price impact is treated in an analytically convenient, but simplistic manner. In this case, as a linear function ([Cont and Schaanning, 2017](#), p.16):

“A key assumption in models of fire sales spillovers concerns the impact of asset liquidations on market prices. This may be summarized in the choice of a market impact function Ψ_μ , which defines the correspondence between the liquidation size q (in monetary units) and the relative price change for each asset class μ ... An adequate choice for Ψ_μ should be increasing, concave, satisfy $\Psi_\mu(0) = 0$ and lead to non-negative prices. Common specifications are the linear model”

Indeed, we also find that linear models of price impact are commonly deployed in models of shared-asset contagion (see Table 2.1). Given the critical importance of price impact in shared asset models, it might be expected that this aspect of models receives a great deal of attention. However, as Table 2.1 illustrates, this aspect of shared-asset models is often treated in a simplified manner. This point is also raised by Benoit et al. (2017). Braverman and Minca (2018) offer an improved price impact model whereby prices respond to trader demand via a market clearing demand-supply optimisation. This model does not account for details of order placement, market liquidity, or crisis price impact.

Paper	Price impact	Assets	Investment
Nier et al. (2007)	exponential†	single	long
Gai and Kapadia (2010)	exponential†	single	long
Arinaminpathy et al. (2012)	exponential†	multiple‡	long
Cont and Wagalath (2013)	linear	multiple	long/short
Huang et al. (2013)	linear	multiple	long
Caccioli et al. (2014)	exponential†	multiple	long
Greenwood et al. (2015)	linear	multiple	long
Caccioli et al. (2015)	linear	single	long
Battiston et al. (2016a)	linear	multiple	long
Serri et al. (2017)	linear	multiple	long
Cont and Schaanning (2017)	linear	multiple	long
Braouezec and Wagalath (2018)	linear	single	long
Braverman and Minca (2018)	endogenous	multiple	long
Di Gangi et al. (2018)	linear	multiple‡	long
Zhou and Li (2019)	linear	multiple	long
Ma et al. (2019)	exponential†	multiple	long

Table 2.1: A survey of price impact functions and portfolio investment constraints utilised in shared-asset models of financial contagion. †corresponds to “linear market impact for log-prices” (Caccioli et al., 2014). ‡represents the set of e.g., equities or bonds by a single representative *asset class*. *Long* and *short* refer respectively to positively and negatively signed investments.

What is perhaps particularly surprising is that empirical studies have shown that price impact under equilibrium market conditions does indeed have a simple form — however it is a concave function of order size such as a square root or logarithm, rather than linear (Zarinelli et al., 2015; Bouchaud, 2018). Models that are capable of endogenously producing these kind of price-impact functions are the purview of *market microstructure* analysis and are discussed in detail in Section 2.5.3.

What is clear from the preceding analysis is that models of shared-asset contagion do not consider the sub-daily (*intraday*) fine-grained mechanics of buying and selling on stock markets that gives rise to asset price movements in reality. The incorporation of realistic price impact functions in the form of models of the supply and demand of financial assets remains an important direction for future research.

Finally, we note in Table 2.1 that the majority of papers surveyed do not consider *short selling* (the action of borrowing assets to sell in the hope of repurchasing at a lower future price, equivalent to taking negative positions). Short sellers face additional risks that could act as crisis amplification mechanisms. A particular concern is the *short squeeze* (Dechow et al., 2001). Because asset prices can rise without limit, potential losses on short positions are unbounded. Even in the absence of leverage constraints, rising prices could force investors to buy to cover their short positions, pushing prices higher still.

2.4.4 Risk Management

Systemic risk associated with coordinated trading activity has been studied in the *crowded trades* literature (Stein, 2009; Sias et al., 2016). As previously mentioned, Khandani and Lo (2007) and Cont and Wagalath (2013) implicate strategy crowding in models of the Quant Quake of August 2007. Rather than look at individual investments, Khandani and Lo (2007) consider the overlap of risk factors (Fama and French, 1993), suggesting that hedge funds were crowded into (*exposed to*) a specific risk factor. Both Khandani and Lo (2007) and Cont and Wagalath (2013) invoke portfolio size management at hedge funds as the contagion amplification mechanism, though Khandani and Lo (2007) propose a leverage management mechanism (Brunnermeier and Pedersen, 2009) whereas Cont and Wagalath (2013) consider investors withdrawing capital from a *mutual fund* (a collective investment pool). A similar model, not specifically related to the Quant Quake, is proposed by Delpini et al. (2019). Braun-Munzinger et al. (2018) investigate the systemic impact of strategy crowding using an ABM. The model is able to generate endogenous price cycles, and the authors claim a strong ability of the model to reproduce stylised facts (we discuss model validation in Section 2.6). Menkveld (2015) relates crowding to systemic risk, and derives analytic expressions for the extent of crowding. However, the approach relies on strong assumptions, e.g., regarding the Gaussianity of price movements.

Sias et al. (2016) perform an empirical analysis of the hedge fund sector, and argue that hedge funds are no more crowded than other sectors of the investment community. This analysis is based on the cosine similarity between portfolios as a

crowdedness metric, and it contributes the important observation that “‘no portfolio overlap’ is not the null hypothesis” (Sias et al., 2016, p.8).

Finally, Bruno et al. (2015) propose a model where prices follow random walks, but “copycat” risk management practices cause crowding to emerge regardless. The authors find that systemic risk is maximised when all institutions share the same risk management model.

Strategy crowding implies position crowding, but the converse does not necessarily hold⁵. Models that consider strategy crowding as a source of systemic risk have demonstrated the importance of this metric (Cont and Wagalath, 2013), however there are few such models. Further work in this area could provide insights that would help inform policy — a particularly difficult regulatory challenge, given the secrecy with which market participants guard their trading strategies.

2.5 Micro-level Phenomena

What is missing from the macroscopic investigations reported in the previous section is the specification and utilisation of a detailed and comprehensive model of crisis price impact. Such an impact model falls within the domain of market microstructure analysis.

Linear price impact functions fail to capture the emergent and time-dependent nature of market liquidity (the ability to freely buy and sell assets at current prices), the study of which falls within the market microstructure literature (Gould et al., 2013). Events such as the Flash Crash of 6th May 2010 (CFTC and SEC, 2010) demonstrate that price impact during crisis periods can be extreme. On that occasion, asset prices exhibited unprecedented volatility as the orderly provision of market liquidity broke down. The event was also characterised by the high speed at which prices fell and subsequently returned almost to pre-crash levels. Automated algorithmic trading systems were implicated in all phases of the Flash Crash, including triggering the event, sustaining price falls, and facilitating eventual reversals (CFTC and SEC, 2010).⁶

⁵We thank Albert Menkveld for a useful discussion on this topic.

⁶Flash crashes are characterised in part by behaviours that no rational human would exhibit: for example Apple shares trading for \$100,000 on May 6th 2010. As pointed out by Madhavan (2012, p.34) “Security prices can exhibit sharp price movements over short time intervals even in a centralized market structure ... but unlike the flash crash, stocks did not trade at absurd prices”. For this reason we focus our analysis on electronic markets such as equity and futures markets, where intraday dynamics are heavily influenced by the behaviour of trading algorithms.

2.5.1 Market Microstructure and Limit Order Books

In the large majority of modern financial markets, asset prices are determined via the interaction of traders placing orders to buy and sell via *continuous double auctions* (CDAs, [Wellman, 2011](#)). In such an auction, two main kinds of orders are accepted by the *exchange* (the business providing the infrastructure to conduct the CDA, for example the New York Stock Exchange).

- Limit orders – orders specifying quantity and a maximum (minimum) price for buys (sells).
- Market orders — orders specifying quantity only, that may match at any available price point.

Orders to buy and orders to sell are stored in separate queues, sorted by price (descending for buy orders, ascending for sells) and then by time. Thus orders arriving earlier have an earlier position in the queue than later orders at the same price. The data structure representing the sorted queues is known as the *limit order book* (LOB), ([Gould et al., 2013](#)), and is depicted in Figure 2.2.

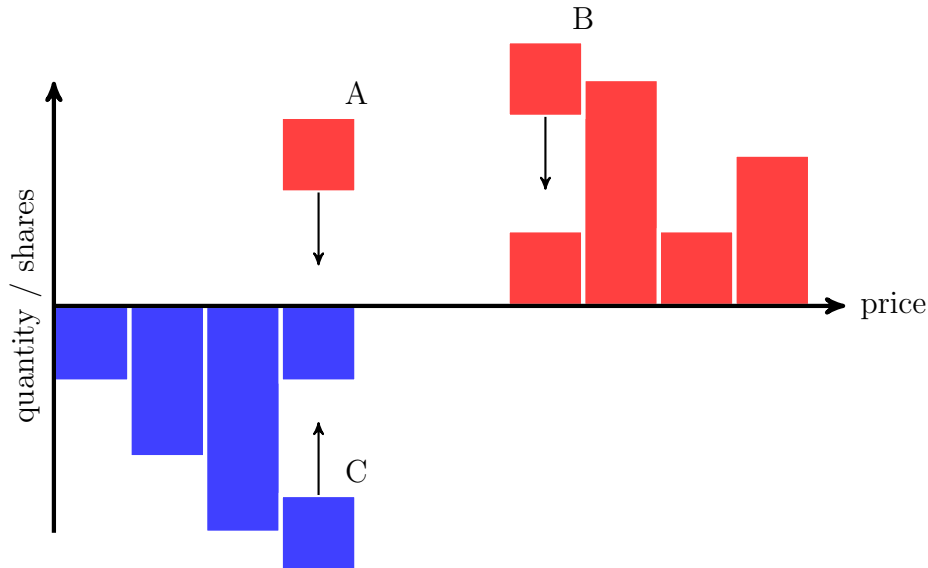


Figure 2.2: Schematic representation of the limit order book (LOB). Orders to buy (sell) are shaded in blue (red).

Figure 2.2 schematically illustrates the operation of the LOB. The figure represents the quantity of buy orders (blue, negative by convention) and sell orders (red, positive by convention) available on the LOB at a particular time. The gap between the highest-priced buy order (the *best bid*) and the lowest-priced sell order (the *best ask*)

is called the *spread*. Since permissible order prices are quantised, the spread has a minimum size of one *tick* (Figure 2.2 depicts a spread of three ticks). The size of the spread is a common metric by which to judge the quality of markets. To see this we note that in order for a transaction to occur, it is necessary for one party to *cross the spread*, i.e., a buyer must accept at least the price offered at the best ask, and conversely for sellers. Figure 2.2 represents a spread-crossing sell order (A). Such orders are described as *aggressive*, whereas the orders designated (B) and (C) in Figure 2.2 will join the queue of orders that is already established at their respective *passive* price levels. A narrow spread implies the cost of transacting is minimised for both buyers and sellers, which is a characteristic of a *liquid* market. Thorough discussions of the operation, modelling and statistical properties of LOBs are provided by Wellman (2011) and Gould et al. (2013), and we refer the reader to these works for further discussion of the mechanics of order interaction via CDA trading. It is important to note that the LOB is active for a fixed time window of each trading day, and not on a continuous basis, though in some markets the window is close to 24 hours, for example *index futures* (derivative contracts constructed to track the combined price movement of a set of underlying assets). Thus, the time domain of interest for microstructure analysis is strictly intraday.

2.5.2 Microstructure Modelling

The use of agent-based models to investigate the mechanisms of asset price formation is well-established in the literature e.g., Gode and Sunder (1993); LeBaron (2000); Tesfatsion (2002); Farmer et al. (2005). Indeed, heterogeneous mixtures of zero-intelligence agents interacting through realistic price formation mechanisms have been repeatedly shown to accurately reproduce the stylised facts of financial markets (Tesfatsion, 2002; Zovko et al., 2002; Farmer et al., 2005; Chen et al., 2012; Gould et al., 2013; Jacob Leal and Napoletano, 2019; Kononovicius and Ruseckas, 2019).

Chiarella et al. (2002) introduce an ABM featuring zero-intelligence traders along with populations of trend-following *chartists* and *fundamentalists* who possess a private valuation for the asset being modelled. In this model, chartists act to destabilise markets by enhancing bubbles and crashes, whereas fundamentalists act to stabilise markets by returning the price to long-run averages (so-called *mean reversion*). In their later work, Chiarella et al. (2009) argue that chartists, through their destabilising actions, give rise to the fat-tailed distribution of price changes observed in real markets. Thurner et al. (2012), however, construct a single-asset model containing only fundamentalists, which also exhibits fat-tailed distributions of price changes.

The authors find that leverage management of the kind discussed in Section 2.4.2 forces funds to occasionally sell into falling markets, exacerbating the falls via margin spirals (Brunnermeier and Pedersen, 2009). In an influential study, Darley and Outkin (2007) demonstrate an ABM that reproduces stylised facts of the NASDAQ stock exchange. The model is used to predict the effect that a pending change to market structure (decimilisation of price increments) will have on market quality. The model featured sophisticated agents that could adapt their trading based on profit expectations, and also adapt at a population level via evolutionary selection. Collver (2017) demonstrates the US regulator’s continued interest in microstructure modelling, in this case for understanding the effect of proposed changes to asset tick sizes, by producing a single-asset microstructure ABM.

Other, non-ABM, approaches to microstructure modelling include pure econometric models of patient and impatient traders under equilibrium conditions (Foucault et al., 2005) and Hawkes “jump” processes (Calcagnile et al., 2018). It is not clear how these alternative approaches might be integrated with complex network theory, and we proceed with consideration of microstructure ABMs in this thesis.

2.5.3 Liquidity and Price Impact

Of particular relevance to our discussion is the study of price impact models. Works such as Tóth et al. (2011), Zarinelli et al. (2015) and Bucci et al. (2019) have been very successful at characterising price impact of orders and *meta-orders* (large orders broken into small pieces in an attempt to minimise impact). However, these models assume equilibrium markets that are not under distress. Two key and related mechanisms mean that crisis price impact is different from equilibrium price impact. Cristelli et al. (2010) and Bookstaber et al. (2016) argue that, during crises, market participants refrain from placing orders. Brogaard et al. (2018) analyse extreme price movements in US markets and offer evidence that HFT firms switch from providers to consumers of liquidity during crises. Easley et al. (2012) relate these phenomena to the inferred presence of *toxic order flow*, i.e., strongly directional price movements that are outside non-crisis bounds. This removal of liquidity leads to sparsely populated LOBs, and as discussed by Farmer et al. (2004), gaps on the LOB can lead to further large price movements. Over longer timescales, there is considerable evidence that market participants regulate their liquidity provision in response to asset price shocks (Anand et al., 2013; Cheng et al., 2017), toxic order flow (Da et al., 2011) and volatility (Nagel, 2012).

Hence price impact during crises has a very different form from equilibrium price impact, and strongly depends on the details of liquidity provision. If market liquidity dries up and the LOB becomes sparsely populated, arbitrarily small market orders can result in arbitrarily high price impact. To see this, it is useful to consider the phenomenon of *stub quoting*. When financial intermediaries known as market makers (Gould et al., 2013) are contractually required to maintain open orders on the LOB at all times yet do not wish to trade, they may move their orders far away from *fundamental* (economically rational and prevailing) prices, and place orders at fractions of a penny, or indeed for extremely high prices, with no expectation that their orders will be matched in the market. However, during crises such as the 6th May 2010 Flash Crash, such stub quotes were indeed matched, resulting in wild price fluctuations (*volatility*) (CFTC and SEC, 2010).

2.5.4 Flash Crash Simulation

In the context of market microstructure, studies of financial stability focus on the extreme price impact of events such as flash crashes (Johnson et al., 2013). Recent research has demonstrated the success of agent-based approaches in investigating intraday market stability, and in particular investigating the role of HFT in market quality and flash crashes e.g., Hanson (2011); Johnson et al. (2013); Paddrik et al. (2012); Vuorenmaa and Wang (2014); Budish et al. (2015); Jacob Leal and Napoletano (2019). Although these models are very successful at reproducing stylised facts of financial markets, including during crises, attempts to extend these models to multi-asset or multi-venue settings have been extremely limited.

Kirilenko et al. (2017) and Golub et al. (2012) take purely empirical approaches to understanding the causes of the Flash Crash of 6th May 2010. Kirilenko et al. (2017) uses regression analysis on a unique dataset labelled with the identities of market participants. This analysis demonstrated that HFT activity resulted in a hot-potato effect exacerbating price declines. Golub et al. (2012) looks at order placement activity in markets one minute either side of mini-flash crash events (Johnson et al., 2013) finding HFTs withdraw liquidity when it is most needed, at the same time as market makers and other liquidity providers. Ben-David et al. (2012), Madhavan (2012) and Menkveld and Yueshen (2019) each explore a prominent contagion channel, implicated in the events of 6th May 2010 by CFTC and SEC (2010), whereby the arbitrage of S&P 500 e-mini futures against the “SPY” ETF led to distressed selling pressure in the e-mini market spilling over into, most prominently, equities. There is strong evidence that the breakdown of arbitrage, exacerbated by venue fragmentation,

played a major role in propagating distress between assets on 6th May 2010 ([Menkveld and Yueshen, 2019](#)). However, other contagion channels for flash crash propagation are possible ([Abrol et al., 2016](#)), and have been implicated in more recent flash crash events. For example, in its official report into the Sterling Flash Crash of 7th October 2016, the Bank for International Settlements found evidence that mechanistic portfolio management practices, rather than arbitrage, amplified distress: “Correlated and mechanical responses to shocks, including an unwind of crowded positioning or derivatives hedging flows, can amplify price movements and jeopardise market liquidity.” ([Jackson et al., 2017](#), p.15). It is surprising that such alternative channels have previously received little attention in the flash crash simulation literature, as these channels have the potential to coordinate large scale deleveraging by institutional market participants, and thus act as another channel capable of giving rise to systemic flash crash events. We are unaware of any investigations into the interaction between alternative flash crash contagion channels.

[Hanson \(2011\)](#), [Paddrik et al. \(2012\)](#), [Vuorenmaa and Wang \(2014\)](#), and [Jacob Leal and Napoletano \(2019\)](#) use combinations of agents representing low and high frequency market participants implementing fundamental (holding opinions regarding true underlying asset values) and chartist (trend following, without reference to true underlying value) strategies⁷ and are successful at reproducing the characteristic price fall and recovery observed during flash crashes. Indeed, the price recovery observed during flash crashes is another way in which crisis price impact differs from the linear models used in the fire-sales literature. [Virgilio \(2020\)](#) develops an ABM featuring low, and high-frequency traders, and finds that HFTs contribute to market volatility by exploiting their informational advantage over slower market participants and creating circumstances for destabilising orders by low-frequency traders.

During the Flash Crash of 6th May 2010, financial distress was observed to spread throughout equity and futures markets, making it a systemic event. However, microstructure models typically consider a single asset in isolation and do not incorporate network effects. This is a significant limitation of current approaches, and further research is necessary if the systemic risk due to flash crashes is to be successfully modelled and analysed.

⁷The need for such a mixture of agent types is well-established in the literature. For example, [Chiarella et al. \(2002, p.350\)](#) note that “all three trader types are necessary in some form to generate realistic looking return dynamics”.

2.5.5 Splash Crash Simulation

Buchanan (2015, p.163) reports:

“Some investors speculate about a ‘splash crash’, in which a massive spike in one market disrupts or freezes trade in foreign exchange, futures, commodities, bonds and other assets, potentially triggering a global economic crisis.”

This notion of intraday financial distress *splashing* onto related assets and markets represents a new channel of systemic risk, facilitated by algorithmic trading. Cliff and Northrop (2012) describe a mechanism by which a flash crash could become a splash crash. The authors conjecture that if the 6th May 2010 Flash Crash had occurred shortly before the close of trading, such that prices had not yet had time to rebound, the global impact would have been significantly higher as investors in other markets (Asian markets first, then European) would have been forced to respond to the losses they had accrued on their US investments.

Empirical Studies

Ben-David et al. (2012) consider *arbitrage* as a channel of shock propagation between assets. In an arbitrage trade, investors exploit relative price changes between financial instruments. Ben-David et al. (2012, p.4) suggest that arbitrage trading of Exchange Traded Funds against the set of underlying assets that they represent propagated distress during the 6th May 2010 Flash Crash: “We hypothesize, and find consistent evidence, that arbitrage between the ETFs and the underlying stocks contributed to propagate the initial shock to the spot market for stocks.” Abrol et al. (2016) include ETF arbitrage among a set of proposed intraday channels by which distress could propagate. Their list also includes transmission due to risk management of the kind discussed in Section 2.4.3 above. This work, however, does not support its observations with either model or empirical research. Jain et al. (2016) perform a purely empirical investigation looking at the impact of the introduction of HFT on the Tokyo stock exchange. The authors conclude that HFT activity increases correlation between stocks, and furthermore suggest that “special attention is required to deal with the adverse impact of [HFT] during tail events” Jain et al. (2016, p.5).

Multi-Asset Microstructure Models

Microstructure models featuring multiple assets are rare in the literature. [Gerig \(2015\)](#) takes a model of informed (fundamentalist) and uninformed (zero-intelligence) traders across multiple assets and finds that stocks with higher simulated HFT activity are more highly correlated. The authors also find that this higher correlation allows pricing errors to propagate more readily. However, the model does not specifically consider network topology, financial stability, systemic risk or flash crashes. [Torii et al. \(2015\)](#) consider a three-asset agent-based model of arbitrage where an index future is traded simultaneously with the set of underlying assets. [Torii et al. \(2015\)](#) find that price shocks can propagate from asset to asset via movements in the index future. The model does not consider the topology of the asset-index network, or flash crashes. However, this is a notable result, given the paucity of microstructure models considering the contagion of financial distress. [Xiong et al. \(2015\)](#) propose a closely-related model consisting of five equity assets connected by a futures index asset. This model is used to assess the effectiveness of price limits on market volatility, but does not consider network topology or flash crashes. [Cespa and Foucault \(2014\)](#) consider a two-asset model for distress propagation where an “illiquidity cycle” emerges as traders avoid changing their investments when price volatility leads to uncertainty and a reduced “risk appetite”. This model does not feature a realistic model of trading, a necessary assumption to provide analytic tractability in this non-ABM work. The authors use the model to test a heterodox trading halt policy that relies on volatility predictions, though the practicality of this policy remains unclear.

Multi-Market Microstructure Models

Microstructure models featuring multiple markets are also rare. [Wah and Wellman \(2016\)](#) consider a single-asset, two-market ABM of *latency arbitrage* (HFT attempting to profit from speed advantages due to technological innovations) though this model does not consider stability. The paper proposes frequent batch auctions as a policy intervention to improve welfare in the model ([Budish et al. \(2015\)](#) make similar recommendations, though in a single-asset single-market context, see Section 2.7). [Schmitt and Westerhoff \(2014\)](#) introduce a two-market, multi-asset ABM with chartists and fundamentalists and again find that “Speculators relying on technical analysis tend to destabilize markets, while speculators using fundamental analysis stabilize them.” ([Schmitt and Westerhoff, 2014](#), p.4). The model is able to reproduce a number of

stylised facts observed in real markets (Section 2.6). The authors do not consider systemic risk or flash crashes. Mo et al. (2013) study a multi-asset, multi-market ABM of *dark pools* (stock exchanges that do not publish data feeds). Although this model is both micro-level, and multi-market, it does not consider stability, systemic risk or flash crashes. The study does, however, make interesting use of real-world market data to drive the simulation in a hybrid back-test/simulation framework. Finally, Van Kervel and Menkveld (2019) develop a single-asset, two-market model with fast and slow traders, and use it to investigate cross-venue trade cancellation practices observed in an empirical dataset. The authors offer evidence that HFT traders switch from liquidity providers to liquidity consumers when asset prices exhibit persistent trends. This is related to the notion of toxic order flow, implicated in the Flash Crash of 6th May 2010 (Easley et al., 2012).

2.5.6 Summary of Microstructure Models of Systemic Risk

In conclusion, we find that micro-level models are generally adept at reproducing empirically observed regularities of financial markets, both at times of crisis (e.g., flash crashes) and non-crisis. However, much further research is required to bridge the gap between microscopic models of trading activity and macroscopic models of systemic risk. The small number of microstructure models featuring either multiple assets, or multiple markets (or both) represent a promising start, but many unanswered questions remain. As stated by Gould et al. (2013, p.1733) “From a regulatory standpoint, it is crucial to understand how electronic trading algorithms affect market stability, yet almost nothing concrete is currently known.”

2.6 Calibration and Validation

As discussed in Section 2.2, calibration and validation of ABMs are critical operations that have proved challenging for the community to codify. However, in order for ABM to deliver on the promise of offering digital laboratories for the *in silico* analysis of policies, regulators need to be assured that the models faithfully capture real-world dynamics. Assessing the extent to which this is the case is a multi-faceted and highly complex problem, and the question of how to effectively integrate real-world data is by no means a new one in the ABM literature. Axtell and Epstein (1994, p.28) provide the following taxonomy:

Level 0: The model is a caricature of reality, as established through the use of simple graphical devices (e.g., allowing visualization of agent motion);

Level 1: The model is in qualitative agreement with empirical macro-structures, as established by plotting, say, distributional properties of the agent population;

Level 2: The model produces quantitative agreement with empirical macro-structures, as established through [...] statistical estimation routines; and finally,

Level 3: The model exhibits quantitative agreement with empirical micro-structures, as determined from cross-sectional and longitudinal analysis of the agent population.

An example of Level 0 is [Schelling \(1971\)](#)’s seminal model of housing segregation. The model is highly abstract, and functions as a cellular automaton ([Langton, 1986](#)). Level 1 models are the most common in the ABM literature. [Paddrik et al. \(2012\)](#) is an example of such a model, and the authors demonstrate qualitative reproduction of a series of stylised facts, which correspond to [Axtell and Epstein \(1994\)](#)’s “distributional properties”. The ABM literature has in recent years progressed to Level 2 models, for example with models of housing markets ([Baptista et al., 2016](#)), and bond markets ([Braun-Munzinger et al., 2018](#)). The availability of sufficiently detailed micro-level data presents significant obstacles to the realisation of a Level 3 model.

In addition to the immense technical challenges posed by calibration and validation — such as the need for data of sufficient quality and quantity, the identification of appropriate methods, and access to sufficient computational resources — it has even proved difficult for the community to arrive at a set of consistent terminology to describe what it is that is being attempted. We adopt the terminology promoted by [Fagiolo et al. \(2019\)](#) and [Gatti et al. \(2018\)](#), which, in brief, is:

- Input validation: asserting that the model’s components, agent behaviours, and initial conditions, are consistent with observed reality.
- Output validation: asserting that the model’s observable outputs are in quantitative agreement with real-world observables.
- Calibration: adjusting model parameters such that model output is brought into the closest possible agreement with reality.

The main terminology difficulties in the literature centre on the distinctions between output validation and calibration. The matter remains unresolved, for example [Iskrev \(2019, p.55\)](#) uses “calibration” to refer to input validation (directly fixing initial model conditions using empirical data), and “estimation” to refer to calibration: “estimating the unrestricted model is almost certain to result in point estimates of the previously fixed parameters that are different from the calibration values”. Similar difficulties are apparent in [Recchioni et al. \(2015, p.2\)](#) who promise “a rigorous calibration procedure designed to validate agent-based models. To this end, we undertake the estimation of the parameters of the agent based model”. [Gatti et al. \(2018\)](#) offer a subtle distinction between calibration and estimation based on the consistency properties of statistical estimators, and thus argue that estimation is a subset of calibration. This latter distinction is not material to this thesis, and so we choose to refrain from using the term “estimation”.

In straightforward terms, calibration allows for the optimal selection of free model parameters, according to a quantitative comparison between model output and reality. (Output) validation allows us to determine whether the resulting output is close enough to be deemed consistent with reality. A key trend in recent ABM literature is the migration from qualitative to quantitative output validation ([Axtell and Epstein, 1994](#)), and the radically enhanced model realism this requires and implies. The rest of this discussion therefore refers to strictly quantitative output validation.

Both calibration and (quantitative) output validation require the definition of a distance metric, i.e., a function of model output and real-world output. Alternative approaches to this distance metric have gained traction in the ABM literature ([Fagiolo et al., 2007](#)). Of particular prominence is the *method of simulated moments* ([Winker et al., 2007](#)). In this approach, ABM and real-world outputs are reduced to vectors each consisting of a set of statistical moment measurements (e.g., mean, skewness, kurtosis). By comparing moments between simulation and reality, it is possible to quantify the model’s ability to exhibit statistical regularities, i.e., stylised facts, observed in the real world. Stylised facts of macro-level models include, for example, economic growth, recession, and the distribution of firm sizes ([Haldane and Turrell, 2019](#)). However, micro-level stylised facts are of more direct relevance to this thesis, and are summarised in Table 2.2.

ABM approaches seek to jointly reproduce as many stylised facts as possible. The ability of models to jointly capture these stylised facts is regarded as a strong indicator of the power of the model ([Franke and Westerhoff, 2012](#)). It is interesting to note that in order to capture a large subset of these stylised facts, models have been found

Paper	Stylised fact	Description
Mandelbrot (1963)	Excess Kurtosis	Price movement distribution has power-law tail
Kirman and Teyssiere (2002)	Aggregational Gaussianity	Price movements measured over increasing timescales tends to a Gaussian distribution
Cont (2001)	Return Independence	Price changes are not autocorrelated
Kirman and Teyssiere (2002)	Volatility Clustering	Absolute price changes have slowly decaying autocorrelation
Bouchaud et al. (2002)	Mean Relative Depth	Volume of shares at increasing depth away from best prices has a characteristic profile
Zovko et al. (2002)	Relative Price	Power-law distribution of order placement prices

Table 2.2: Statistical regularities (stylised facts) of financial markets.

to require significant heterogeneity. For example, Paddrik et al. (2012)’s model of the Flash Crash of 6th May 2010 features six classes of agent, however it still fails to reproduce the Relative Price stylised fact (Zovko et al., 2002). As an alternative to statistical moments, the GSL divergence (Lamperti, 2018) uses Shannon entropy to compare two alphabetised time series, and Barde and van der Hoog (2017) propose to use the Markov Information Criterion to characterise ABM output.

The process of calibration is essentially a search problem, where the goal is to identify the optimal location in a *parameter space* (Blum and Roli, 2003). Such numerical optimisation has a long history that is entirely independent of ABM (for example, Newton’s iterative minimisation method), however, inherent properties of ABMs make the selection of appropriate calibration methods a challenging problem (Section 2.6.1).

The process of output validation requires researchers to define a *threshold* distance between simulated and real-world observables, that allows for the classification of a model as either consistent, or not consistent, with reality. Franke and Westerhoff (2012) offer a principled approach to selection of this threshold by defining the *Moment Coverage Ratio* (MCR). The MCR quantifies the model’s ability to generate a measurable output that is statistically similar to empirical observations. This is further extended in the joint moment coverage ratio (jMCR) which assesses a model’s ability to produce a set of outputs that are simultaneously consistent with their empirical counterparts (e.g., a set of stylised facts). Braun-Munzinger et al. (2018) and Kårvik et al. (2018) successfully deploy the jMCR to validate their models against

intraday market data in the bond and foreign exchange markets, respectively.

2.6.1 Difficulties of Calibration and Validation

Calibration and validation are inherently challenging for ABMs due to key properties of the models themselves — emergence, stochasticity, the presence of tipping points, and computational expense. As stated by [Kukacka and Barunik \(2017, p.23\)](#) a difficulty is posed by the “roughness of the surface of the objective function, which further embarrasses standard methods of optimisation searches”. Not only is the calibration objective function likely to exhibit discontinuities and local minima, it is also difficult to even measure, due to the stochastic nature of ABMs and the need to perform repeated Monte Carlo sampling. Computational costs can thus become highly material. This expense has led to practitioners seeking more efficient methods of searching parameter spaces for locations that produce desirable model outcomes. The application of Machine Learning techniques such as surrogate models to ABM validation ([Lamperti et al., 2018](#)) represents an interesting current research direction ([Fagiolo and Roventini, 2017](#)). [Gratiet et al. \(2016, p.32\)](#) offers the following summary of the surrogate modelling process: “Classical estimation methods based on Monte Carlo simulation are computationally expensive . . . The required costs, in the order of $10^3 - 10^4$ model evaluations, are often not compatible with the advanced simulation models encountered in engineering applications. For this reason, surrogate models may be first built up from a limited number of runs of the computation model (the so-called experimental design), and the sensitivity analysis is then carried out by substituting the surrogate model for the original one”. [Salle and Yıldızoğlu \(2014\)](#) and [Bargigli et al. \(2018\)](#) use *kriging*, or Gaussian Process Regression (GPR, [Bishop, 2006](#)), to model uncertainty in the model output across the parameter space. By interpolating between locations with known utility, this approach allows the efficient selection of subsequent candidate parameter vectors for evaluation in the full (expensive) ABM. A side benefit of this approach is that the resulting surrogate model can also be cheaply queried during *sensitivity analysis* ([Saltelli et al., 2019](#)). However, the discontinuous nature of the objective function also poses difficulties for GPR, and this approach does not scale well to larger numbers of evaluations that may be necessary for complex large-scale models.

A second major difficulty with ABM calibration and validation is a lack of appropriate data. Financial data, in particular, are often highly proprietary, and either expensive or impossible to access ([Fagiolo et al., 2007](#); [Chen et al., 2012](#); [Fagiolo et al., 2019](#)). As we discuss in more detail in the following section, micro-level datasets

available to previous researchers have suffered from at least one of limited date range, limited asset scope, or limited granularity.

Regulators increasingly require banks to report balance sheet data, however as these data are often in aggregate form, researchers must resort to network reconstruction techniques (Cimini et al., 2015; Di Gangi et al., 2018; Newman, 2018). For example Squartini et al. (2016) attempt to reconstruct the asset portfolios of financial institutions using maximum entropy reconstruction. Gualdi et al. (2016) take a similar approach and compare empirical fund–asset networks to stochastically generated networks in order to determine the relative degree of crowding. Anand et al. (2018) provide a comparative study of alternative reconstruction approaches on interbank data, concluding (Anand et al., 2018, p.3) “large interbank networks tend to have a low density, a low diversity, and only a small number of financial institutions in the core.”

Regulatory filings are an important source of information regarding the investments of banks and funds. In the US, institutions with more than \$100mn under management are required to file quarterly *13-F* reports, detailing their long investments in US companies. These data are publicly available (García and Norli, 2012). Guo et al. (2016) provide an interesting analysis of the longitudinal evolution of network metrics based on 13-F data, finding certain network theoretic metrics are correlated with *excess returns* (i.e., financial assets outperforming the market average).

Some datasets on interbank lending have been made available to researchers. For example, Caccioli et al. (2015) and Poledna and Thurner (2016) validate their models against 800 banks in the Austrian interbank network. Serri et al. (2017) validates against a network of 180 Dutch banks. Successful validation results in simulated networks with statistically similar properties to real networks, a prominent example is the distribution of the *degrees* of nodes in the network.

Data on leverage usage by hedge funds has only recently become subject to regulatory disclosure, and the data are not made public⁸. Ang et al. (2011) source this data from a hedge fund that invests exclusively in other hedge funds (a *fund of funds*). Such studies seem to be extremely rare. Bian et al. (2019) access proprietary brokerage data for the Chinese equity market, and others have attempted to indirectly infer leverage by considering other observable properties of funds such as diversification (Jiang, 2018) and market volatility sensitivity (Boguth and Simutin, 2018).

⁸<https://www.sec.gov/about/forms/formpf.pdf>, accessed 2019-03-28.

2.6.2 Calibration and Validation in Practice

The sophistication of calibration and validation procedures has increased steadily since [Kermack and McKendrick \(1927\)](#)'s seminal model of epidemic spreading, which was calibrated using data from the 1905 Bombay plague outbreak, with accelerating progress during the past 10–15 years. Early ABMs exhibited qualitative agreement with empirical reality, and this level of output validation remains the minimum standard for academic ABM publications ([Axtell and Epstein, 1994](#)). However, advances in data accessibility, computational power, and methodology, have fostered ABM calibration and validation as a fertile research area.

Early successes were provided by [Gilli and Winker \(2003\)](#) who use a modified version of the Nelder-Mead simplex (NMS) optimisation algorithm ([Nelder and Mead, 1965](#)) to calibrate a daily-level foreign exchange ABM. The optimisation function incorporates two stylised facts (based on return autocorrelation and kurtosis) and demonstrates convergence to a local minimum, however the resulting model is not validated. [Franke and Westerhoff \(2012\)](#) perform calibration and validation of a daily model of S&P 500 index returns, and, as mentioned above, make an important contribution in the form of the joint moment coverage ratio, which provides a principled way to choose the validation success criterion. [Calvez and Hutzler \(2006\)](#) utilise genetic algorithms ([Goldberg, 2006](#)) to successfully calibrate the parameters of a non-financial ABM. [Fabretti \(2013\)](#) also makes use of genetic algorithms for the calibration of a daily microstructure ABM, and compares the effectiveness of the NMS method. Both methods are found to be effective, but the author cautions about the “long execution times required for computation. Any procedure is so time-consuming that each experiment must be designed carefully before starting” ([Fabretti, 2013](#), p.291). More recently, Bayesian approaches such as Markov chain Monte Carlo have started to gain traction. [Grazzini et al. \(2017\)](#) and [Lux \(2018\)](#) apply a variety of Bayesian approaches to macro-level ABMs, with mixed success. [Lux \(2018\)](#) finds Bayesian approaches are extremely computationally expensive, and do not lead to parameter vectors that give rise to successfully validated model behaviour. [Grazzini et al. \(2017\)](#) are more successful, but their results rest on strong assumptions regarding model ergodicity. [Kukacka and Barunik \(2017\)](#) calibrate a daily micro-level ABM to six market indices including the S&P 500 and FTSE 100 using simulated maximum likelihood. This approach is of interest because the non-closed-form nature of ABMs would otherwise be expected to preclude calibration via maximum likelihood methods. However, the approach suffers from the curse of dimensionality, and attempts to characterise the entire parameter space evenly, with a fixed computational budget.

Comparisons of the effectiveness of different calibration methods would be particularly useful. Existing approaches, while promising, could further benefit from a wider range of methods evaluated (Fabretti, 2013), and from the availability of real-world data (Platt, 2019).

Calibration and validation examples using higher frequency (sub-daily) market data are rare, quite possibly because of data access difficulties. Platt and Gebbie (2018) use a single day of market microstructure data for a single asset on the South African exchange at the one-minute frequency. The results achieve modest success calibrating a micro-level ABM of fundamental and chartist traders using the NMS algorithm. However, the resulting validation is not successful for all empirical moments tested. In their model of the 6th May 2010 Flash Crash, Paddrik et al. (2012) access S&P 500 index data at 1-minute frequency for the month of June 2011. This work does not feature calibration, and the output validation is qualitative. Farmer et al. (2005) access individual transaction level data for 11 London Stock Exchange assets between August 1998 and April 2000, and present a model, featuring impatient and patient traders, that is shown to be in quantitative agreement with a substantial set of stylised facts. Kononovicius and Ruseckas (2019) access data on 26 NYSE assets at one-minute frequency from January 2005 to March 2007 and calibrate a micro-level *herding* model using simulated annealing (Kirkpatrick et al., 1983). Output validation is performed by graphically comparing stylised facts between simulations and empirical data. Finally, Aloud et al. (2017) access transaction level data for 48 currency pair assets from January 2007 to March 2009, and quantitatively validate a model against a suite of micro-level stylised facts. The authors also examine stylised fact sensitivity to model parameters.

In summary, we could not find any evidence of previous microstructure calibration or validation experiments that feature all of (1) fine-grained (intraday) data, (2) a substantial set of assets, and (3) a substantial date range. It is rare to find a study that exhibits more than one of these desiderata. Consequently, we did not find evidence of any prior attempts to calibrate or validate a hybrid micro-macro model (i.e., a large scale microstructure model with a substantial number of assets), though this is unsurprising as such models are themselves under-represented in the literature, notwithstanding the added complexity of model validation.

Gualdi et al. (2015) and Barde and van der Hoog (2017) each argue for the importance of exploring model behaviour over the full range of its feasible parameter space. There are two advantages to this. Firstly, this exploration allows the full characterisation of the model's expressive capability, and secondly, it allows for determination of

the importance and relevance of input parameters. For example, [Gualdi et al. \(2015, p.30\)](#) propose “the very first step in exploring the properties of an Agent-Based model should be, we believe, to identify the different possible phases in parameter space and the location of the phase boundaries.” Both aspects of parameter space exploration relate to sensitivity analysis, which, as a field, arose within the life sciences, where individual experiments can be extremely costly to perform. Under such conditions, principled methods of exploring the effect of experimental parameters demand the careful budgeting of experimental trials. Sampling-based approaches offer such a parsimonious Design of Experiments (DoE) ([Helton and Davis, 2002](#); [Saltelli et al., 2010, 2019](#)). The growing computational costs of ABMs, which typically feature a large number of input parameters, renders combinatorial evaluation of the complete parameter space infeasible. The use of sensitivity analysis within ABM is, therefore, growing in popularity ([Recchioni et al., 2015](#); [Broeke et al., 2016](#); [Barde and van der Hoog, 2017](#); [Llacay and Pepper, 2018](#); [Lamperti et al., 2018](#)). A key benefit of sensitivity analysis is that it offers a defence against claims that ABMs are prone to over-fitting. Sensitivity analyses can reveal whether any model parameters are redundant and do not contribute to output model variance, either on their own, or in combination with other parameters. This level of diagnostic checking offers the possibility of elevating the rigour and power of ABM input validation.

2.6.3 Summary of ABM Calibration and Validation

The discipline of ABM is in the process of replacing the qualitative assessment of model validity with a quantitative approach i.e., progressing from Level 1 to Level 2 in the influential taxonomy of [Axtell and Epstein \(1994\)](#). This progress is being fuelled partially by influences from within the field, but also by a recognition of the relevance of methods from the wider scientific community. While approaches such as the method of simulated moments, stylised facts, and Monte Carlo sampling are by now well-established in the ABM practitioner’s tool-box, recent progress has been made incorporating sensitivity analysis, non-linear optimisation methods, Bayesian reasoning, and machine learning techniques ([Fagiolo et al., 2019](#)). However, consensus as to best practice is yet to emerge, and this situation presents a barrier to the wide-scale adoption of ABM ([Llacay and Pepper, 2018](#)).

Our survey suggests the following potential remedies:

- Adopt consistent terminology for calibration, input validation, and output validation.

- Consistently apply state-of-the-art methodology incorporating sensitivity analysis, stylised fact definitions, and quantitative output validation criteria.
- Establish and strengthen existing involvement with the regulatory community and industry, to ensure wide access to datasets of sufficient scope, fidelity, range of data types, and range of dates.
- Evaluate the relative performance of alternative calibration procedures for complex models and large datasets representative of the current generation of ABMs, including using real-world high frequency market data over a wide range of assets and a wide, and recent, range of dates.

Of direct relevance to this thesis, we do not find evidence of substantial attempts to calibrate or validate hybrid micro-macro models of the financial system. Such models, representing the complex interactions of agents at widely differing scales, may require the synthesis of multiple interrelated datasets, not all of which may be readily accessible. Previous attempts to calibrate and validate intraday ABMs have suffered from a lack of relevant empirical data, in terms of data granularity, or the range of assets and dates considered. There is also limited data on the effectiveness of competing calibration methods for intraday financial market ABMs.

Addressing the aforementioned difficulties will serve to greatly enhance the quantitative rigour of calibration and validation within the field of ABM. This in turn will inspire confidence that model outputs are truly representative of the real world, with quantifiable limitations on model applicability established through sensitivity analysis. Only then will the full potential of ABM be realised as essential utilities in the regulator's tool-box for the *in silico* evaluation of novel policies.

2.7 Regulatory Policy and Interventions

The failure of DSGE models to predict or mitigate the Global Financial Crisis of 2007–2009 gave rise to calls for alternative modelling approaches, from within the regulatory community (Bookstaber, 2012; Haldane and May, 2011; Haldane and Turrell, 2019). The academic ABM community have enthusiastically taken up the challenge (Farmer and Foley, 2009; Levin and Lo, 2015; Bouchaud, 2019). However, partially for reasons discussed in the previous section, and partially due to inertia within mainstream economics (Haldane and Turrell, 2018), ABM is not yet established as an orthodox approach for evaluating regulatory policy.

The rich potential for simulating counterfactual scenarios and novel policies makes ABM an attractive proposition, particularly for crisis scenarios which are, by definition, under-represented in empirical data. Promising macro-level research directions include the improvement of bank stress-testing practice⁹ (Aymanns et al., 2018; Enriques et al., 2019), the evaluation of banking stability measures such as the Basel accords (Poledna et al., 2014; Cheng and Wellman, 2017), and bank capital requirements (Braouezec and Wagalath, 2018).

At the micro-level, the Flash Crash of 6th May 2010 brought the potential for algorithmic trading to give rise to systemic risk to the attention of regulators (UK Government Office for Science, 2012; Ackert, 2012), and regulatory interest remains strong, with the adoption of the MiFID II regulation in Europe¹⁰ which represents a major increase in regulatory scrutiny of algorithmic traders. Danielsson and Zer (2012, p.13) summarise the regulator's concern: "there are however good reasons to believe that under special circumstances high frequency extremes might interact with trading algorithms to generate endogenous feedback loops, resulting in extreme medium frequency outcomes and even on to a systemic event." Microstructure ABMs have been developed to assess the effectiveness of a wide range of intraday policies. A substantial body of research investigates the impact of replacing the standard continuous double auction mechanism with alternative *frequent batch auctions* (Lee et al., 2011; Wah and Wellman, 2016; Brewer et al., 2013; Budish et al., 2015). There appears to be academic consensus that frequent batch auctions reduce the ability of HFT to prey on slower market participants, and there are commercial stock exchanges offering this auction style as an alternative trading venue, as summarised by the European market regulator ESMA (2019)¹¹.

A second major body of micro-level research concerns *price limits* and *circuit breakers* (Sifat and Mohamad, 2019). These strictly intraday regulatory policies are intended to suspend or limit trading in assets during times of distress (at the asset, or wider market level). Price limits and circuit breakers have been controversial since their introduction to the markets following major crashes of the late 1980s, and academic opinion remains divided as to their effectiveness (Subrahmanyam, 2013; Dalko, 2016; Sifat and Mohamad, 2019; Hautsch and Horvath, 2019). Open problems include

⁹Stress tests "assess the resilience of an individual institution, a set of similar institutions (e.g., banks), or the financial sector as a whole by devising a hypothetical (but plausible) stress scenario and determining the losses that would ensue for the relevant component if it materializes." (Enriques et al., 2019, p.24).

¹⁰<https://bit.ly/2G4OCMU>, accessed 2019-10-23.

¹¹<https://bit.ly/2PaECcE>, accessed 2019-10-23.

the appropriate level of market distress at which to trigger a circuit breaker (Lee et al., 2011); whether volatility is reduced following activation or whether the circuit breaker simply postpones distress (Kobayashi and Hashimoto, 2011); and whether the existence of a circuit breaker can increase volatility due to *magnet effects*, as investors pre-emptively sell when a trigger level is approached, accelerating the decline (Subrahmanyam, 2013; Sifat and Mohamad, 2019). Conversely, circuit breakers seem to offer intuitive advantages by providing cooling-off periods that bring the timescale of market dynamics back into a more human regime. Subrahmanyam (2013, p.6) suggests that this benefit could potentially mitigate a systemic crisis: “in sharply falling markets traders may receive margin calls. Without a halt traders may be caught by surprise and may not be able to satisfy their margin calls . . . A halt that collects and matches orders after trading resumes may allow for more time for margin calls to be fulfilled”.

Part of the problem is that circuit breaker activations are rare, indeed the US circuit breaker failed to trigger during the Flash Crash of 2010, resulting in regulators subsequently tightening the threshold levels at which falls in the reference S&P 500 index trigger a market-wide halt¹². The relatively small number of circuit breaker activations limits the relevance of empirical analysis as Subrahmanyam (2013, p.6) argues: “since they [circuit breakers] are triggered so rarely, at the present moment, it is challenging to ascertain their costs or benefits with any degree of reliability. Any discussion of potential benefits has to be based on logic, rather than evidence”. The ABM community has offered substantial contributions to improving this situation via simulated price limits and circuit breakers. The opportunity is summarised by Wellman and Wah (2017, p.106)

“Analysis of available data is ultimately limited, however, with respect to counterfactual questions, such as the response of financial markets to rarely occurring shocks or the effects of alternative market rules and regulations. Answering such questions inherently requires models that incorporate causal premises, specifically, assumptions as to how trading behavior is shaped by environmental conditions.”

Yeh and Yang (2010) test the effectiveness of price limits in a single asset microstructure ABM. The model is not calibrated (according to our terminology), but it is qualitatively validated against a small set of stylised facts. Price limits are shown to effectively reduce price volatility, with the side effect of slowing down adjustment

¹²<https://www.sec.gov/rules/sro/nyse/2019/34-85560.pdf>, accessed 2019-08-26.

of prices to reflect changes in fundamental value. Lee et al. (2011) propose a qualitatively validated ABM with three equity assets and a composite index asset. The authors find that circuit breaker activation following price falls of 20–30% is effective at reducing market volatility. Smaller thresholds are associated with reduced liquidity and hence increased post-halt volatility. Kobayashi and Hashimoto (2011) propose a single-asset ABM that is again only qualitatively validated, and find that circuit breakers stabilise markets leading to fewer bankrupt traders, at the expense of reduced market volume. The authors highlight the need for future studies to fully consider “the micro-macro loop including market institutions.” (Kobayashi and Hashimoto, 2011, p.370). Xiong et al. (2015) present an ABM notable for the relatively large number of microstructurally simulated assets (five, in addition to an index asset). Although this set of assets falls well short of representing the complex network of interconnected assets at real-world scale, the model extends the state of the art in micro-macro financial ABM simulation scope. The price limits in the model faithfully reflect their empirical counterparts in the Chinese futures market, however other aspects of the model are not calibrated or validated, even qualitatively. The authors find that price limits in an intermediate regime offer the optimal balance of trade volume and volatility. Finally, Jacob Leal and Napoletano (2019) compare what they term *ex ante* (price limits that prevent extreme price falls) with *ex post* (circuit breakers triggered as a result of falls) in a single-asset, qualitatively-validated ABM. Flash crashes in this model are effectively prevented by price limits, and the authors conclude that price limits are more effective regulatory policies than circuit breakers. This conclusion, however, reveals a problem. In reality, circuit breakers trigger in response to distress at the level of *markets*, whereas price limits control volatility at the level of *assets* (Sifat and Mohamad, 2019). Indeed, referring to the current US *Limit-Up-Limit-Down* (LULD) single stock regulation, BlackRock (2015, p.10) demonstrate the deficiency of a pure *ex ante* price limit approach: “On August 24 [2015], market mechanisms were not working properly. Securities experienced sharp price dislocations and nearly 1,300 LULD halts were triggered as the market lacked transparency and clarity, demonstrating a limitation of the LULD mechanism during market-wide events.”

What is lacking in the studies presented here is a *systemic* view of circuit breaker activation. Given market-wide circuit breakers (1) trigger in response to falls in an index composed of large numbers of assets (500 assets in the case of the US financial system¹³), (2) impact the entire market simultaneously by halting trading in all

¹³<https://www.nyse.com/markets/nyse/trading-info>, accessed 2019-10-06.

equity, future, options and fixed income markets, and (3) are intended to prevent systemic collapse of kind described by [Subrahmanyam \(2013\)](#), we find an almost total absence of models that connect the micro- and macro- market levels that could facilitate investigation of the systemic effects of circuit breaker activation. With such a micro-macro model, that considers the complex network of interrelated financial assets, regulators may be able to gain a deeper understanding of the interplay between single-asset price limits and market-wide circuit breakers. If such a model were quantitatively calibrated and validated against empirical microstructure data, regulators could go further, and potentially quantify the effectiveness of these regulatory policies that are deeply rooted in the micro-level behaviour of market participants, yet trigger responses at the level of entire markets, with global visibility.

2.8 The Need for a New Micro-Macro Modelling Approach

We have considered the existing bodies of literature covering both macroscopic and microscopic financial agent-based models in the context of financial stability. We have found that models from both sets of approaches perform well in reproducing statistical properties associated with their modelling domain. Macroscopic interbank network models utilising systemic risk metrics perform well in identifying *systemically important financial institutions* (SIFIs) during the 2007–2009 crisis ([Battiston et al., 2012b](#)), and microstructure models perform well in reproducing extreme price impact events such as flash crashes ([Paddrik et al., 2012](#)).

We have identified an important limitation of interbank models in the form of unrealistic linear price impact functions, whilst microstructure models have demonstrated that crisis price impact can be far from linear. Conversely we have identified a lack of microstructure models that consider more than one financial asset or market. We join [Hu et al. \(2015, p.4\)](#) in highlighting the advantages of “combining econometric-based and topology-based approaches”.

We argue, therefore, that in the context of financial stability modelling, both literatures have something to offer the other. By providing a realistic crisis price impact model, microstructure models could enhance the accuracy and realism of macroscopic network models of shared-asset distress. Conversely, by connecting assets together in realistic fund ownership networks, macroscopic network models could facilitate the study of realistic financial distress contagion at short term, intraday

timescales thus providing alternative methods to analyse flash crash contagion. Figure 2.3 illustrates the proposed hybrid modelling framework.

We claim that such compound models are inadequately represented in the literature, and that significant further research is justified to work towards addressing this shortcoming.

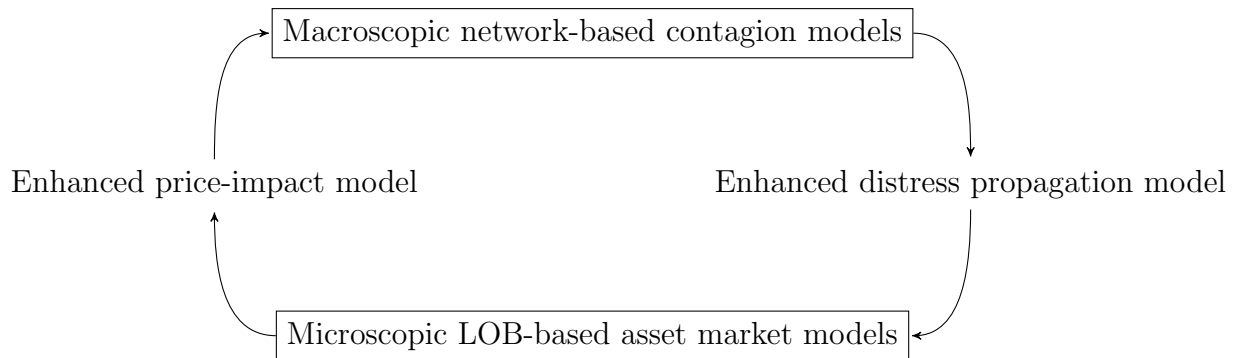


Figure 2.3: An illustration of our argument that (1) macro models can improve micro models by providing the machinery of financial distress contagion, and (2) micro models can improve macro models by providing realistic models of crisis price impact.

2.9 Summary of Previous Research

Our survey of existing literature has revealed that ABM is making significant progress towards a level of quantitative realism that will facilitate the adoption of ABM as an orthodox methodology by mainstream financial regulators and practitioners. Some of this progress is due to advances in the availability of suitably detailed datasets, and increasing computer power, and these trends are expected to continue. While there has also been great progress in advancing ABM methodology, we have identified key weaknesses in current approaches:

1. Existing macro-level systemic risk models rest on unrealistic assumptions regarding price impact, and ignore details of liquidity provision and order placement.
2. Existing micro-level market models ignore the network of interconnected assets and market participants, and so are unsuitable for the investigation of multi-asset, systemic, phenomena.

3. This lack of comprehensive micro-macro modelling is also apparent in previous research into intraday policies such as price limits, and circuit breakers, where the multi-asset nature of circuit breaker activation is ignored, or simulated with a small number of assets that are not representative of real-world complex financial networks.
4. Previous attempts to calibrate micro-level ABMs are limited in terms of (often more than one of) data granularity, range of assets, or range of dates. We do not find evidence of previous microstructure calibration experiments at a scope or scale representative of the real-world.

We thus establish the need for a new class of hybrid micro-macro models, calibrated at the scale of entire markets, and capable of faithfully representing the complex real-world network of interconnected funds and assets. With the inclusion of realistic regulatory mechanisms such as market-wide circuit breakers, such models would allow advanced in-depth analyses of intraday systemic risk and flash crash contagion.

“In the future, we must clarify the market mechanism of interactions between institutions and market participants by dealing with the micro-macro loop including market institutions.”

— Kobayashi and Hashimoto (2011, p.370)

3

A New Approach to Simulating Systemic Risk and Flash Crash Contagion

IN THE PREVIOUS CHAPTER we surveyed the (macroscopic) systemic risk literature, and contrasted methodological approaches with those in the (microscopic) flash crash and price impact literatures. Our conclusion was that very few attempts have been made to bridge the gap between the micro and the macro — to understand how micro-level algorithmic trading gives rise to macro-level systemic risk, and how macro-level network structures affect the propagation of micro-level flash crash distress.

Our objective in the current chapter is to develop a micro-macro hybrid approach to financial market modelling, that seeks to address the aforementioned limitations of previous research. To reiterate, the question that underpins our research is:

What role do algorithmic trading systems have in the contagion of distress through a network of financial institutions, connected via shared-asset investments?

In order to begin to answer this question, we propose a novel synthesis of established approaches from the macro and micro ABM literatures. With such a model in hand, we are able to investigate the interplay between micro-level agent behaviours and macro-level network properties, in the context of systemic risk and flash crash contagion.

This chapter proceeds as follows. We briefly discuss our epistemological position in Section 3.1. In Section 3.2, we present a high-level overview of the micro-macro hybrid approach, including a detailed discussion of the modelling assumptions underpinning the different system components. In Section 3.3, we describe in detail our agent-based methodology, including the behaviours implemented by agents in the model, and constraints by which agents are bound. Section 3.4 provides formal definitions of the metrics that we use to quantify systemic risk emerging from simulated instantiations of the ABM. Sections 3.5 and 3.6 complete our methodological presentation with details regarding our simulation protocol, and a summary of notation used in this chapter. We empirically validate the model in Section 3.7. In Section 3.8, we present results demonstrating the effect of distressed selling on asset prices, and identify the circumstances under which the model exhibits flash crash phenomena. Results concerning the impact of leverage, capitalisation, and margin tolerance on systemic risk are presented in Section 3.9, and the effect of fund–asset network topology on systemic risk and flash crash contagion is investigated in Section 3.10. We conclude in Section 3.11.

3.1 A Note on Epistemology in this Thesis

As discussed in Chapter 2, the stochastic and emergent nature of ABMs renders closed-form analysis infeasible. It is instead necessary to perform repeated Monte Carlo sampling of a given ABM scenario in order to assess the distribution of possible model outcomes. In the analyses that follow in this, and subsequent, chapters, we thus make use of statistical methods in order to draw inferences regarding model behaviour. Our approach is not, therefore, to formulate and prove theorems. Instead, our epistemological position is to accept as knowledge the outcome of statistical hypothesis tests.

Unless otherwise stated, in all results we present the mean, with 95% confidence interval, of the distribution of Monte Carlo outcomes. The interval $[a, b]$ represents the *standard error in the mean*, and is given, for a Monte Carlo distribution of n observations, with mean μ and standard deviation σ , by

$$[a, b] = \mu \pm t_{0.975, \infty} \frac{\sigma}{\sqrt{n}}, \quad (3.1)$$

where $t_{0.975, \infty} = 1.96$ is the asymptotic critical value of Student's t -distribution for a two-sided hypothesis test at $p < 0.05$ (Boas, 1996).

3.2 Micro-Macro Approach

Our model features the synthesis of a macroscopic fund–asset network with a microscopic limit order book (LOB) trading simulation. When funds in the model adjust their asset holdings, they do so by submitting orders to a realistically input-validated (Section 2.6) stock market model, where heterogeneous trading agents interact with the fund’s orders. Price impact arises endogenously from our simulation, and resulting asset price movements affect the profits and losses of the funds. In order to represent the composite micro-macro structure it is necessary to introduce a set of heterogeneous interacting entities:

1. Banks: to model the provision of leveraged financing to institutions.
2. Funds: to model institutions holding leveraged positions in risky assets.
3. Assets: to model liquidity provision via the limit order book.
4. Traders: to model realistic asset market activity and response to market distress.

A summary of the agents and other entities in the model is presented in Figure 3.1. Dashed arrows indicate secured margin loans extended by the single bank to the fund agents (for the sake of clarity, only a small number of funds and assets are represented). Solid arrows represent fund holdings in assets, and this component of the system is what we refer to throughout this thesis as the “bipartite fund–asset network”. Finally, dotted arrows indicate the order placement activity of non-leveraged trader agents.

In the sections that follow, we will discuss how we capture these entities in our model, but we first present the key assumptions underpinning the approach.

3.2.1 Asset-related Assumptions

Asset microstructure is represented using limit order books. Transactions occur when orders to buy (sell) match against previously submitted orders to sell (buy) (Gould et al., 2013). Our asset model is supported by the following key assumptions.

We assume:

- A single currency for all financial transactions in the model.
- That trading is costless in the sense that there are no commissions, exchange fees or rebates involved in the model. Such costs are likely to be immaterial in comparison to the extreme price impact encountered during flash crash type phenomena.

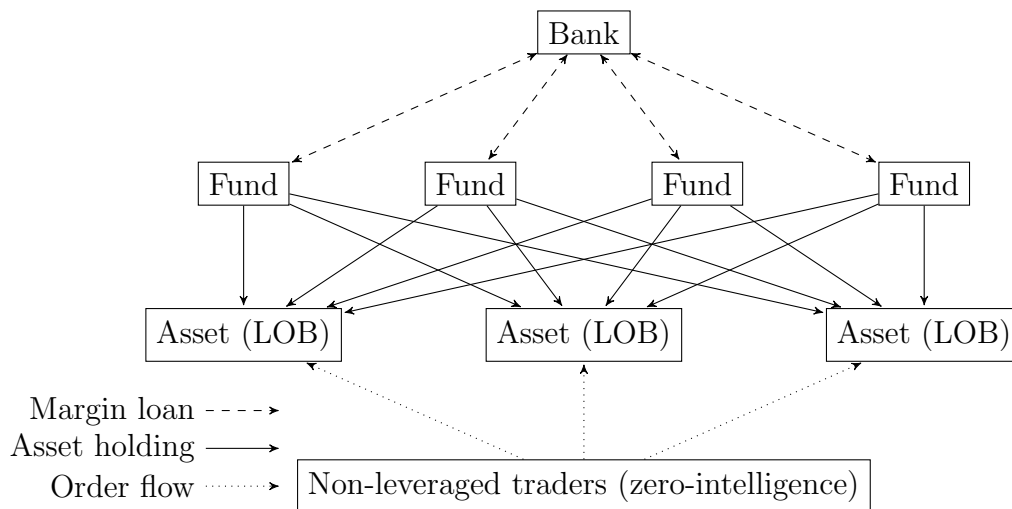


Figure 3.1: Relationships between entities in the model. Arrows represent balance-sheet relationships. Between banks and funds, these relationships refer to secured margin loans and leverage. Between funds and assets, the relationships refer to asset ownership within fund portfolios. Similarly, between traders and assets the relationship again refers to asset ownership in trader portfolios, though in the present model we assume that traders are non-leveraged and so not subject to margin calls. Note that we do not model banks directly investing in assets.

- That assets have no corporate actions such as splits or dividends (a minor assumption given the timescale of corporate actions is weeks rather than intraday).
- That assets have the same liquidity and trading characteristics, as well as holding the same start price at the beginning of simulation runs. The liquidity assumption can be compared to restricting our asset universe to members of a major index such as the Eurostoxx 50 or S&P 500.
- That assets are not directly linked via derivative contracts such as options or via direct asset–asset ownership. Such links constitute additional potential channels for distress propagation, and are outside the scope of this chapter.
- That all trading takes place at a single trading venue. Market participant confusion over conflicting data from alternative trading venues, as well as venue fragmentation, was implicated in the Flash Crash of 6th May 2010 (CFTC and SEC, 2010; Madhavan, 2012). However, flash crash liquidity and price impact dynamics arise in the present model without needing to simulate at this level of detail.

3.2.2 Fund-related Assumptions

Our research questions relate to how the fund–asset network impacts systemic risk (Sections 3.9 and 3.10), with a particular focus on the impact of alternative network topologies of overlapping asset portfolios. We construct networks with specific properties in order to simulate alternative economic scenarios. The investments made by funds are, therefore, not validated against real-world data. Such data is available via regulatory filings and the model can, in principle, be validated against these data to further enhance the model’s applicability to policy-making (Chapter 4).

We characterise the funds in the model as large institutions that take a long time (many days at least) to accumulate positions, and so over the short intraday timescales we wish to consider, such as the duration of a flash crash (minutes), their pre-crisis position-building behaviour is not relevant to our research questions. However, if a fund enters distress via the enactment of a margin call, it becomes active and seeks to rapidly reduce leverage.

We assume:

- That funds already hold a set of asset positions before the simulation begins and do not adjust positions unless experiencing financial distress.
- That funds do not invest directly in other funds (so-called *fund-of-funds*). As with asset-asset direct linkage, such connections would constitute an additional contagion channel that is beyond the scope of this chapter.
- That funds deploy all of their capital as collateral for leverage loans and so have no cash reserves available in order to satisfy margin calls. This implies that funds must liquidate (reduce in absolute terms) asset positions in order to reduce leverage. Capital injection is equivalent to increased margin tolerance, which we investigate comprehensively in this chapter.
- That funds liquidate aggressively (quickly, by taking liquidity), similarly to the empirical observations by [Adrian and Shin \(2010\)](#) and the model by [Greenwood et al. \(2015\)](#).
- That when funds liquidate, they liquidate uniformly across their assets. This approach follows the methodology established by [Caccioli et al. \(2014\)](#).
- That all funds in the simulation have equal starting capital (also as in [Caccioli et al., 2014](#)).
- That, in line with established methodology (see Table 2.1), asset positions are long-only (positive). This strong assumption means that our present work fits into the fire-sales literature ([Shleifer and Vishny, 2011](#); [Greenwood et al., 2015](#))

and does not attempt to explain long-short phenomena such as the 2007 “Quant Quake” (Khandani and Lo, 2007)¹.

3.2.3 Bank-related Assumptions

In our model, financial distress propagates between funds connected via shared asset holdings. It is not, therefore, necessary to consider interbank contagion in order to address our research questions. Neither is it necessary for us to represent banks as directly owning assets; this behaviour is encapsulated entirely within funds and so provides a clear separation of responsibilities without loss of generality. Our model thus incorporates a single bank that provides leverage loans to all funds in the simulation.

We assume:

- That all funds in the simulation are offered equal leverage terms, following established methodology (Caccioli et al., 2014).
- That fund portfolio constitution is ignored when banks set leverage limits. In reality, different assets attract different levels of permissible leverage (Lillo and Pirino, 2015), and more generally leverage provision varies as a function of banks’ balance sheets (Adrian and Shin, 2013). Increased portfolio volatility reduces asset collateral value, and so represents an additional channel of distress propagation. Our abstraction follows the methodology of related approaches (Huang et al., 2013; Caccioli et al., 2014; Cont and Schaanning, 2017).
- That banks monitor fund portfolios on an intraday basis and issue margin calls to funds whose current leverage breaches some threshold. This assumption is realistic (Brunnermeier and Pedersen, 2009; Cont and Schaanning, 2017).
- That if a fund defaults (trading losses exceed capital) the bank will take over and liquidate the fund’s portfolio. Margin terms typically provide for this re-possession.
- That there are no other funding costs or interest charged by banks to funds. We justify this on the basis of timescale — such fees are, in reality, accounted on a daily or monthly basis and are not material to the short intraday timescales relevant to our research questions.

¹Khandani and Lo (2007) and Cont and Wagalath (2013) argue that crowding at the trading strategy (or *risk factor*, Fama and French, 1993) level describes the financial contagion observed during the “Quant Quake” more clearly than consideration of individual investments in single financial securities.

3.3 Agent-Based Model

In order to simulate the emergent properties of the interconnected, heterogeneous system of banks, funds, assets and traders, we implement an agent-based model. The ABM consists of software agents representing banks, funds, and traders. The model further incorporates assets, and the markets for the trading of those assets, but these are not considered agents in their own right. Our notion of which entities in the model constitute agents follows Wooldridge (2001, p.21):

“An *agent* is a computer system that is *situated* in some *environment*, and that is capable of *autonomous action* in this environment in order to meet its delegated objectives.”

In the sections that follow, we describe the behaviour of, and interconnections between, agents in the model.

3.3.1 Banks

As discussed above, we model the activity of a single bank whose primary purpose is to provide leveraged financing to funds. The bank is modelled by considering its balance sheet of assets and liabilities, in line with existing approaches (Gai and Kapadia, 2010). The balance sheet for the bank in our model is presented in Table 3.1.

Assets	Liabilities
Margin lending	Loss on margin loans
Cash	Equity

Table 3.1: Bank balance sheet model. Equity is defined as the sum of assets minus the sum of liabilities.

As shown in Table 3.1 and discussed previously, we do not model interbank lending in this chapter. The function of the bank in our model, therefore, is to issue debt to funds. This debt is secured using the fund’s initial capital $C^0 \in \mathbb{R}_+$ as collateral. Banks offer funds identical initial leverage $\lambda_i^0 = \lambda^0, \forall i \in (1, \dots, n_f)$, $\lambda^0 \in \mathbb{R}_+$. The total value L_i of a loan made by the bank to the i th fund is therefore given by $L_i = C_i^0 \lambda_i^0$. This expression demonstrates an important convention regarding leverage: the total cash available to funds for the purposes of investment, V , is given by

$$V_i = L_i + C_i^0 = C_i^0(\lambda_i^0 + 1), \quad (3.2)$$

where leverage of zero ($\lambda^0 = 0$) corresponds to the case where $L_i = 0$, that is, no cash is loaned.

As explained above, we assume that each of the n_f funds utilise all of their available capital for the purposes of funding asset positions. If we represent the investment of fund i in asset j as an element of a matrix A_{ij} then we may write

$$C_i^0(\lambda_i^0 + 1) = \sum_{j=1}^{n_a} A_{ij}. \quad (3.3)$$

If we further represent asset positions as the product of asset price at time t , $p_j^t \in \mathbb{R}_+$, and position size in shares, $S_{ij}^t \in \mathbb{R}$, then for time $t = 0$ we can write

$$C_i^0(\lambda_i^0 + 1) = \sum_{j=1}^{n_a} S_{ij}^0 p_j^0, \quad (3.4)$$

which makes explicit the relationship between leverage, capital, price and fund asset holdings. As prices fluctuate, the total fund portfolio value will also of course vary. This price variation will result in profit and loss accruing to the fund at different times (this approach is known as *mark-to-market* valuation). This profit and loss is combined with the fund's capital, yielding an updated capital at time t , C_i^t . We can thus derive the leverage for fund i at time t , λ_i^t , which is given by

$$\lambda_i^t = \frac{1}{C_i^t} \sum_{j=1}^{n_a} S_{ij}^t p_j^t - 1. \quad (3.5)$$

As in [Cont and Schaanning \(2017\)](#), we assume that banks continuously measure the ratio $\tau = \lambda^t/\lambda^0$ for all funds, and issue margin calls when this ratio exceeds a critical hysteresis threshold $\tau_c \in [1, \infty)$ (where the infinite limit implies banks never issue margin calls, and the lower limit of 1 implies banks issue margin calls for infinitesimally small fund losses). The margin call state remains in force at a fund until either the fund reduces its leverage λ_i^t such that $\lambda_i^t < \lambda_i^0$, or the fund defaults (which occurs when $C_i^t < 0$). Note that in our model the effect of a bank issuing a margin call to a fund is exactly equivalent to a fund managing its leverage to a target – a common industry practice ([Cont and Schaanning, 2017](#)). If a fund defaults, we

invoke the bank's realistic prerogative to seize its assets and to continue to liquidate the positions.

Banks act autonomously in their decisions as to whether and when to issue margin calls. Although we do not explore bank decision-making in the present work, the possibility justifies our classification of banks as agents in the model.

3.3.2 Funds

The primary purpose of funds in the model is to accept leverage loans from banks and to use these loans to finance a portfolio of assets (see Section 3.3.1). The model incorporates many funds, each with its own randomly allocated portfolio (details of the allocation process are provided below). As asset prices vary as a result of the microstructure simulation described in the next section, so the leverage of funds also varies as profit and loss accrues to the fund's account. If leverage increases beyond a critical threshold (see Section 3.3.1 for details) the fund will enter a margin call state and will aggressively liquidate its portfolio (details of the liquidation mechanism and how this relates to the microstructure model are presented in Section 3.3.5 below). Funds can, in principle, decide autonomously which assets they elect to liquidate. Indeed, [Capponi and Larsson \(2015\)](#) find that the strategic selection of asset liquidation behaviour has a large impact on systemic risk in a macroscopic model. We consider this strategic behaviour in Chapter 5. Further fund autonomy manifests in the selection of distressed liquidation parameters, and we explore systemic sensitivity to this in Section 3.8.

Funds are represented by their balance sheet, as shown in Table 3.2. As described previously, we do not simulate the acquisition of fund asset portfolios, rather we specify portfolios as part of the initial conditions of simulation runs. This approach allows us to explore portfolios with specific characteristics and their effect on systemic risk.

Assets	Liabilities
Capital	Margin loan
Portfolio of assets	Portfolio profit/loss
	Equity

Table 3.2: Fund balance sheet model.

3.3.3 Fund–Asset Bipartite Network

Following the approaches of Caccioli et al. (2014) and Huang et al. (2013) we define the random bipartite graph (Newman, 2003) $\mathcal{G} = (X \cup Y, E)$ in which edges E represent investments made by funds (bottom nodes, X) in assets (top nodes, Y) such that $|X| = n_f$ and $|Y| = n_a$. The structure of $\mathcal{G}$ can be represented by the matrix $\mathbf{A} \in \mathbb{R}^{n_f \times n_a}$, defined in Section 3.3.1. A positive element A_{ij} corresponds to the presence of an edge $E \in \mathcal{G}$ between fund i and asset j . $\mathbf{A}$ is thus the *adjacency matrix* representation of $\mathcal{G}$ (Newman, 2003). The size of investment, in units of US dollars, is represented by the edge *weight*. In the present work all investments are positive, as discussed in section 3.2.2.

We introduce two parameters to control the topology of $\mathcal{G}$. The fraction of assets in which a fund invests (the *diversification*, or *density*, Caccioli et al., 2014) is controlled by parameter $\rho \in [0, 1]$. Funds are restricted such that they hold investments in at least one asset, that is, un-weighted fund degree $k_{fund} = \max(\lfloor \rho n_a \rfloor, 1)$. Since the same value of ρ is used for all funds, as $\rho \rightarrow 1$, all funds converge to the same portfolio (such that every fund invests in every asset). Conversely, $\rho \rightarrow 0$ means each fund invests in a single asset.

In order to generate networks with a controllable level of overlap, we introduce the notion of preferential attachment (Barabási and Albert, 1999) governed by a parameter $\beta \in \mathbb{R}$ controlling the non-linear strength of preferential attachment (Krapivsky et al., 2000). Starting with the set of assets, network construction proceeds by adding funds sequentially. Each additional fund makes weighted connections to k_{fund} different assets. The probability that the fund connects with a given asset is related to the total already invested in that asset, modulated by the crowding parameter β . $\beta = 0$ generates a fully stochastic Erdős-Rényi random graph (Newman, 2003). $\beta = 1$ generates the linear Barabási-Albert model, and $\beta > 1$ generates a network in which funds have a very high probability of selecting assets that are already held by other funds. This can leave some assets disconnected from the network in our model, and we account for this in the results that follow. Our model also supports $\beta < 0$, which generates networks in which funds prefer to invest in assets with the smallest existing investment. We refer to such network configurations as *dispersed*. Where $n_f \leq n_a$ it is possible for each fund to invest in a different asset producing a maximally disjoint network with no possibility of contagion between assets. Dispersed network topologies give rise to portfolios that satisfy an “orthogonality” condition (Cont and Wagalath, 2013) whereby “distressed selling of investors in a large fund may also exacerbate the

volatility of funds with similar allocations, while leaving funds verifying an ‘orthogonality’ condition unaffected” (Cont and Wagalath, 2013, p.737). Chen et al. (2014) find that such a network is the most stable configuration in high-leverage scenarios and we seek to test this assertion in our micro-macro model.

Unlike Caccioli et al. (2014), we do not assume that funds make equal investments to each asset in their portfolio. Instead, for each fund we sample k_{fund} random variates from a LogNormal distribution $\text{Lognormal}(0, 0.66^2)$, normalising such that the variates sum to unity². Edges are added between the fund and k_{fund} assets with edge weights assigned in decreasing size order from the set of normalised variates. The implication of size ordering is that when $\beta > 0$, funds are likely to allocate their largest investment to the same asset. We explore the effect of alternative asset allocation distributions in section 3.10.4. Algorithm 1 in Appendix B presents the network construction method detail.

Figure 3.2 demonstrates the range of topologies that can be created using our bipartite preferential attachment method (for clarity we present small networks in which $n_f = n_a = 4$). Panels on the same row have equal diversification parameter ρ , and panels in the same column have equal crowding parameter β . Figure 3.2(a) shows a low density, dispersed configuration in which each fund holds a different asset. This network is disconnected, prohibiting default cascades. As crowding parameter β increases in Figure 3.2(b), the network becomes somewhat more connected with two funds sharing a common asset. The effect of preferential attachment becomes apparent in Figure 3.2(c) where all funds invest in the same single asset, leaving most assets disconnected from the network. Figures 3.2(d–f) illustrate higher diversification networks with $\rho = 0.5$. In each case the network consists of a single connected component (disconnected assets are not considered to be network components as they are effectively irrelevant to our simulation). It is possible in connected networks for distress at any fund to affect (directly, or indirectly) all other funds. Finally, Figures 3.2(g–i) demonstrate that as $\rho \rightarrow 1$ networks become *complete* (Newman, 2003). For complete networks it may appear that crowding parameter β becomes degenerate, however, for weighted networks this is not necessarily the case. Figures 3.2(h) and 3.2(i) demonstrate cases where some nodes receive higher investment than others, even though all nodes have the same (unweighted) degree.

²We fix LogNormal mean at 0 and LogNormal standard deviation at 0.66, which generates moderately skewed positive variates.

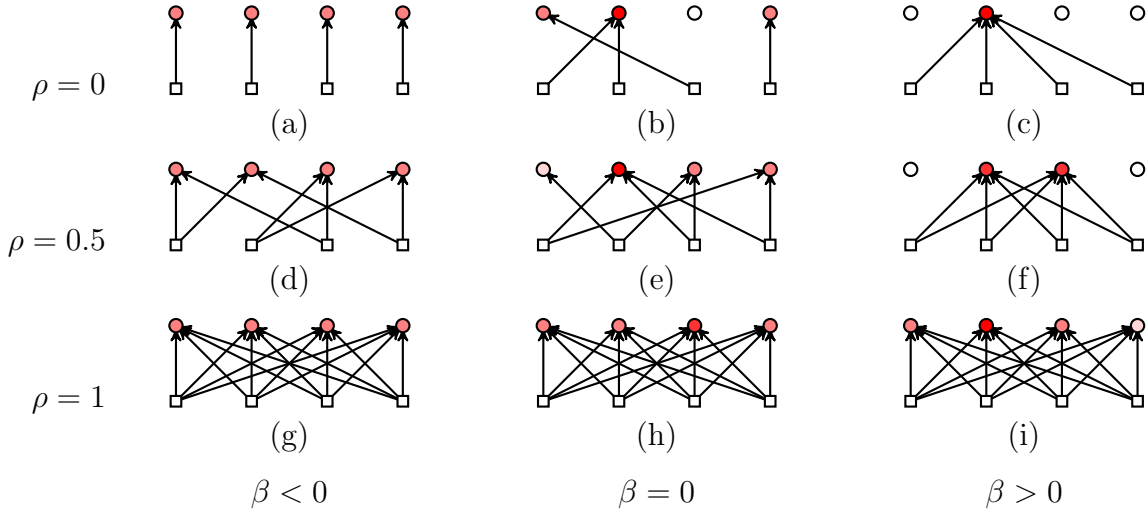


Figure 3.2: Illustrative examples of the range of bipartite topologies that can be generated with our network construction method. Top nodes (circles) represent assets, bottom nodes (squares) represent funds. An edge between two nodes represents asset ownership by a fund. Panels on the same row have equal diversification parameter ρ , and panels in the same column have equal crowding parameter β . Darker asset node shading indicates a higher proportional amount invested in the asset.

3.3.4 Traders

Trader agents in our model operate in a single-market, single-asset capacity in line with most of the ABM microstructure literature. That is, a given trader agent is active in a single asset, even though it may be controlled by an entity with interests in multiple assets. Such considerations are of peripheral importance in the present model since funds do not take an active trading role. Trader agents, furthermore, are not considered to be leveraged, and so will not be subject to margin calls.

Following well-established principles in the agent-based computational finance literature, we minimise behavioural assumptions by endowing trader agents with minimal intelligence. This approach is known as zero-intelligence (ZI), and is characterised by the lack of forecast capability or adaptation in agents (Gode and Sunder, 1993; Farmer et al., 2005). Despite their apparent simplicity, such models are remarkably successful at reproducing stylised facts of financial markets (Fagiolo et al., 2007; Panayi et al., 2012). To be more specific, we reproduce the model presented in Padrik et al. (2012) which features a heterogeneous set of trader agents, each of which is in turn based upon behavioural observations of real world trading behaviour on the day of the 6th May 2010 Flash Crash presented by Kirilenko et al. (2017).

Other models were also considered as a basis for the microscopic LOB simulation

(Vuorenmaa and Wang, 2014; Hanson, 2011; Jacob Leal and Napoletano, 2019). All of these models feature a combination of low and high-frequency trading and so are broadly similar, but the model presented in Paddrik et al. (2012) had the advantage of removing any strategic choice on the part of agents and by doing so remaining closer to the zero-intelligence agent principle. Paddrik et al. (2012) has the further advantage of featuring a realistic order matching mechanism which is closer to real-world stock exchanges than any of the other models mentioned above. A final and notable feature of Paddrik et al. (2012)’s model is that the microstructure simulation is calibrated to real-world wall-clock timings. The agent market entry schedules and resulting asset price dynamics thus operate over realistic timescales, with flash crashes occurring over a period of seconds to minutes (the model validation is presented in full in Section 3.7). This stands in contrast to more familiar event-time models (Gode and Sunder, 1993; Chiarella et al., 2002; Hanson, 2011; Jacob Leal and Napoletano, 2019), and allows us to deeply investigate the intraday consequences of a deleveraging event. All of the results in this thesis are generated with a model using an update time increment (δt) of 50 milliseconds. All agents, including banks and funds, are updated at this same high level of resolution, though our results are robust in the case where banks and funds operate 100 times slower on a 5-second update schedule.

As in Paddrik et al. (2012), we assume agents arrive at the market according to a Poisson process with a characteristic mean interaction interval specified per agent type (see Table 3.3). This approach is standard for micro-market models (Farmer et al., 2005).

Type	Timescale (s)	Inventory (shares)	Population	Inputs
Small	7200	$(-\infty, \infty)$	6500	–
Fundamental B/S	60	$(-\infty, \infty)$	2500	Toxic order flow
Opportunistic	120	$[-120, 120]$	1600	Stochastic signal
Market Maker	20	$[-120, 120]$	320	Toxic order flow
HFT	0.35	$[-120, 120]$	16	LOB imbalance
Distressed	10	$(-\infty, \infty)$	$(1, \dots, n_f)$	Market volume

Table 3.3: Trader agents in our model are based on Paddrik et al. (2012)’s model, which was validated against real market data presented by Kirilenko et al. (2017).

We summarise the main features of the trader agents in Table 3.3. Detailed descriptions of the trader agent types are provided in Appendix A, and are closely related to the implementations described in (Paddrik et al., 2012).

The connection between the behaviour of the distressed selling agent and the liquidation behaviour of a controlling fund agent is discussed in the next section.

3.3.5 Distressed Sellers

A single aggressive seller executing a large set of trades was identified as a key causative factor in the 6th May 2010 Flash Crash ([CFTC and SEC, 2010](#)). Distressed seller agents in our model place market orders (orders with quantity but no price limits) with an order size limited to a fixed percentage (η) of the previous one-minute trading volume. This facilitates a positive feedback loop if the placement of such orders causes market volume to increase, leading to the placement of even larger orders as observed in the 6th May 2010 Flash Crash ([CFTC and SEC, 2010](#)).

The present model enhances the role of distressed selling agents from [Paddrik et al. \(2012\)](#). Distressed sellers in our model come from two sources:

1. Exogenous distressed sellers, introduced to shock the system, and
2. Distressed sellers implementing the liquidation order placement of distressed funds.

First, as in [Paddrik et al. \(2012\)](#), we introduce a distressed seller with a given position size at a specific time point in the simulation purely to generate an asset price shock. We allow a period of 20% of the simulation duration to elapse before shocking assets to allow initial transient effects to decay. This distressed seller will continue to attempt to liquidate until it has sold its entire position, after which it becomes inactive.

The second source of distressed sellers in our model are distressed funds. When a fund is in a margin call or default state it is compelled to reduce leverage by liquidating asset positions. We model this by introducing a distressed seller agent to the market for every asset that the fund chooses to liquidate (in this chapter, this is all assets held by the fund). The distressed seller attempts to reduce its parent fund's position in each asset to zero by placing repeated volume-limited market orders according to a characteristic timescale controlled by parameter δ (sensitivity to this parameter is investigated in section 3.8). Critically, however, if fund leverage falls back below the tolerance threshold, liquidation is immediately suspended, and the distressed agents representing the fund are deactivated. It is possible that the fund may re-enter distress later and here the distressed selling agents would reactivate. Our justification for this deleverage mechanism is that margin call terms usually

stipulate that leverage must be brought under control within a single trading day, thus necessitating immediate and robust action (Brunnermeier and Pedersen, 2009). It is not unreasonable to assume that portfolio managers accept real-time feeds of market data in order to value their portfolios intraday.³

In the results we will present, banks and funds update their estimate of fund leverage by tracking asset prices in real-time. In our simulation this means that banks and funds operate at the same fine-grained time fidelity as the trading agent microstructure simulation ($\delta t = 50$ milliseconds). However, the timing of funds and banks is not critical, and our results were unchanged when fund and bank update frequencies were reduced by a factor of one hundred, to 5 seconds. Whichever timescale they utilise, implied fund leverage is updated according to equation (3.5), with distressed sellers being added to- or removed from the simulation immediately in response. This approach to the simulation of leverage management follows established practice (Brunnermeier and Pedersen, 2009; Cont and Schaanning, 2017). No complex decision making is undertaken by banks or funds over short timescales. Such strategic considerations are outside the scope of the present work (as discussed in Section 3.2).

3.3.6 Assets

Assets are not treated as agents in the model. Rather, assets represent the environment in which the other agents are situated (Wooldridge, 2001). Assets can be bought and sold by trader agents utilising market and limit orders via a realistic continuous double auction matching engine. The market clears according to the industry-standard price-time priority (Gould et al., 2013), and agents are able to view the current state of the market as represented by the LOB. Matched orders result in a trade, and the current price of the asset is thus deemed to have been updated. The simulation is enhanced by the realistic inclusion of an opening auction period and intraday volatility auctions. In both of these situations, market clearing is suspended for a period of time to allow liquidity to accumulate on the LOB and hence aid price discovery. We assume an initial one-minute opening auction phase and trigger five-second volatility auctions when prices fall by 1.3% in a single second, in line with prevailing intraday volatility limit mechanisms at the time of the Flash Crash, consistent with Paddrik et al. (2012).

³Appropriate real-time market data products are offered by Bloomberg <https://www.bloomberg.com/professional/product/market-data>, accessed 2018-09-17 and Thomson Reuters <https://financial.thomsonreuters.com/en/products/data-analytics/pricing-data-services.html>, accessed 2018-09-17.

3.4 Systemic Risk Metrics

For a comprehensive survey of systemic risk metrics the reader is referred to [Bisias et al. \(2012\)](#) and [Acharya et al. \(2017\)](#). For the present work we adopt two notions of systemic risk. The first is tailored to distress at *funds*, and the second is tailored to distress in *assets*.

3.4.1 Financial Contagion

We have available several measures of fund distress: (1) fund profit and loss, (2) funds entering margin call state, and (3) funds defaulting (when losses exceed capital). In the analysis that follows, we adopt funds entering default as our fund-level systemic risk metric. The advantage of this metric is that it represents an extreme situation, more-so than merely entering the margin call state, and also represents a terminal state of the fund (funds do not participate further in the simulation once they are in default). Using profit and loss would require the specification of the value of loss necessary to describe a fund as distressed. By using the notion of default we automatically determine this threshold (the relationship between profit/loss and default is stated in Section 3.3.1).

Following [Caccioli et al. \(2014\)](#) we measure financial contagion by considering the fraction of funds that enter default in each simulation run. If at least $\gamma = 5\%$ of funds are in default by the end of the simulation, we join [Caccioli et al. \(2014\)](#) in considering a *default cascade* to have occurred. We define the *probability of contagion* as the fraction of Monte-Carlo trials (see Section 3.5) in which a default cascade occurs. Again following [Caccioli et al. \(2014\)](#), we define the *extent of contagion* as the fraction of funds in default, conditional on a default cascade having occurred. Formally, the probability of contagion is given by

$$P_{contagion} = \frac{1}{M} \sum_{m=1}^M \mathbb{I}_{cascade}(m), \quad (3.6)$$

where M is the number of Monte-Carlo trials for a single location in the parameter space. Here, $\mathbb{I}_{cascade}(m)$ is an indicator variable representing the occurrence of a default cascade during trial $m \in (1, \dots, M)$, and is defined as

$$\mathbb{I}_{cascade}(m) = \begin{cases} 1 & \text{if } f_{default}(m) > \gamma, \\ 0 & \text{otherwise,} \end{cases} \quad (3.7)$$

where γ is a scalar threshold controlling the fraction of funds that must default before we consider a default cascade to have occurred. $f_{\text{default}}(m)$, the fraction of funds entering default during Monte Carlo trial m , is given by

$$f_{\text{default}}(m) = \frac{1}{n_f} \sum_{i=1}^{n_f} \mathbb{I}_{\text{default}}(i, m), \quad (3.8)$$

where $\mathbb{I}_{\text{default}}(i, m)$ is an indicator variable representing the default of fund i in trial m , and is defined as

$$\mathbb{I}_{\text{default}}(i, m) = \begin{cases} 1 & \text{if fund } i \text{ defaults in trial } m, \\ 0 & \text{otherwise.} \end{cases} \quad (3.9)$$

The extent of contagion, $\Omega_{\text{contagion}}$, is defined by [Caccioli et al. \(2014\)](#) as the expected fraction of funds entering default, given a default cascade occurs. Formally, let us define $\mathcal{D}$, the set of Monte Carlo trials in which a default cascade occurs, such that $\mathcal{D} = \{m \in (1, \dots, M) : \mathbb{I}_{\text{cascade}}(m) = 1\}$, then we may write

$$\Omega_{\text{contagion}} = \frac{1}{|\mathcal{D}|} \sum_{m \in \mathcal{D}} f_{\text{default}}(m). \quad (3.10)$$

Note that $\Omega_{\text{contagion}}$ is undefined if no default cascades occur, i.e., if $\mathbb{I}_{\text{cascade}}(m) = 0 \forall m \in (1, \dots, M)$, giving $|\mathcal{D}| = 0$.

3.4.2 Flash Crash Occurrence

The second systemic risk indicator concerns distress at assets, and here again we have a choice of possible metrics. [Paddrik et al. \(2012\)](#) look at the lowest price attained by simulated assets, which has the benefit of simplicity but lacks consideration of the time dimension so characteristic of flash crash phenomena. Although price falls are indeed material to our investigation, our research questions are also concerned with the speed at which distress propagates, and the temporal characteristics of sudden price movements are better-described by metrics that are specifically designed to detect flash crashes. [Johnson et al. \(2012\)](#) consider both price movement magnitude and timescale, but also require an unbroken sequence of trade prices in a given direction ([Johnson et al., 2012, p.3](#)):

“For a large price drop to qualify as an extreme event (i.e., black swan crash) the stock price had to tick down at least ten times before ticking up and the price change had to exceed 0.8% . . . In order to explore timescales which go beyond typical human reaction times, we focus on black swans with durations less than 1500 milliseconds”

We encountered difficulty using [Johnson et al. \(2012\)](#)’s method. In particular the metric was not found to be robust to simulation scaling (Section 3.5).

[Vuorenmaa and Wang \(2014\)](#) define a flash crash as a price fall of 10% followed by recovery of at least 50% of the size of the fall over an unspecified short timescale. [Jacob Leal et al. \(2016\)](#) use a similar metric based on a price fall of 5% followed by an unspecified degree of reversion within a period of 30 minutes. The dependence on price reversion in each of these metrics was found to be problematic for our model where prices may not revert within the simulated timescale.

We define a flash crash as a price fall of 5% over a rolling 5-minute window. No further constraints are applied. This metric was found to be robust to simulation scale, and to provide a better false positive rate than using the occurrence of volatility auctions (triggered by falls of 1.3% in a single second, see Section 3.3.6). Formally, let us define a variable, $\mathbb{I}_{FC}$, indicating the presence of a flash crash in asset $j \in (1, \dots, n_a)$ at time $t \in (1, \dots, T)$, as

$$\mathbb{I}_{FC}(j, t) = \begin{cases} 1 & \frac{(p_j^t - p_j^{t'})}{p_j^{t'}} < -0.05, \\ 0 & \text{otherwise,} \end{cases} \quad (3.11)$$

where $p_j^{t'}$ is the price of asset j at $\max(t - 5 \text{ minutes}, 0)$.

3.4.3 Flash Crash Propagation Speed

In addition to considering the fraction of assets experiencing flash crashes according to the above definition, we also consider the speed of flash crash propagation through the system. Propagation speed is only well-defined where the number of assets experiencing flash crashes is at least two. We measure the time since the start of the simulation run at which assets experience their first flash crash (individual assets may go on to experience several flash crashes but we do not need to account for this in our propagation speed metric). We then sort these time intervals in ascending order and take the time difference between the first flash crash and the 80th percentile flash

crash as a measure of the time taken for flash crash propagation in the system. We scale this time interval inversely according to the number of assets that experience flash crashes, giving a resulting metric with units of flash-crashes per minute.

Let us define the set $\mathcal{F}^j$ of flash crashes experienced by asset j such that $\mathcal{F}^j = \{t \in (1, \dots, T) : \mathbb{I}_{FC}(j, t) = 1\}$, where T is the total number of simulated time steps, and $\mathbb{I}_{FC}(j, t)$ is the indicator variable defined above.

Let us further define a total order $\leq_j$ over $\mathcal{F}^j$, then we may label the elements $t_i^j \in \mathcal{F}^j$ with indices $i \in (1, \dots, |\mathcal{F}^j|)$ such that $t_1^j \leq_j t_2^j \leq_j \dots \leq_j t_i^j \leq_j \dots \leq_j t_{|\mathcal{F}^j|}^j$. We define the set $\mathcal{F}$ of first flash crashes experienced by each asset in a single Monte Carlo trial as

$$\mathcal{F} = \bigcup_{j=1}^{n_a} \{t_1^j\}. \quad (3.12)$$

We also define a total order $\leq$ over $\mathcal{F}$. We label the elements $t_i \in \mathcal{F}$ with indices $i \in (1, \dots, |\mathcal{F}|)$ such that $t_1 \leq t_2 \leq \dots \leq t_i \leq \dots \leq t_p \leq \dots \leq t_{|\mathcal{F}|}$, where p is the index of the ζ th fractile of the ordered set, i.e., $p = \lfloor \zeta |\mathcal{F}| \rfloor$. We fix $\zeta = 0.8$ in the present work. Finally, we define the speed s of flash crash propagation by

$$s = \frac{c\zeta(|\mathcal{F}| - 1)}{(t_p - t_1)}, \quad |\mathcal{F}| > 1, \quad (3.13)$$

where c is a dimensionless constant that allows us to convert from units of (time steps) $^{-1}$ to (minutes) $^{-1}$. If δt is the duration of one simulated time step in milliseconds, we have $c = 60,000\delta t^{-1}$.

An important feature of this metric is that it smooths out propagation speed for a given Monte Carlo trial. We observed that propagation speed exhibits non-uniform temporal structure including bursts of activity. However, we also observed that the speed results are noisy and depend strongly on network topology (we return to this matter in Section 3.10). The smoothing implicit in this metric therefore yields a more robust estimate of the propagation speed than naively using the maximum speed attained in a given trial.

3.5 Simulation Protocol Summary

Our simulation proceeds according to the following steps:

1. *Setup (meta)*: Exogenous selection of parameter space region to explore and the number of stochastic Monte Carlo trials to perform for each parameter setting. Monte Carlo instances dispatched to high-performance computing cluster.⁴
2. *Setup (per-trial)*: Construction of agents with stochastic parameters determined via pseudorandom number generation, including fund–asset network construction.
3. *Intraday model*: Simulate $T = 100,000$ periods of market activity, each step lasting $\delta t = 50$ milliseconds.
4. *Single period*: Agents are selected for order placement according to a Poisson process. LOBs are updated in industry standard price-time priority.
5. *Shock*: After $T^* = \lfloor 0.2T \rfloor$ periods to allow for the decay of initial transient effects, a distressed selling agent is introduced for a single asset. We then observe subsequent system dynamics, and measure systemic risk.
6. *Distributional outputs*: Repeated stochastic Monte Carlo trials facilitate the collection of distributional results.

Due to the stochastic, path-dependent nature of the model (a common feature of agent-based models), it is necessary to perform repeated experimental trials in order to build statistical confidence in the observed systemic dynamics (Fagiolo et al., 2019; Franke and Westerhoff, 2012). This approach, widely known as Monte Carlo sampling, induces a considerable computational burden. Our simulation protocol accommodates the computational cost in two ways — firstly by making highly-parallel use of a distributed high-performance computation cluster, and secondly by introducing judicious granularity into the agent simulation⁵. Facing similar computational constraints, Paddrik et al. (2012) introduce granularity by scaling the number of trader agents in their simulation⁶. Table 3.3 gives unscaled agent populations. Following Paddrik et al. (2012) we divide these populations by a factor of 32 (we found our systemic risk metrics were robust to simulation at $1/32$ or $1/4$ scale). Since this scaling results in reduced LOB volume, it is also necessary to scale agent inventory limits by the same factor. Agent activity timescales and order placement sizes are not scaled. We adopt the same time granularity used in the flash crash model of Vuorenmaa and

⁴We would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work <http://dx.doi.org/10.5281/zenodo.22558>.

⁵Performance of the LOB software component was also found to be critical. We extended the implementation at <https://github.com/ab24v07/PyLOB>, accessed 2017-01-31, available under the MIT license and developed as part of Booth (2016).

⁶Hayes et al. (2014) provide a reference software implementation at <https://github.com/uva-financial-engineering/JinSup>, accessed 2017-02-23, available under a permissive license.

Wang (2014) and set $\delta t = 50\text{ms}$. It is possible that the Poisson arrival processes governing agent timing result in more than one agent being selected to place orders on a single LOB during a single time period. In such cases, agent activation order is randomised. Since agents are therefore unable to have perfect knowledge of the LOB at the exact time of their order placement, this randomisation places a lower-bound on agent trading latency. The most active agents (HFTs) arrive at the market according to a characteristic timescale of 350ms (see Table 3.3), about an order of magnitude slower than the minimal order placement latency supported in our simulation.

3.6 Mathematical Notation

Table 3.4 presents a summary of the parameters used in this chapter. The sampled domain for each parameter is discussed at point of parameter usage in the results sections that follow. Table 3.5 summarises other non-parameter notation.

Symbol	Name	Domain	Sampled domain
n_f	Number of funds	$\mathbb{N}_+$	$(0, \dots, 195)$
n_a	Number of assets	$\mathbb{N}_+$	$(1, \dots, 195)$
A_{ij}	Asset portfolio values ($n_f \times n_a$)	$\mathbb{R}$	$[0, \sim \$10^8]$
S_{ij}	Asset portfolio positions ($n_f \times n_a$)	$\mathbb{N}_+$	$(0, \dots, \sim 10^4)$
M	Number of Monte Carlo replicates	$\mathbb{N}_+$	$(1, \dots, 100)$
T	Number of simulated time steps	$\mathbb{N}_+$	100,000
δt	Time step duration / milliseconds	$\mathbb{N}_+$	50
ρ	Erdős-Rényi network density (or diversification)	$[0, 1]$	$[0, 1]$
k	Network degree of fund nodes	$\mathbb{N}$	$(1, \dots, 195)$
β	Preferential attachment coefficient	$\mathbb{R}$	$[-7.5, 7.5]$
λ^0	Initial fund leverage	$\mathbb{R}_+$	$[1, 20]$
C^0	Initial fund capital / \$millions	$\mathbb{R}_+$	$[0.1, 5]$
τ_c	Margin hysteresis critical value	$[1, \infty)$	$[1.001, 1.2]$
θ	Parameter triple (λ^0, τ_c, C^0)	-	-
α	Fund portfolio uniformity coefficient	$[0, 1]$	$[0, 1]$
η	Distressed seller volume fraction	$\mathbb{R}_+$	$[0.05, 0.18]$
δ	Distressed seller timescale / s	$\mathbb{R}_+$	$[1, 60]$
γ	Default cascade indicator threshold	$(0, 1]$	0.05
ζ	Flash crash speed measurement percentile	$[0, 1]$	0.8

Table 3.4: Parameters in the model. The sampled domain for each parameter is discussed at point of parameter usage in Sections 3.7–3.10.

Symbol	Name	Range
V_i	Total investment value for fund i	$\mathbb{R}_+$
L_i	Total loan value made to fund i	$\mathbb{R}_+$
p_j^t	Price of j th asset at time t	$\mathbb{R}_+$
Q_j^{bid}	Total shares of j th asset at best bid price	$\mathbb{N}_+$
Q_j^{ask}	Total shares of j th asset at best ask price	$\mathbb{N}_+$
$P_{contagion}$	Probability of contagion	$[0, 1]$
$\Omega_{contagion}$	Extent of contagion	$[0, 1]$

Table 3.5: Other notation used in this chapter.

3.7 Empirical Validation

Figure 3.3(a) demonstrates the output from a single evaluation of the ABM in which we simulate a single asset and a single fund. The simulation comprises $T = 100,000$ time steps each representing $\delta t = 50\text{ms}$ for a total of approximately 1.5 hours of simulated trading. The simulation was performed at $1/4$ scale, following the method established by Paddrik et al. (2012).

The system is shocked after an initialisation period at $t = 20,000$ by forcing the fund to liquidate all of its inventory. The initial asset holdings of the shocked fund and its liquidation behaviour are selected to generate equivalent sell order placement activity to that exhibited by the distressed trader in the 6th May 2010 Flash Crash (CFTC and SEC, 2010). Figure 3.3(a) shows the price evolution of the system (black, left hand axes) and the one-minute binned trade volume (red, right hand axes). We can see the characteristic crash and recovery signature of a flash crash occurring around $t = 60,000$. Market volume and market volatility both increase while the distressed selling is taking place and return to pre-crash levels after the distressed selling has concluded. The zero-intelligence nature of the trader agents precludes any temporally-extended response to volatility and so, unlike in the real Flash Crash of 6th May 2010, the model does not capture the extended period of volatility that followed the real crash (CFTC and SEC, 2010).

Figure 3.3(b) shows the total volume on the LOB during the simulation disaggregated into bid (black) and ask (red) volumes. We can see that the available bid volume collapses following the introduction of the distressed selling agent at $t = 20,000$. As discussed in Section 3.3.4, the distressed selling agent places sell market orders which aggressively remove liquidity from the bid side of the LOB. This figure clearly shows the cause of the eventual flash crash — just before $t = 60,000$ there is a brief period

where buy-side liquidity collapses to almost zero. In such a regime, the orderly provision of liquidity to the market has failed, and *stub quotes* (orders placed far from prevailing prices that are not intended to transact) may be executed. This is because market orders do not have price limits and so trades can occur at arbitrarily low prices (Cristelli et al., 2010). Once the distressed selling has completed after $t = 60,000$, buy side volume returns to pre-distressed levels, and fundamental buyer agents are able to restore the price to pre-crash levels.

Figures 3.3(c–f) demonstrate that the model reproduces several important stylised facts of the market (Cont, 2001; Kirman and Teyssiere, 2002; Chen et al., 2012). Figure 3.3(c) demonstrates that the model produces heavy-tailed returns with a power-law tail. We give the mean cumulative distribution function across 160 Monte-Carlo trials with 95% confidence intervals for trials with (red) and without (black) flash crashes. For comparison, we also show Gaussian return distributions with mean and variance matched to the observed results. Whether a flash crash occurs or not, the returns are heavy tailed. The heavy-tailed nature is much enhanced where flash crashes do indeed occur.

Figure 3.3(d) demonstrates *aggregational Gaussianity* (Kirman and Teyssiere, 2002), whereby returns taken over longer time horizons appear more Gaussian and less heavy-tailed. Again, we show 95% confidence intervals around the mean across 160 Monte-Carlo trials disaggregated into trials with and without flash crashes. We plot the kurtosis of the return distribution (a Gaussian distribution has *kurtosis* = 3). Both series converge towards the Gaussian kurtosis level as we increase the measurement horizon over 10,000 simulated time-steps, and the decay is significantly more pronounced for trials with a flash crash.

Figure 3.3(e) demonstrates a lack of autocorrelation of price returns except at very short intraday timescales, and Figure 3.3(f) demonstrates a long tail in the decay of autocorrelation of volatility (absolute return). Both panels plot mean results across 160 Monte-Carlo trials with 95% confidence intervals. The slow decay of autocorrelation is markedly slower for trials exhibiting flash crash phenomena. These results are consistent with empirically-observed stylised facts (Cont, 2001). Indeed, strong negative autocorrelation of returns at short timescales is reported by, for example, Hayes et al. (2014) and Kalnina (2011). The decay of absolute return autocorrelations produced by the model is faster than may be expected in reality. This may be attributable to the zero intelligence nature of the trading agents who lack either memory or adaptive behaviours.

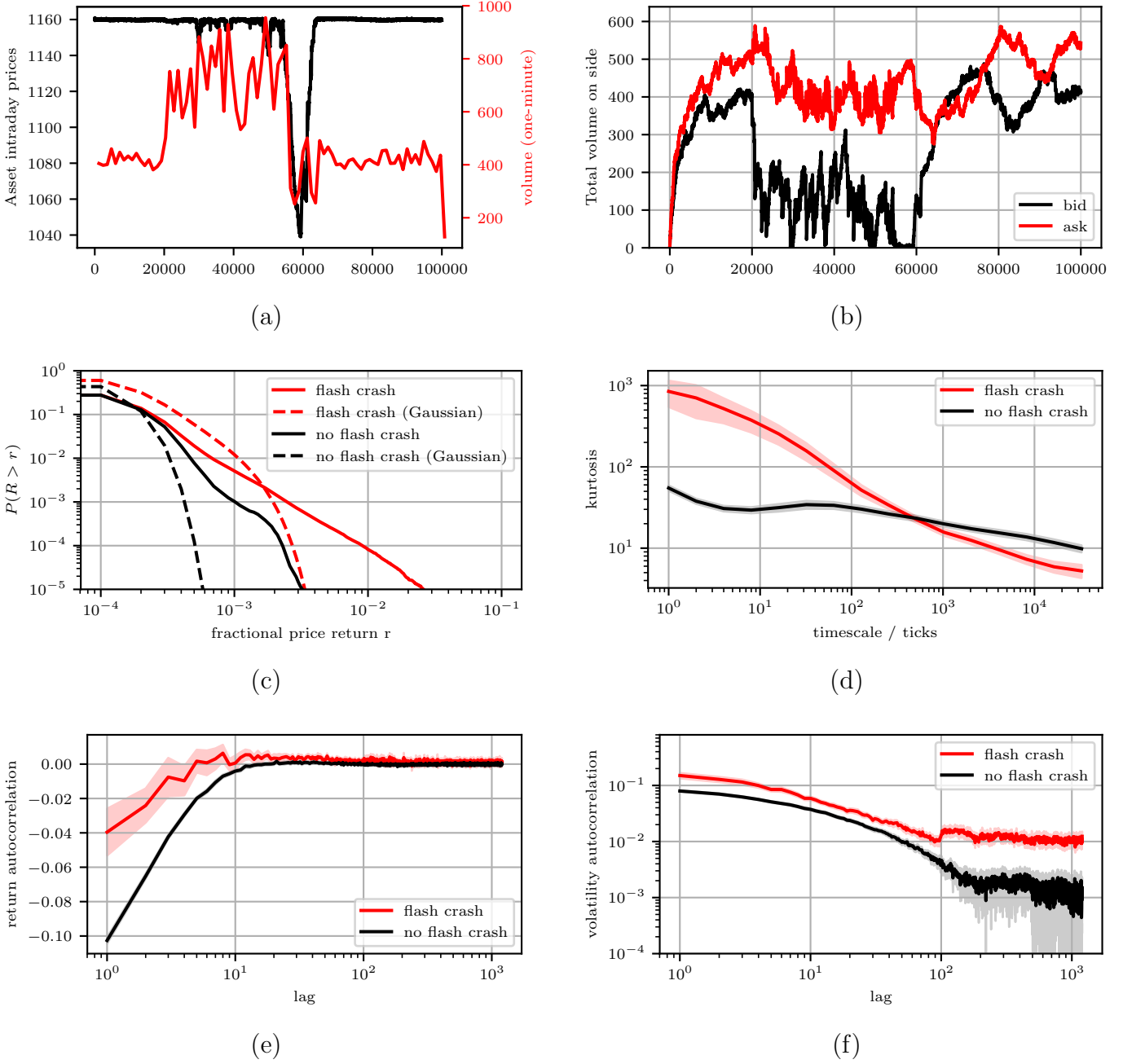


Figure 3.3: Sample model output for a single asset. (a) price and volume series demonstrating the occurrence of a flash crash event. (b) Total volume on bid (black) and ask (red) sides of the LOB demonstrating the collapse of buy-side volume prior to the flash crash. The simulation lasts for $T = 100,000 \times \delta t = 50\text{ms}$ simulated time steps, therefore represents approximately 1.5 hours of market activity. (c) cumulative distribution function over price returns (log-log scale). (d) return kurtosis measured over increasing time horizons (log-log scale). (e) demonstrating lack of autocorrelation of price returns except at very short horizons (x-axis log scale). (f) demonstrating slowly-decaying autocorrelation of volatility (log-log scale). For panels (c)–(f) we plot the mean across 160 Monte-Carlo trials with 95% confidence interval.

Paddrik et al. (2012)’s model formulation requires all order placement by all agents to be within ten *ticks* (quantised minimum price increments on the LOB) of the current best price. Their model is therefore precluded from generating the long tailed relative price stylised fact reported by Zovko et al. (2002). A consequence of that modelling decision is that their model does not generate stub quotes. As a result, the Paddrik et al. (2012) flash crash price gradient appears shallow relative to an empirical flash crash or those generated by other models such as Vuorenmaa and Wang (2014). Our model relaxed the ten-tick assumption for the small trader agents only, permitting them to place orders at up to one-thousand ticks away from best. This resulted in an as-expected improvement in the model’s ability to reproduce the relative price stylised fact. Flash crashes generated under this relaxed assumption exhibit a steeper price gradient, in better agreement with empirical data (Figure 3.3(a)).

For all of these simulations we held-fixed the distressed selling agent properties. Flash crash occurrence is, however, strongly dependent on the price impact of the distressed selling agent’s orders, which we would expect in turn to strongly depend on order size and order placement rate (Paddrik et al., 2012). We explore this matter in the next section.

3.8 Trading Behaviour and Asset Price Shocks

The distressed selling agent in the Flash Crash of 6th May 2010 reportedly placed a programme of liquidity-removing orders sized according to a fixed fraction of rolling one-minute market volume (Menkveld and Yueshen, 2019). They furthermore executed this sequence of orders, intended to achieve their total desired sell volume, over a period of 20 minutes. Paddrik et al. (2012) and Paddrik et al. (2017) each investigate sensitivity to the volume limit (η in our notation). In the analysis that follows we additionally investigate sensitivity to the placement rate (δ in our notation). We find that flash crash occurrence is strongly dependent on both parameters.

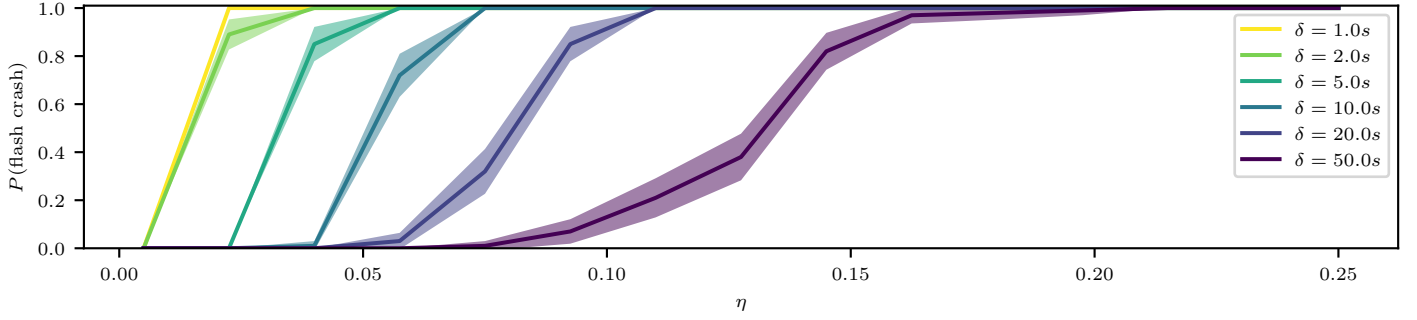
In this experimental scenario we retained the single-fund, single-asset structure adopted in the previous section. The method of shock delivery via the complete liquidation of the fund portfolio is also retained. We performed a combinatorial evaluation over the parameter space $\eta \in [0.005, 0.25]$ and $\delta \in [1, 50]$ sampling 77 parameter pairs. For each parameter pair we performed 100 Monte-Carlo trials. Each trial was classified according to whether a flash crash was observed in the simulated

asset price time-series. We define flash crash occurrence as a price fall exceeding 5% in a period of 5 minutes (see Section 3.4.2 for details regarding this metric).

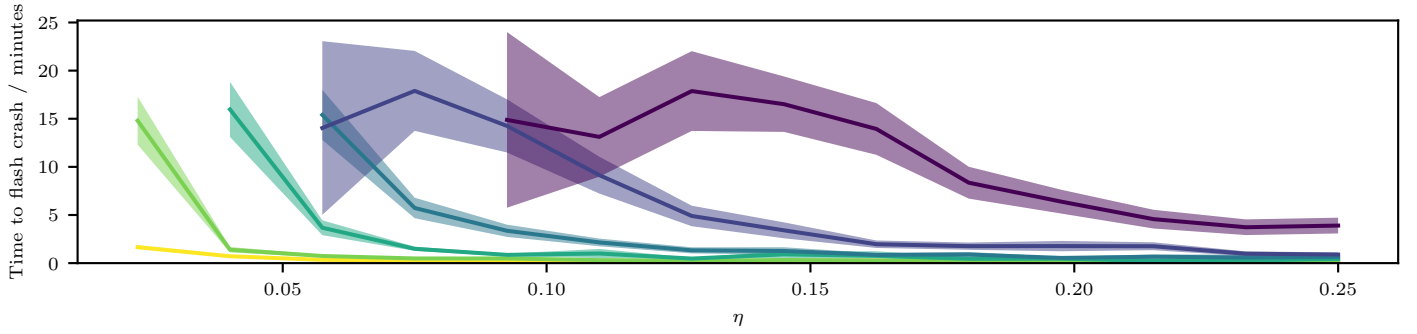
The mean and 95% confidence intervals for Monte Carlo results are plotted in Figure 3.4(a). What can be clearly seen in this figure is that flash crashes tend to occur with higher probability as the sizes of orders placed increase, and as the interval between order placements decreases. Both of these findings are consistent with the notion that a flash crash occurs when there is insufficient buy-side liquidity to meet elevated sell-side demand, possibly exacerbated by algorithmic traders withdrawing liquidity from the market. We find that, for any given order size limit in the regime studied, it is possible to reduce the probability of flash crash occurrence by increasing the interval between order placements. Figure 3.4(a) also demonstrates that, for a given order placement interval, δ , there exists a critical value of the order size limit, η , below which flash crashes occur with negligible probability. Two traders placing equally-sized orders at equal timescales δ is exactly equivalent to a single trader placing orders with a timescale $\delta' = \delta/2$.

The distressed seller implicated in the Flash Crash of 6th May 2010 reportedly used a value of $\eta = 0.09$ (Menkveld and Yueshen, 2019). We initialise this parameter setting in our model using this reported value, and retain this setting for the remainder of the results in this chapter. We select $\delta = 10$ s as a reasonable setting based on the results in Figure 3.4(a). Figure 3.4(b) implies that this choice of η and δ is sufficient to induce a flash crash within approximately two minutes of the onset of distressed selling.

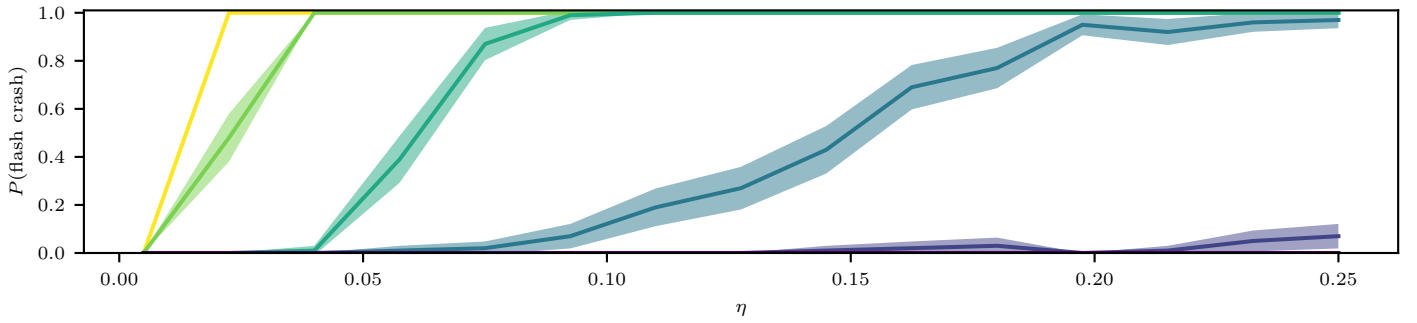
The distressed seller in our model places market orders, following methodology established by Foucault et al. (2005), Paddrik et al. (2012), Filimonov and Sornette (2012), and Vuorenmaa and Wang (2014). However, more recent research by Menkveld and Yueshen (2019) facilitated by “unique access to a data set that contains the large seller’s trades” (Menkveld and Yueshen, 2019, p.1), indicates that during the Flash Crash of 6th May 2010, the large distressed seller instead used liquidity-taking (*marketable*) limit orders. Figures 3.4(c) and 3.4(d) demonstrate that the model exhibits qualitatively similar behaviour under two alternative limit order placement scenarios, as when using market orders (Figure 3.4(a)). Figure 3.4(c) presents the probability of occurrence of flash crashes when the distressed seller places limit orders with prices drawn from a uniform distribution $\sim \text{Uniform}(bid - 10 \text{ ticks}, bid)$, where a *tick* refers to the minimum permitted price increment on the limit order book. It is apparent that at larger order placement intervals δ , flash crashes become less likely when placing limit orders instead of market orders. The results in Figure 3.4(d) were



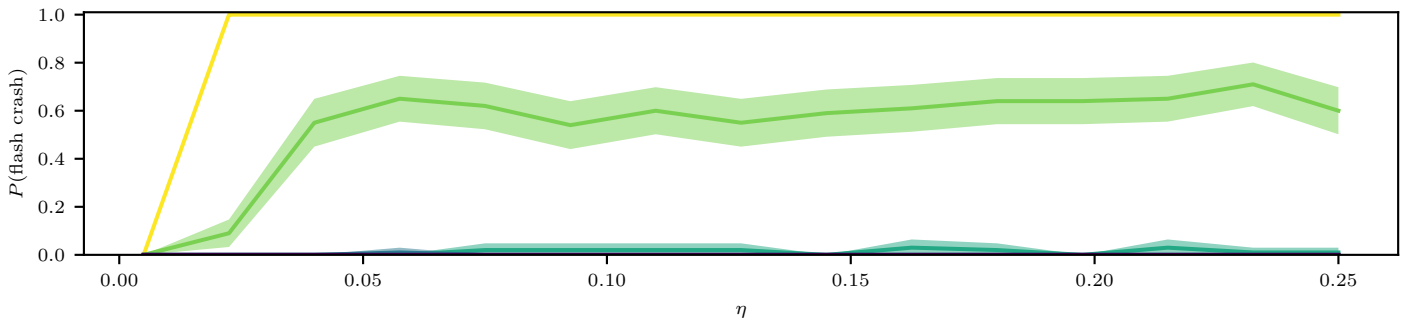
(a) Probability of flash crash occurrence (market orders).



(b) Time interval prior to flash crash occurrence following the onset of distressed selling.



(c) Probability of flash crash occurrence (limit orders).



(d) Probability of flash crash occurrence (pegged limit orders).

Figure 3.4: Single asset flash crash (a,c,d) occurrence and (b) time interval prior to occurrence as a function of distressed selling agent properties for a range of order types. η refers to the order size limit as a fraction of rolling one-minute market volume. δ refers to the mean interval between sell order placements in seconds. We show the mean and 95% confidence intervals sampled over 100 Monte Carlo trials.

obtained by fully replicating the distressed seller behaviour described by [Menkveld and Yueshen \(2019, p.12\)](#) “The large seller paid for demanding liquidity, not by taking liquidity beyond the best bid (none of his trades walked down the order book)”. We implemented this behaviour by introducing an additional limit order placement method in which distressed sell orders execute exactly at the best bid price, and are therefore constrained by the volume available at that price level on the limit order book⁷. Here we find that flash crashes only occur for distressed seller order placement intervals $\delta \leq 5$ seconds. Our result is consistent with the empirical observation that the real distressed seller was active with $\delta \leq 2$ seconds, as inferred from Figure 4 in [Menkveld and Yueshen \(2019\)](#). In summary, although an alternative model specification using rapid placement of limit orders close to the prevailing market price more closely matches the observed behaviour of the large distressed seller during the 2010 Flash Crash, the qualitative model behaviour, for the research questions considered in this chapter, is unchanged under such a specification. In order to promote comparison with existing literature, we therefore assume market order driven distressed seller behaviour.

3.9 Leverage, Capital, and Margin Tolerance

Previous sections have established the effectiveness of the model at reproducing stylised facts of real financial markets, as well as its ability to exhibit flash crashes. We now proceed to consider a full simulation of multiple connected funds and assets, as opposed to the single-fund single-asset simulations studied in previous sections.

In this section, we consider the response of the system to changes in fundamental parameters related to margin tolerance (τ_c), leverage (λ^0) and fund capital (C^0). We find that the stability of the multi-asset system is strongly dependent on each of these parameters, moving from a stable regime where it is unlikely for fund defaults and flash crashes to occur, to an unstable regime where catastrophic collapse becomes highly likely.

Formal definitions of our systemic risk metrics for fund default cascades and asset flash crash occurrence are given in Section 3.4. In brief, we consider a default cascade to have occurred if more than $\gamma = 5\%$ of funds end the simulation in default. We further define the extent of a default cascade to be the fraction of funds in default, conditional on a default cascade having occurred. Both of these metrics follow the

⁷We found similar results regardless of whether any residual un-filled order volume is allowed to rest on the book, or is cancelled. We report results from the latter case, which we described as a *pegged* limit order.

methodology of [Caccioli et al. \(2014\)](#). We also introduce our own metric for flash crash occurrence, which we define as an asset price fall of 5% within a rolling five minute window.

3.9.1 Experiment Parameters

For this experimental scenario, we perform a systematic exploration over parameter triples $\theta(\lambda^0, \tau_c, C^0)$ for $\lambda^0 \in [1, 20]$, $\tau_c \in [1.002, 1.1]$ and $C^0 \in [0.1, 5.0]$ \$millions (210 total combinations). The analysis performed by [Ang et al. \(2011\)](#) suggested that real-world λ^0 typically lies within $[1, 20]$ for hedge funds trading equities. Since we do not have access to an independent data source for our calibration, we adopt [Ang et al. \(2011\)](#)'s suggested domain. Analysis of the 6th May 2010 Flash Crash implied the size of sell orders sent by the distressed trader totalled 75,000 ([CFTC and SEC, 2010](#)). We sample $C^0 \in [0.1, 5.0]$ such that in combination with λ^0 (see Section 3.3.1 for details), the positions held by funds extend from the region whereby full liquidation of a single fund could not induce a flash crash even if all of its holdings were allocated to a single asset, to the region whereby liquidation of a single fund could indeed induce a flash crash. We do not have access to empirical data on margin hysteresis τ_c and so here we perform a logarithmic sampling across three orders of magnitude⁸ in order to determine sensitivity to this parameter. In combination, these parameter domains are sufficient to determine the boundary in (λ^0, τ_c, C^0) space separating stable from unstable behaviour. Each parameter triple, θ , is evaluated over five Monte-Carlo trials⁹ resulting in a total of 1050 ABM evaluations. Only the pseudorandom number seed is varied between Monte Carlo trials.

For each evaluation we generate a new set of fund–asset allocations according to the algorithm presented in Section 3.3.3. Allocations are stochastic with a fixed diversification coefficient $\rho = 0.5$ and no preferential attachment ($\beta = 0$). The topology may therefore be represented as an Erdős-Rényi random graph ([Newman, 2003](#)) in which, on average, edges exist between each fund node and 50% of the available asset nodes. We consider the effect of topology in the next section. Finally, we fix the number of assets and funds at $n_a = n_f = 50$. These selections strike a balance between computational tractability and providing fine enough grained distinction for metrics such as flash crash propagation speed and extent. For the present study, all funds have the same level of capital C^0 and initial leverage λ^0 .

⁸i.e., we sample $\tau_c - 1 = \tau'_c \in [0.002, 0.1]$.

⁹The low number is due to computational constraints which can be improved in future studies.

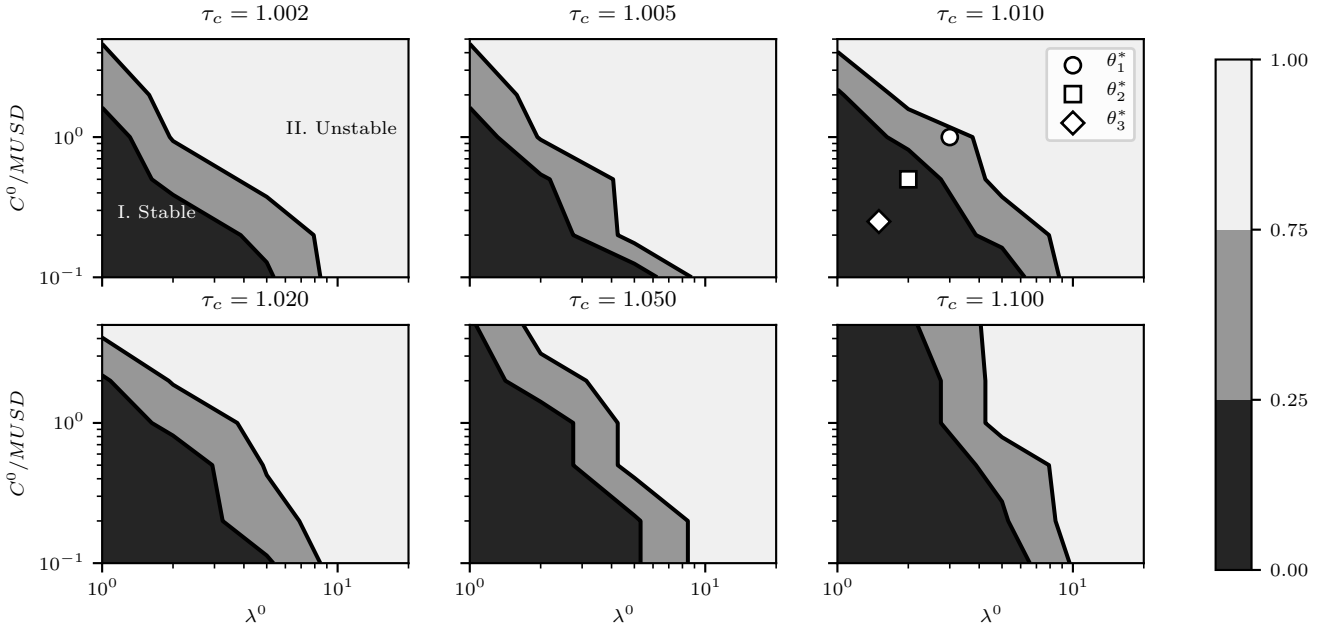


Figure 3.5: Probability of contagion as a function of leverage (λ^0), margin hysteresis (τ_c) and fund capital (C^0). Plotted values are the mean across Monte-Carlo trials as a function of λ^0 and C^0 . τ_c increases from panel to panel thus we present slices of the full three-dimensional parameter space at fixed values of τ_c . Contour lines are drawn to denote fractional default at the 25% and 75% levels. In region I most funds avoid default, but in region II it is likely that all funds end up in default. In the upper-right panel we label three points in (λ^0, τ_c, C^0) parameter space θ_1^* , θ_2^* , and θ_3^* . We refer to these in detail in Section 3.10.

The system is perturbed with an exogenous shock according to the same protocol as that presented in previous sections. We introduce a distressed seller to a single asset (selected from a uniform distribution across all assets that are held by at least one fund). The distressed agent places a sequence of sell orders with a total quantity of shares that is sufficient to trigger a flash crash in the shocked asset. No further exogenous interventions are performed, and further episodes of distressed selling by funds and asset flash crashes arise endogenously.

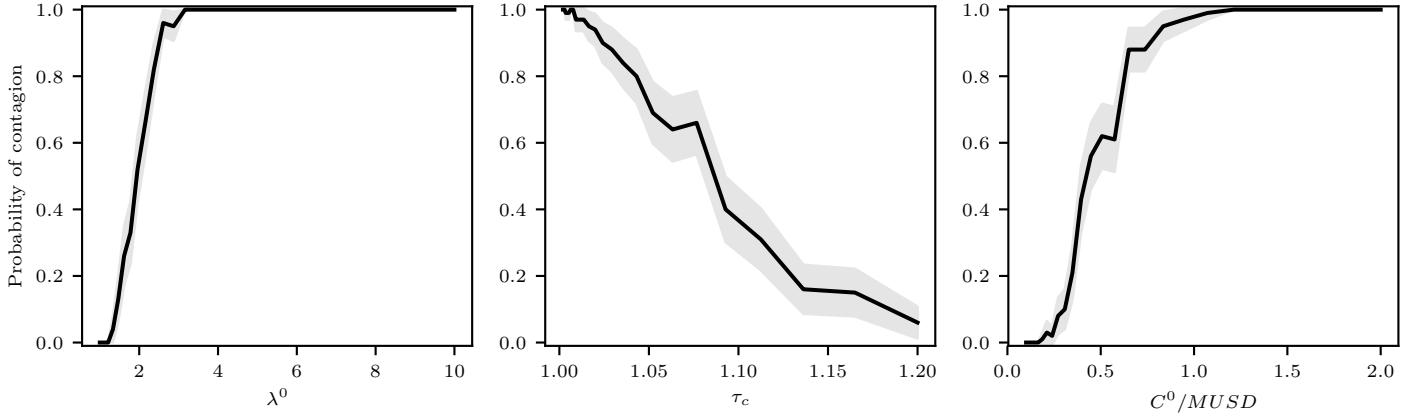
3.9.2 Parameter Space Exploration

Figure 3.5 shows the fraction of funds entering default (mean across Monte-Carlo trials) as we perform the combinatorial search. τ_c is increased from panel to panel thus we present slices of the full three-dimensional parameter space at fixed values of τ_c . It is clearly seen from this figure that leverage-capital space is partitioned into regions

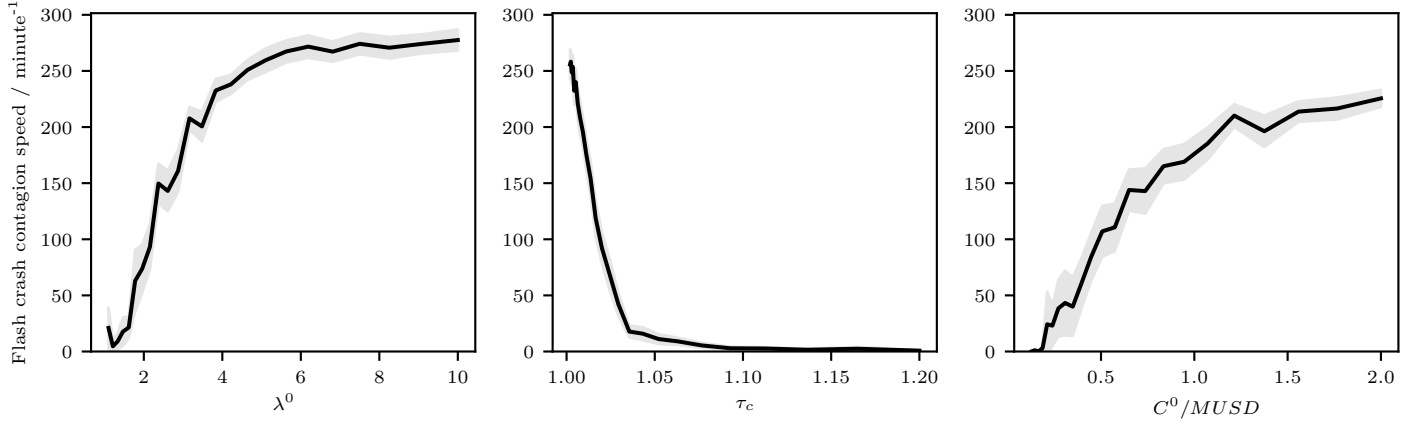
giving rise to low and high systemic risk, whereby the likelihood of funds entering default increases from approximately zero (region I) to approximately 100% (region II). This result is consistent with expectations because a larger value of leverage λ^0 magnifies fund losses in response to a given asset price move, making margin calls more likely. This effect is directly mitigated by increasing margin tolerance τ_c . Conditional on a margin call being in force, larger fund capital C^0 means that a greater quantity of shares will need to be sold by the fund, which therefore implies a greater price impact when that fund elects to sell, propagating greater losses to other funds that hold positions in the same assets. Systemic sensitivity to the topological properties of overlapping (crowded) asset portfolios are explored in detail in Section 3.10.

Although it is not easy to see in Figure 3.5, systemic risk changes significantly over the domain of τ_c . In order to make this more apparent, we performed linear sensitivity analyses of the three variables in isolation, holding the other two variables fixed. The three sampled lines intersect at $\theta_1^* = \theta(\lambda^0 = 3, \tau_c = 1.01, C^0 = 1)$. Figure 3.6(a) plots the probability of contagion for these linear explorations, showing 95% confidence intervals around the mean. This figure demonstrates that systemic risk is strongly dependent on all three variables under consideration. The change from stability to instability occurs as leverage and capital increase and as margin tolerance decreases. We further observed that the extent of contagion was always 100%, conditional on the occurrence of a default cascade. This result adds evidence in support of existing studies. Similarly to Huang et al. (2013), we find that increasing margin tolerance (τ_c) reduces the likelihood of systemic failure, and also that increasing price impact correspondingly increases systemic risk. In our model, price impact is a function of the distressed liquidity entering the market which, by construction, is an increasing function of leverage (λ^0) and capital (C^0).

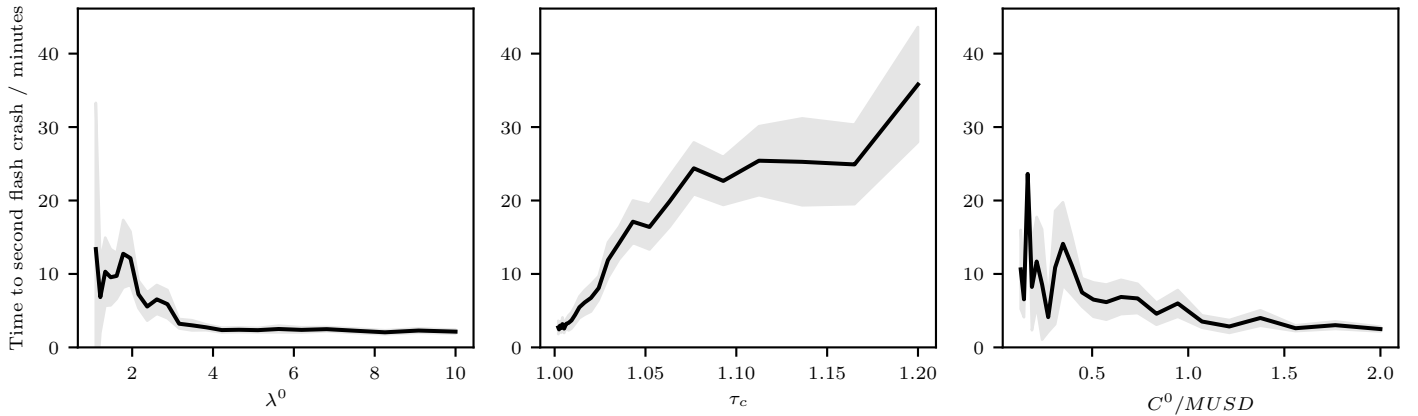
Figure 3.6(b) plots the speed of flash crash propagation for the linear explorations. As discussed in Section 3.4.3, the speed we report here is inversely related to the time taken for 80% of subsequent flash crashes to occur, following the observation of the first flash crash. These results demonstrate that flash crash propagation speed is an increasing concave function of leverage and capital, and a decreasing convex function of margin hysteresis, similarly to the probability of contagion result. Margin hysteresis has a dramatic damping effect on flash crash propagation speed, reducing propagation speeds close to zero for $\tau_c > 1.03$. These results are straightforward to explain. Increasing leverage makes funds more sensitive to price changes at an asset, making funds more likely to commence distressed selling for a given asset price move, and hence more readily propagating price impact between assets. Increased capital



(a) Probability of contagion as we systematically adjust one variable, holding the other two fixed.



(b) Flash crash propagation speed as we systematically adjust one variable, holding the other two fixed.



(c) Time taken for the second flash crash to occur following the introduction of the distressed selling agent, as we systematically adjust one variable, holding the other two fixed.

Figure 3.6: Exploring financial stability as we systematically adjust one agent parameter, holding the other two fixed. The exogenous variables are Left panels: leverage λ^0 , Middle panels: margin hysteresis τ_c , Right panels: capital C^0 (units millions of US dollars, *MUSD*). We plot the mean across 100 Monte-Carlo trials, along with 95% confidence intervals. The intersection in parameter space is at θ_1^* ($\lambda^0 = 3, \tau_c = 1.01, C^0 = 1$). The results demonstrate that systemic risk is strongly dependent on each parameter.

means that, conditional on distressed selling taking place, a fund will sell a larger quantity of shares, resulting in steeper price impact in assets that are being sold. This in turn means that the conditions satisfying our flash crash metric (a fall of 5% within a rolling 5 minute window) will be satisfied sooner than under lower capital conditions. The effect of increasing margin hysteresis is predominantly in opposition to the effect of increasing leverage. However, conditional on flash crashes occurring under high margin tolerance, the total quantity of shares to be unwound will be higher than when tolerance is small. This leads to a mixed effect of the margin tolerance parameter.

Finally, Figure 3.6(c) illustrates the time taken for the second flash crash to occur following the introduction of an exogenous shock to the system. The first flash crash typically occurs in the asset receiving the exogenous shock, so by measuring the time of the second flash crash we probe the systemic nature of flash crash propagation across the fund–asset network. These results are again consistent with expectations and demonstrate that higher values of λ^0 and C^0 destabilise the system, causing flash crashes to occur more rapidly following commencement of distressed agent selling. Increasing the margin tolerance τ_c stabilises the system, leading to larger intervals between shock delivery and flash crash onset.

Our model continues to support the conclusions of models with a similarly-structured macroscopic network component. Our work goes further by allowing detailed consideration of distress propagation speeds and of the intraday mechanism by which contagion occurs. Price impact in real markets does not occur instantaneously, rather it manifests across longer time periods while trading strategies are in operation. The realistic input validation (Section 2.6) of the microstructure component of our model allowed us to show that the market quickly reacts to initial distress, leading to contagious spreading of flash crashes across the network in a matter of minutes. This short time interval places an upper bound on the window for the deployment of *ex post* regulatory interventions. Our results demonstrate that increasing margin tolerance provides the most effective route to increasing the size of the *ex post* intervention window. However, the short timescale associated with this window across large regions of the parameter space suggests that *ex ante* precautions may have higher efficacy than *ex post* alternatives.

3.10 Bipartite Network Topology

Previous works such as [Acemoglu et al. \(2015b\)](#) and [Gai and Kapadia \(2010\)](#) have established the importance of network topology when considering the systemic risk associated with financial distress propagation. As discussed in Chapter 2, [Gai and Kapadia \(2010\)](#) introduce the notion of *robust-yet-fragile* networks in which higher network connectivity is advantageous for stability in the presence of small shocks as banks are able to share losses and avoid default. On the other hand, high interconnectedness becomes problematic during high shock conditions by providing a large number of channels for distress propagation. [Acemoglu et al. \(2015b\)](#) similarly argue that in low-shock regimes, more highly-connected networks are more stable. In this section we explore the extent to which these findings apply to a joint micro-macro model of systemic risk due to overlapping portfolios.

In Section 3.9 we represented the fund–asset allocation network as an Erdős–Rényi random graph ([Newman, 2003](#)). We held fixed both the topology and the graph density which was characterised by the fund’s average asset diversification $\rho = 0.5$. We now discuss systemic risk when network density and topology are varied.

3.10.1 Experiment Parameters

Using the bipartite network generation algorithm, we construct networks in which funds exhibit varying degrees of asset diversification (controlled by parameter ρ) and also varying degrees of preferential asset attachment (controlled by parameter β).

As in Section 3.9 where we looked at leverage dependence, we continue with a 50-asset, 50-fund model¹⁰. We perform a combinatorial set of evaluations for $\rho \in [0, 1]$ and $\beta \in \{-7.5, -2, 0, 2, 7.5\}$ yielding a total of 35 parameter pairs. Each pair was evaluated over 100 Monte-Carlo trials in which only the pseudorandom number generator seed varied between simulation instances.

Based on the results from Section 3.9, we chose three fixed points in (λ^0, τ_c, C^0) space. These locations are depicted in the upper-right panel of Figure 3.5 and are presented in Table 3.6. We performed a total of 10,500 ABM evaluations, each consisting of $T = 100,000$ time steps, each step with a simulated resolution of $\delta t = 50\text{ms}$.

As in previous sections, we introduce an exogenous shock to a single asset in the form of a distressed selling agent, which places a sequence of orders sufficient to trigger a flash crash in the selected asset. We select the asset to shock uniformly from the

¹⁰The effect of variable network size was found to be similar to the effect of diversification parameter ρ and so we choose to frame our discussion in terms of ρ for the sake of brevity.

	λ^0	τ_c	C^0	Leverage regime	Stability†
θ_1^*	3	1.01	1	High	Typically unstable
θ_2^*	2	1.01	0.5	Medium	Moderately stable
θ_3^*	1.5	1.01	0.25	Low	Typically stable

Table 3.6: Locations in (λ^0, τ_c, C^0) space which are fixed during topology experiments. These points are depicted in Figure 3.5. †We report the stability at $\rho = 0.5, \beta = 0, n_f = n_a = 50$, which is consistent with parameter selection in Section 3.9.

set of assets that are held by at least one fund. In this way, we avoid shocking assets that are disconnected from the network. Formally, let us define the set of connected assets $\mathcal{A}^c = \{j \in (1, \dots, n_a) : \sum_{i=1}^{n_f} A_{ij} > 0\}$, then the probability of selecting an asset $j' \in \mathcal{A}^c$ as the single asset to receive the exogenous shock is equal to $1/|\mathcal{A}^c|$.¹¹

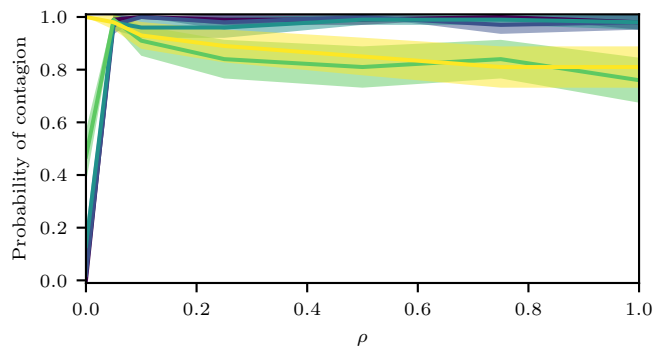
3.10.2 Probability and Extent of Contagion

Figure 3.7 presents the probability (left column) and extent (right column) of contagion as we vary network diversification (ρ) and crowding parameter (β). We present results for three leverage scenarios: high leverage (θ_1^* , top row), medium leverage (θ_2^* , middle row) and low leverage (θ_3^* , bottom row). We find that systemic risk is strongly dependent on each of these system parameters.

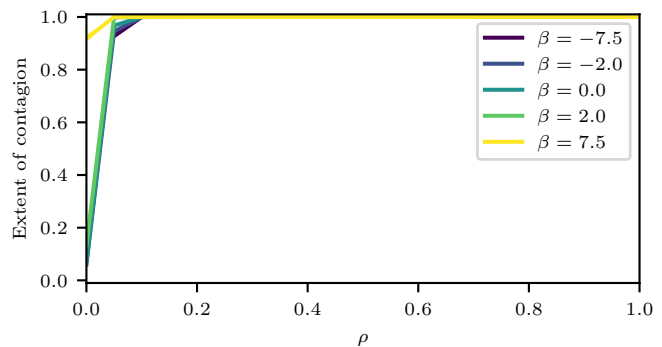
Figure 3.7(a) shows that default cascades occur with high probability for all but the smallest values of diversification parameter ρ . When ρ is less than 3%, funds invest in a single asset¹². At low diversification, we find that the system is most stable for dispersed configurations ($\beta < 0$). This result supports the findings from a purely analytical model by Chen et al. (2014), who point out that in dispersed topologies, each fund invests in a different asset, and so it is unsurprising that distressed selling by one fund cannot affect the portfolios of other funds (see Table 3.7). When diversification is low, we also observe that crowded systems ($\beta > 1$) are the least stable. This corresponds to a topology in which all funds invest in the same single asset. This has two implications — first that a falling price of this asset will simultaneously affect all funds, and second that distressed selling by any fund will necessarily affect the single jointly-held asset. Both factors contribute to increased price impact at the jointly-held asset and hence distress propagation between funds in

¹¹We further require that if more than 80% of funds are connected in a giant network component, then the shock must be delivered to an asset that is also connected to the same component. This is to reduce the impact of spurious outliers and noise in the results and does not affect our conclusions.

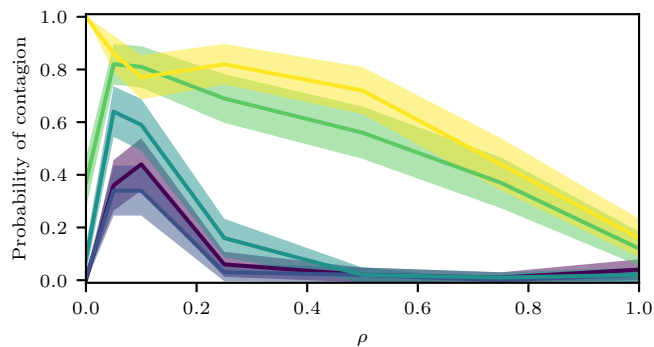
¹²Recall $k_{fund} \approx \max(\rho n_a, 1)$, and $n_a = 50$ in the present experiment.



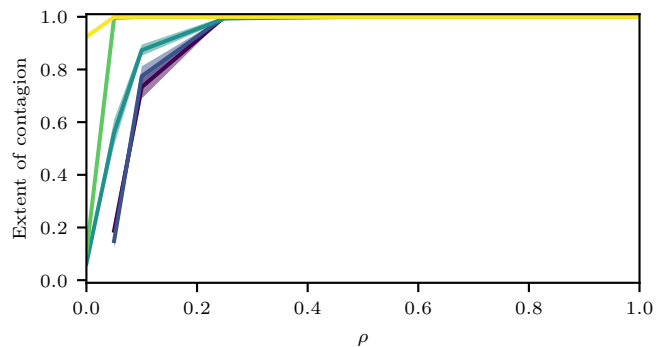
(a) θ_1^* (high leverage)



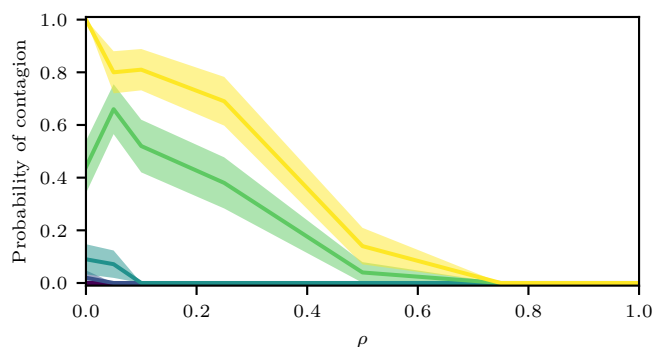
(b) θ_1^* (high leverage)



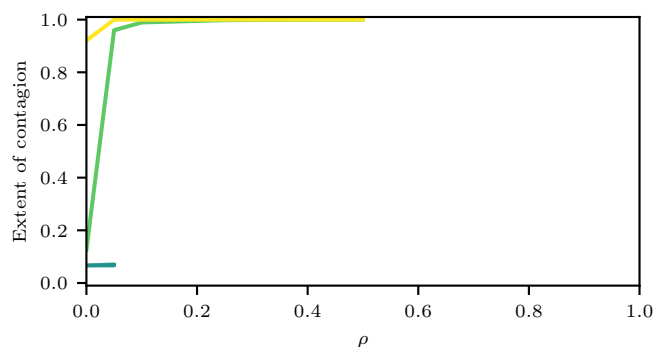
(c) θ_2^* (medium leverage)



(d) θ_2^* (medium leverage)



(e) θ_3^* (low leverage)



(f) θ_3^* (low leverage)

Figure 3.7: Probability (a,c,e) and extent (b,d,f) of a default cascade as a function of portfolio diversification (ρ) and crowding (β) parameters. The analysis is repeated for three leverage regimes denoted $\{\theta_1^*, \theta_2^*$ and $\theta_3^*\}$ in decreasing order of magnitude.

β	$\langle n_f^C \rangle$	95% interval
-7.5	1.52	[1.42, 1.62]
-2.0	2.56	[2.45, 2.67]
0.0	3.80	[3.64, 3.96]
2.0	11.2	[10.6, 11.7]
7.5	45.9	[45.1, 46.7]

Table 3.7: Presenting the mean and 95% confidence interval of n_f^C , the number of funds in the largest connected component of the $n_f = n_a = 50$ fund–asset network when $\rho \rightarrow 0$, i.e., $k_{fund} = 1$, for different choices of crowding parameter β . Statistics are generated from an ensemble of 100 random networks.

the system. As diversification ρ increases beyond 3%, the system becomes less stable for all crowding regimes. In high leverage scenarios, we therefore find that increasing diversification does not improve systemic stability. Rather, increased diversification allows contagion to spread freely to more of the assets and funds in the network. [Acemoglu et al. \(2015b\)](#) find a somewhat analogous effect when comparing the size of shocks delivered to their system. However, a surprising result emerges as crowding (β) increases. It is not obvious what mechanism acts to increase system stability as a function of higher portfolio overlap for moderate and high levels of diversification in this high leverage regime. This finding is not consistent with existing literature, and we return to consider it in detail in Section 3.10.4.¹³

Figure 3.7(c) presents the probability of contagion in a moderate leverage regime. At very low diversification we again observe that the system is most stable for dispersed configurations ($\beta < 0$). We may attribute this finding to the network topology which consists of disjoint sub-graphs in which each fund invests in a single asset, and each asset is held by exactly one fund. We find that cascades cannot propagate, as the network is too sparsely connected, similar to the findings by [Gai and Kapadia \(2010\)](#) and [Caccioli et al. \(2014\)](#). As diversification increases to 5–10% we observe a marked reduction in systemic stability for all crowding regimes. At this level of diversification the network is sufficiently connected for distress to propagate. However, as diversification increases further we find that systemic risk is reduced which is again consistent with [Gai and Kapadia \(2010\)](#) and, in particular, [Caccioli et al. \(2014\)](#), who found that increased diversification means that distress at a fund will be spread

¹³This result is consistent when considering alternative indicators of systemic risk including the drawdown (maximum to minimum price return) of affected assets and the profit and loss of affected funds. In both of these cases the effect of enhanced crowding ($\beta > 0$) for $\rho \geq 0.03$ is to reduce the severity of systemic risk.

between a larger number of assets and so have a smaller effect on any one asset. In the moderate leverage regime, we find that distress at a fund is less likely to cause a flash crash at a given asset. This stands in contrast to the high leverage scenario discussed previously, where even fully diversified funds can still have sufficient price impact to trigger flash crash cascades. Highly-crowded regimes ($\beta > 0$) also benefit from increased diversification, however such regimes are always less stable than non-crowded regimes. This result is to be expected: when portfolios are strongly-overlapping, distress at a single asset affects a large fraction of funds. Distress at a single fund is therefore more directly felt by other funds in crowded regimes.

Figure 3.7(e) presents results when leverage is further decreased. We find that crowded regimes remain the least stable, regardless of the level of portfolio diversification. However, this figure shows that the benefits of portfolio diversification manifest at lower levels of diversification parameter ρ compared to higher leverage scenarios. In particular, no default cascades are observed for $\rho \geq 0.75$. When funds utilise low leverage, they are relatively insensitive to price movements of the assets in their portfolios, and so even modest levels of diversification are sufficient to avoid the need for funds to engage in distressed selling. In the low leverage regime, our results are consistent with further findings by [Chen et al. \(2014\)](#) who report that systemic risk in their purely analytical model is minimised when all funds hold the “market portfolio”, which in our notation corresponds to maximum diversification ($\rho = 1$).

Figures 3.7(b), 3.7(d) and 3.7(f) demonstrate that when default cascades are observed, they typically affect the entire network. That is, the extent of contagion is 100% for anything other than very low levels of portfolio diversification. In the moderate and low leverage regimes the system is *robust-yet-fragile* ([Gai and Kapadia, 2010](#)) for higher levels of diversification. Although default cascades are unlikely in such regimes, when they do occur the distress propagation is extensive. In each leverage regime, at low diversification the extent of contagion is highest for more highly-crowded systems. This effect is most pronounced in the moderate leverage regime. In the low leverage regime we did not observe default cascades for certain parameter combinations, and so the (conditional) extent of contagion is undefined in some cases (this results in partial series, particularly noticeable in Figure 3.7(f)). When diversification is low, dispersed network topologies can become disconnected. Contagion is limited to the connected component (sub-graph) containing the exogenously-shocked asset, resulting in an extent of contagion of less than 100%.

The presence of robust-yet-fragile behaviour creates a problem for policy-makers ([Gai and Kapadia, 2010](#)). Although systemic failure can be made less likely by en-

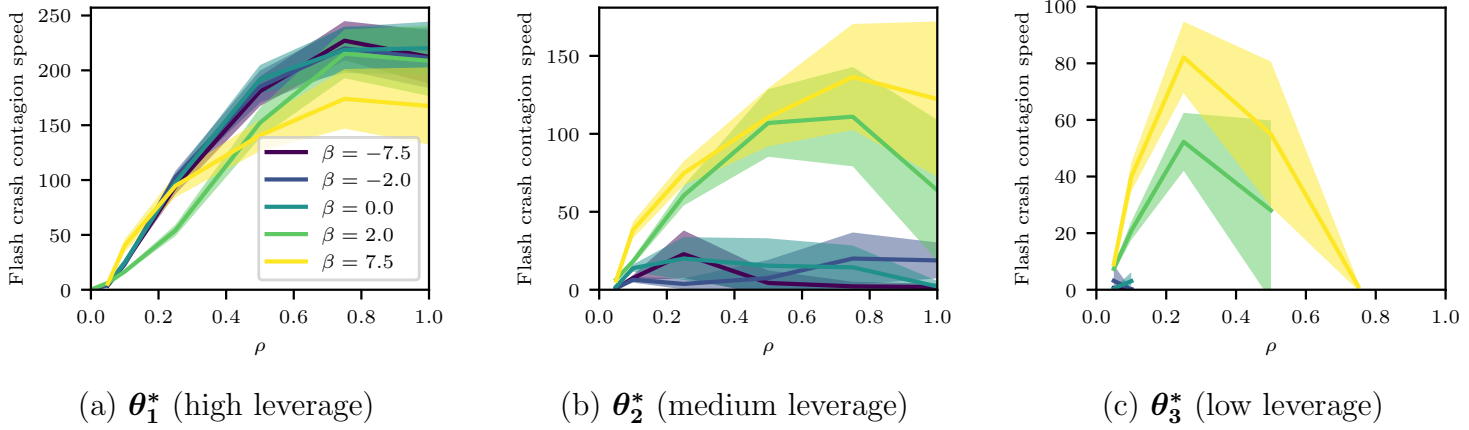


Figure 3.8: Flash crash propagation speed (minutes⁻¹) through the fund–asset network as a function of portfolio diversification (ρ) and crowding (β) parameters. The analysis is repeated for three leverage regimes denoted $\{\theta_1^*, \theta_2^*, \theta_3^*\}$ in decreasing order of magnitude. We plot the mean across 100 Monte-Carlo trials, along with 95% confidence intervals.

couraging higher portfolio diversification at funds, if a cycle of deleveraging takes hold, enhanced diversification means that financial distress spreads more extensively through the network of funds and assets, resulting in increased financial damage to the fund sector overall. In the next section we consider the dynamic nature of the spread of distress and investigate the speed of flash crash propagation as a function of network topology.

3.10.3 Flash Crash Propagation Speed

Figure 3.8 presents the speed of flash crash propagation as a function of fund and asset diversification in the same three leverage regimes described in the previous section. As mentioned previously, and discussed in detail in Section 3.4.3, the speed we report here is inversely related to the time taken for 80% of subsequent flash crashes to occur, following the observation of the first flash crash.

Figure 3.8(a) demonstrates that at high leverage, portfolio crowding makes little difference to the speed at which flash crashes propagate between assets. Instead, the speed strongly depends on network diversification (β , in our notation). The previous section established that under high leverage, enhanced diversification is not beneficial to systemic stability. Figure 3.8(a) shows that enhanced diversification leads to faster flash crash propagation between assets. The apparent stimulating effect of diversification on propagation speed may indicate that if a set of deleveraging funds yields

enough market impact to cause a flash crash in one asset, it likely also wields enough market impact to cause a flash crash in many assets (it is unlikely to have exactly the right level of impact needed to distress one and only one asset) and so enhanced diversification can simply become an enhanced channel for distress propagation. Enhanced diversification also implies that, conditional on entering distress, a fund will send sell orders to a larger fraction of assets in the network. Under high leverage, the resulting price impact causes these assets to undergo flash crashes at similar times during the simulated trading session, resulting in a higher speed measurement. The observed maximum speed of approximately 200 flash crashes per minute corresponds to a time difference of just 5 minutes between the first flash crash and the 80th percentile flash crash. Such a short window places strict constraints on any *ex post* regulatory interventions, and instead indicates that *ex ante* measures may be more useful as policy tools. Calibration of the macroscopic model component against real data would be necessary for a more comprehensive characterisation of the expected efficacy of real-world regulatory interventions.

Figure 3.8(b) presents flash crash propagation speed results in a moderate leverage regime. Here we find that speed remains low regardless of diversification for non-crowded portfolios. In the previous section we observed that default cascades were unlikely for moderate to high diversification for moderate leverage and low crowding (Figure 3.7(c)). However, cascades were observed to occur for low diversification in this regime. The low speed at which flash crashes propagate from asset to asset in this case may be due to the fact that the fund–asset network is only sparsely connected, meaning that the effective network *diameter* (in a graph-theoretic sense) is higher. A higher diameter implies that some assets are more weakly-connected to the network and hence distress takes longer to propagate to such nodes. The results for high-crowding regimes are similar to those for high-crowding regimes at higher leverage, though this observation only holds for low to moderate diversification ρ . When ρ increases beyond approximately 50%, Figure 3.8(b) shows that increased diversification reduces flash crash contagion speed. We note that when portfolios are maximally diversified, there can be limited difference between crowded and non-crowded configurations (we revisit this notion in the next section), and so it is not surprising that the flash crash propagation speed is similar for crowded and non-crowded topologies when $\rho \rightarrow 1$. However, it is not obvious why speed should peak near $\rho = 0.5$. As density increases from $\rho = 0$ to $\rho = 0.5$ in crowded regimes ($\beta > 0$), the number of assets that have non-negligible levels of investment increases. Crucially, however, at low density the effect of this increase is detrimental to systemic stability

(as we saw in the previous section). It is not until density increases further that diversification acts to stabilise lower-leverage systems. It is interesting that the peak of flash crash contagion speed occurs at a higher value of the diversification, ρ , than the peak probability of contagion. This could be due to the conditional nature of the speed measurement. Speed is only defined when more than one asset experiences a flash crash, and this is likely highly correlated with cases where default cascades occur. If we only consider cases where default cascades occur, it does not necessarily follow that the speed, under this condition, will be maximised at the same level of diversification as the unconditional systemic stability.

Figure 3.8(c) presents similar results for a low leverage scenario. Again we see that more crowded systems exhibit higher flash crash propagation speed regardless of portfolio diversification, and that the speed reaches a maximum at an intermediate level of diversification. The figure clearly shows that the maximal propagation speed, for crowded scenarios, occurs at a lower level of diversification coefficient ρ for the low leverage scenario, compared to the moderate leverage scenario. This finding is broadly consistent with Figures 3.7(c) and 3.7(e) which find that probability of contagion reaches a maximum for lower values of ρ for low leverage scenarios compared to moderate leverage scenarios. Since flash crash contagion speed is only defined for Monte Carlo trials where at least two assets experience flash crashes, Figure 3.8(c) features incomplete series¹⁴.

Taken together, Figures 3.8(a)–(c) demonstrate that flash crash propagation speed increases as leverage increases. The results also demonstrate that for high-crowding regimes, speed is a non-monotone function of portfolio diversification. The non-monotonicity is related to similar non-monotonicity observed for the probability of contagion in Section 3.10.2, though the maxima occur at different levels of diversification when considering speed as opposed to stability. We have also found that, similar to the systemic stability results, dispersed configurations and configurations with no crowding have the lowest flash crash propagation speed. This means that they are not only more robust to default cascades, but also that in such scenarios, regulators have a larger time window in which to deploy *ex post* interventions should a deleveraging event occur.

¹⁴We also only report results for Monte Carlo trials with valid distributions. This means that certain parameter settings where only a single Monte Carlo replicate resulted in a valid speed measurement are suppressed as it is not possible to give confidence intervals in such cases.

3.10.4 Effect of Asset Allocation Distribution

In Section 3.10.2 we found the surprising result that, in certain circumstances, increased portfolio crowding is beneficial to systemic stability. We are not aware of previous studies that have encountered this phenomenon. It is not apparent in Caccioli et al. (2014) or Chen et al. (2014). In this section we present evidence in support of the hypothesis that the effect is due to the non-uniformity of fund investments between the assets in their portfolio. We thus establish the importance of considering non-uniform asset allocations in future studies.

With reference to Figure 3.7(a), we consider the probability of contagion in the high leverage scenario θ_1^* . We observe that a higher crowding parameter β results in a reduced probability of contagion for most values of portfolio diversification ρ , and that this effect becomes more pronounced as ρ increases. This is surprising because $\rho \rightarrow 1$ implies that all funds make investments in all assets. The notion of preferentially attaching to a subset of assets therefore has no meaning when funds are simply forced to invest in every asset. This leads us to hypothesise that the non-uniformity of the size of allocations funds make to each asset must have an effect on stability. This would be a potentially important result, because a common assumption in the literature is that funds make equal investments in each of the assets in their portfolio, e.g., (Acemoglu et al., 2015b; Caccioli et al., 2014).

In order to test this hypothesis, we introduce a parameter α that we use to interpolate between funds holding uniformly-allocated portfolios ($\alpha \rightarrow 1$) or LogNormal-allocated portfolios ($\alpha \rightarrow 0$). All previous results in the present chapter set $\alpha = 0$. Figure 3.9 presents the probability of contagion in the high leverage scenario, at maximum diversification ($\rho = 1$ such that all funds invest in all assets), as we vary crowding parameter $\beta \in \{-7.5, 0, 7.5\}$ and interpolate $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$. For each (α, β) parameter pair we performed 100 Monte-Carlo trials, and report the mean and 95% confidence interval from the resulting distribution of default cascade occurrence. As hypothesised, we see that the probability of contagion for different levels of crowding converges as we interpolate from LogNormal to uniform portfolios. This confirms that the surprisingly beneficial impact of increased portfolio crowding in high-leverage scenarios is indeed a result of the non-uniformity of asset allocations in our model.

Although we have established a connection between portfolio allocation uniformity and systemic risk, we now consider the reason that this connection exists. We

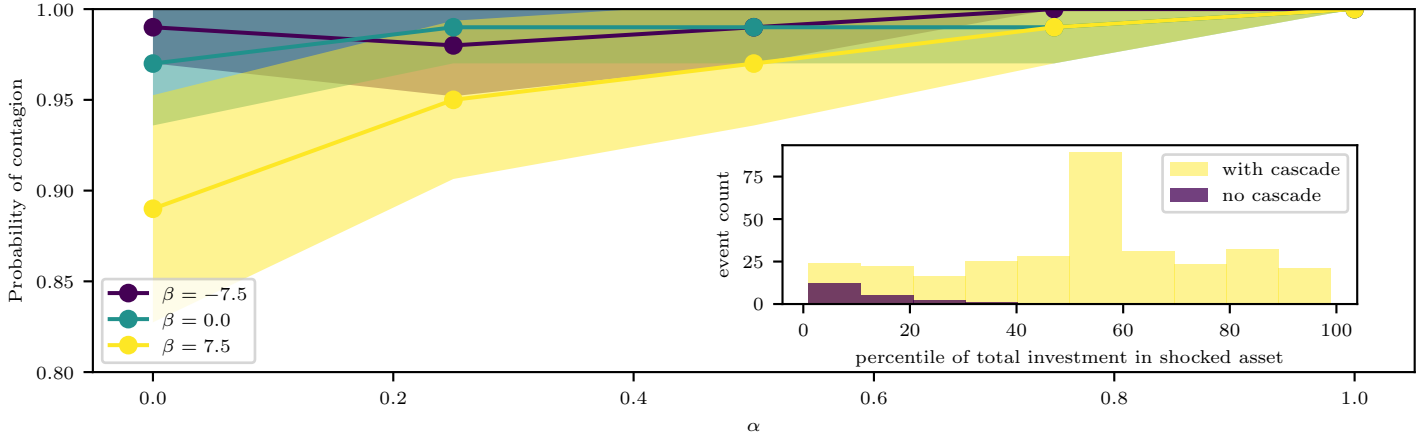


Figure 3.9: Probability of contagion with 95% confidence interval as we interpolate between LogNormal-distributed and uniformly-distributed asset allocations for each fund. $\alpha \rightarrow 1$ implies funds allocate their capital *equally* to each asset in which they invest. We fix $\rho = 1$ here, such that each fund invests in every asset. This study was carried out in the high-leverage scenario, θ_1^* . Inset: The number of Monte Carlo trials exhibiting a default cascade as a function of the total level of investment in the asset that receives the price shock, for the highly-crowded $\beta = 7.5$ case.

hypothesise that the reduced probability of contagion for non-uniform asset allocations in high-crowding compared to low-crowding scenarios is due to the presence of weakly-connected assets in the fund–asset network.

As described in Section 3.3.3, when funds allocate their investments to assets during network construction, they do so in decreasing investment size order. Since the preferential attachment process selects investment targets without replacement, this means that in crowded scenarios ($\beta > 0$) funds will tend to allocate their largest investment to the same asset as have other funds. This leads to dispersion in the overall amount invested in each asset. As a consequence, some assets receive significantly higher levels of total investment than others. We might expect that assets with lower levels of total investment are, *ceteris paribus*, less systemically important than assets with higher levels of investment. This conjecture must hold in the limit $\sum_{j=1}^{n_f} A_{ij} \rightarrow 0$ for some asset $i \in (1, \dots, n_a)$ since by definition such an asset is disconnected from the network.

The inset graph of Figure 3.9 presents the number of Monte Carlo trials exhibiting a default cascade as a function of the total level of investment in the asset that receives the price shock, for the highly-crowded $\beta = 7.5$ case, for all values of α . We rank assets according to their total level of investment and report the normalised rank (percentile) of the asset that receives the initial shock. The figure clearly shows that,

when no default cascade occurs (dark shaded bars), the shocked asset did indeed have a low level of total investment compared to other assets. This strongly supports our hypothesis that the surprising benefit of portfolio crowding on systemic stability in high leverage, high diversification scenarios is due to the presence of *weakly-connected* assets. This, in turn, establishes the importance of considering non-uniform asset allocations in future studies of such phenomena. Figure 3.9 (inset) also demonstrates that the converse does not hold — default cascades frequently occur even when the shocked asset is weakly-connected (light shaded bars). This indicates that the emergent system dynamics are not simply reducible to a statement about the level of investment in a given asset. This finding also has important regulatory consequences. We find that regulators cannot simply ignore the effect of weakly-connected assets when considering systemic stability, as shocks to such assets can readily trigger catastrophic systemic failure.

3.11 Conclusions

The results in this chapter contribute to the rapidly growing literature on systemic risk. We provide a significant early bridge between the macroscopic domain of networks of shared asset holdings and the microscopic domain of algorithmic trading and flash crashes. This study has shown that the detailed (microscopic) behaviour of automated trading strategies has deep implications for (macroscopic) systemic risk. Our results will be of interest to systemic risk researchers in academia, as well as to regulators who desire enhanced tools to assess the stability of global financial systems.

The present study has been one of the first attempts to augment models of shared asset financial contagion with limit order book driven price impact. Our results lend support to key conclusions from previous macroscopic models, including the destabilising effects of leverage, capital and portfolio crowding. Our model also reproduces expected robust-yet-fragile behaviour where systemic collapse is unlikely, but extensive when it occurs.

Our key findings are:

1. Leverage management practice by investment funds engenders a contagion channel for flash crashes to propagate between algorithmically traded assets.
2. The speed of flash crash propagation strongly depends on the amount of leverage and capital deployed by funds and on network topology, but the dependence is not monotonic in high-crowding scenarios.

3. Crowding can, under some circumstances, be *beneficial* to systemic stability.

The validation of the model against realistic behavioural timescales allowed us to move beyond previous macroscopic studies, and to investigate the speed at which distress propagates through the financial network during the trading day. We used this approach to characterise the time window available for regulatory interventions, with results suggesting *ex ante* precautions may have higher efficacy than *ex post* reactions.

We found that financial network structure and leverage management policies each have critical impact on both systemic stability and distress propagation speed. Surprisingly, we found that in high-leverage scenarios, under certain conditions, increased portfolio crowding *enhances* systemic stability. We connected this result to the presence of weakly-connected asset nodes in the financial network. Such circumstances do not occur in previous models that conventionally assume uniform portfolio allocations.

Finally, our results provide new insights into the role played by algorithmic trading strategies in the propagation of systemic risk. We find that leverage management policies by funds with overlapping portfolios provide a sufficient contagion channel for the propagation of flash crashes between algorithmically-traded assets. This contagion channel has received little attention in the flash crash simulation literature. Our findings highlight the need for future systemic risk models to move beyond long-established non-crisis price impact assumptions. The path-dependent nature of liquidity provision and flash crashes in asset markets strongly suggests that agent-based methodology is more appropriate than closed-form analysis for future studies.

In the following chapter, we put our results on a firm quantitative footing by bringing real-world data into both the macroscopic, and microscopic, model components. This undertaking is particularly important if recommendations arising from the model are to be useful to regulators and policy-makers. Many of the design decisions we have taken in producing our model framework are intended to facilitate these calibration and validation exercises, and the model has been developed with extensibility and customisation in mind. At the micro level, the robustness of the microstructure dynamics under varying parametrisation (Section 3.3.4) affords the possibility that the model can be meaningfully calibrated to different markets, asset classes and time periods, while retaining key qualitative features. At the macro level, the fund–asset network is in principle available from public 13-F filings (in the USA) though the size of this dataset demands careful handling in order to keep simulations

tractable. Although the simulation has been optimised for a state-of-the-art HPC cluster, scaling up from tens to tens of thousands of assets represents a significant challenge. With a successfully calibrated and validated model available, the true potential of ABM is unlocked as a tool for *in silico* analysis of “what-if” policy scenarios, which we consider in Chapter 5.

“The validation or estimation of agent based models of financial markets by means of simulation methods based on actual data leads to a highly complex optimization problem.”

— Gilli and Winker (2003, p.299)

4

Calibration and Validation

IN CHAPTER 3 we demonstrated the ability of the market microstructure component of our model to produce output that is qualitatively representative of the real-world. We described this procedure as *empirical validation*. In this chapter we revisit the validation of the model, and subject the model to comprehensive *quantitative* calibration and validation procedures. Our aim is to determine whether the model can, under some suitable parametrisation determined through calibration, give rise to behaviour that reflects that of the real-world with a high degree of statistical confidence. There are many methodological considerations inherent in this process. How are we to measure how close the ABM behaviour is to the real-world? How close is close enough? What real-world and simulated data are required? What measurements should be made on real-world data? How should the parameters of the ABM be identified? How should we deal with the complex stochastic and emergent nature of ABM dynamics? How robust are the resulting models out-of-sample?

These questions arise as a natural consequence of the richness and expressiveness of ABMs and the complexity of the systems they represent. As discussed in Chapter 2, ABMs are not amenable to closed-form analysis, and instead we must resort to repeated Monte Carlo simulations in order for their behaviour to be characterised. The output of such a Monte Carlo experiment is, by definition, distributional. Researchers are therefore required to invoke the tools of statistical analysis to describe the features and behaviour of their models.

The main contribution of this chapter is to progress the model presented in Chapter 3 from Level 1 to Level 2 in the taxonomy of [Axtell and Epstein \(1994\)](#) (section 2.6). In addition to this, we also demonstrate the ability of the model to generalise by representing asset classes beyond that for which it was originally conceived. This adds evidence in favour of the generality of our approach, and suggests that extensions to further related asset classes may be feasible.

The chapter proceeds as follows. In Section 4.1 we present the theoretical foundations of calibration and validation, and introduce notation. In Section 4.2 we describe our data sources and our approach to using these data for calibration and validation, and present the input validation for our model, and in Section 4.3 we discuss the measurements we will make of the model output. We perform comprehensive initial transient and global sensitivity analyses in Sections 4.4 and 4.5. In Section 4.6 we quantitatively compare a set of competing calibration methods, and offer guidance on method selection to future ABM practitioners. In Section 4.7 we calibrate and quantitatively validate our model for an extensive range of dates, assets, and model output features. In Section 4.8 we test the out-of-sample performance of the calibrated and validated model, and in Section 4.9 we analyse the parameters obtained through calibration. Finally, in Section 4.10 we investigate whether, and under what circumstances, the calibrated and validated model is capable of exhibiting flash crashes.

4.1 ABM Calibration and Validation

The problems of calibration and validation are amongst the most fundamental in the ABM literature, and the community is yet to reach a consensus on what it means for a model to be validated, calibrated or estimated. Indeed, different authors ascribe different meaning to these terms, or interchange them freely, for example, [Recchioni et al. \(2015, p.2\)](#) state “the aim of our paper is the development of a rigorous calibration procedure designed to validate agent-based models. To this end, we undertake the estimation of the parameters of the agent based model ...”. In order to avoid ambiguity, and to assist with the process of consensus building within the ABM community regarding calibration and validation, we adopt the terminology arising from the body of work comprising [Fagiolo et al. \(2007\)](#), [Fagiolo et al. \(2019\)](#), [Grazzini and Richiardi \(2015\)](#) and [Gatti et al. \(2018\)](#), described in the next sections.

4.1.1 Preliminary Definitions

Gatti et al. (2018, p.165) present the methodological basis for empirical validation of ABMs in terms of a comparison between a *real-world data-generating process* (rwDGP) and the *model data-generating process* (mDGP). The rwDGP is described as representing the causal mechanism giving rise to real-world observables. The mDGP represents the simplified causal mechanism encapsulated in the ABM. We adapt the terminology presented in Gatti et al. (2018), which is described below. We generalise the notation to incorporate all stateful model entities, including those that lack the capability for autonomous action that would be necessary for them to be described as *agents* (Wooldridge, 2001). Such entities may, nevertheless, maintain (possibly dynamic and/or stochastic) properties that evolve according to (non-behavioural) rules. In our model, asset prices emerge from the order placement behaviour of agents, mediated through a double-auction process. The state of the auction is represented by the limit order book (Chapter 2), and asset prices are made available to all market participants via their views onto this order book. The order book maintains a dynamic state, and updates this state according to rules that are a deterministic function of agent activities. The order book itself cannot, however, be described as possessing *agency*. Rather, the order book represents part of the *environment* in which agents operate in our model. A complete description of an ABM must incorporate a description of all entities of the model, including agents and their environment. For the remainder of this discussion we widen Gatti et al. (2018)'s use of the term *agent* in their Section 3.3, by using the more generic term *entity*.

The state of an entity $i \in (1, \dots, N)$ is fully described at time t by a vector of state variables $\mathbf{x}_{i,t} \in \mathbb{R}^k$. Gatti et al. (2018) present the evolution of entity state variables as the difference equation

$$\mathbf{x}_{i,t+1} = \mathbf{f}_i(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t}, \boldsymbol{\theta}_i, \boldsymbol{\xi}_{i,t}), \quad (4.1)$$

where $\mathbf{x}_{-i,t}$ are the states of all other entities, $\boldsymbol{\theta}_i$ is a (possibly entity-specific) parameter vector in the D -dimensional parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^D$, $\boldsymbol{\xi}_{i,t}$ are stochastic terms and $\mathbf{f}_i$ represents (possibly entity-specific) behavioural or mechanistic rules. Equation 4.1 thus fully-specifies the mDGP (Gatti et al., 2018).

The state of the mDGP can, for all times t , be described by the matrix of individual entity states $\mathbf{X}_t = (\mathbf{x}_{i,t})$, and thus Equation 4.1 can be rewritten as

$$\mathbf{X}_{t+1} = \mathbf{F}(\mathbf{X}_t, \boldsymbol{\theta}, \boldsymbol{\Xi}_t), \quad (4.2)$$

where $\boldsymbol{\Xi}_t$ represents all stochastic terms at time t . In addition to the properties of entities themselves, we will find it useful to measure aggregate properties emerging from the system dynamics, for example asset prices. Continuing with the notation presented by Gatti et al. (2018), we define an aggregate variable at time t , as

$$\mathbf{y}_t = \mathbf{m}(\mathbf{X}_t, \boldsymbol{\kappa}_t), \quad (4.3)$$

where $\mathbf{m}$ is a measurement function and $\boldsymbol{\kappa}_t$ are further stochastic terms representing, for example, measurement uncertainty. Equations 4.2 and 4.3 can be simplified by considering the practical method by which stochastic variates are generated in a computer implementation of an ABM. As discussed in Gatti et al. (2018), digital computers typically generate *pseudorandom* numbers, rather than truly random numbers¹. Pseudorandom number generators (PRNGs) produce high-entropy deterministic sequences parametrised by an initial *seed*, s . If two PRNGs of the same type (for example, our model uses the popular Mersenne Twister algorithm from Matsumoto and Nishimura (1998)) share identical seeds, they will produce identical sequences. This feature facilitates reproducible behaviour in the ABM. We can therefore replace the stochastic terms in Equations 4.2 and 4.3 with the parameter s , giving

$$\mathbf{X}_{t+1} = \mathbf{F}(\mathbf{X}_t, \boldsymbol{\theta}, s), \quad (4.4)$$

$$\mathbf{y}_t = \mathbf{m}(\mathbf{X}_t, s), \quad (4.5)$$

If we treat the PRNG seed as an extra initial condition to the model, $\mathbf{Z}_0 = \{\mathbf{X}_0, s\}$, then we may rewrite Equations 4.4 and 4.5 by recursively substituting $\mathbf{X}_{t+1}$ with $\mathbf{X}_t$, thus

$$\mathbf{X}_t = \mathbf{F}(\mathbf{F}(\cdots \mathbf{F}(\mathbf{Z}_0, \boldsymbol{\theta}) \cdots)) = \mathbf{F}^t(\mathbf{Z}_0, \boldsymbol{\theta}) \quad (4.6)$$

$$\mathbf{y}_t = \mathbf{m}(\mathbf{F}^t(\mathbf{Z}_0, \boldsymbol{\theta})) = \mathbf{g}^t(\mathbf{Z}_0, \boldsymbol{\theta}) \quad (4.7)$$

¹Custom hardware is necessary if true random numbers are required, for example, in specialist cryptographic applications.

Equation 4.7 is the *input-output transformation* (IOT) function of the ABM, which deterministically maps every input to a possibly non-unique output (Grazzini and Richiardi, 2015; Gatti et al., 2018).

With this notation established, we are now able to give formal definitions for the terms validation (both input, and output validation), and calibration.

4.1.2 Input Validation

Gatti et al. (2018, p.169) state that input validation “consists of checking whether the building blocks of the model and its assumptions — concerning for instance agents’ behaviours and interactions — are in line with the available evidence”. This evidence can be generated from expert testimony (of the kind undertaken by US regulators in the aftermath of the 2010 Flash Crash (CFTC and SEC, 2010)) or from experiments with human participants, as is popular in the behavioural finance literature (Lo, 2004).

In terms of the notation introduced in the preceding section, input validation consists of ensuring that the ABM mechanistic and behavioural rules $\mathbf{F}$ align as closely as possible to empirical observations. There is no established convention as to how this task should be undertaken, beyond the application of *Occam’s Razor*.

Input validation also requires that, where possible, initial conditions $\mathbf{X}_0$ and parameters of the model $\boldsymbol{\theta}$ map directly to real-world observables. For example, in Paddrik et al. (2012), ABM parameters governing the number of each class of agent were fixed according to the proportions of participants observed to be active in financial markets during the 2010 Flash Crash (Kirilenko et al., 2017).

4.1.3 Output Validation

Output validation consists of asserting that the model output matrix $\mathbf{Y} = (\mathbf{y}_t), t = (0, \dots, T)$, is *sufficiently* close to the real-world output $\mathbf{Y}_R$, such that we can conclude that the mDGP is representative of the rwDGP. Output validation can be either qualitative or quantitative, corresponding respectively to Levels 1 and 2 of Axtell and Epstein (1994)’s taxonomy (Section 2.6).

There are various approaches to qualitative output validation. In Chapter 3 we made use of *indirect inference* (Fagiolo et al., 2007) when we asserted that certain statistical properties of the model output $\mathbf{Y}$ were in qualitative agreement with statistical regularities observed in the real-world; i.e., the stylised facts. For example, the excess kurtosis (fat tails) of the distribution of price movements is a well-established

empirical feature (Mandelbrot, 1963; Cont, 2001), as is clustered volatility (Kirman and Teyssiere, 2002).

In order to move to a *quantitative* output validation, it is necessary to define a distance, or *loss* function $\mathcal{L}(\mathbf{Y}, \mathbf{Y}_R)$ which attains its minimum value of zero only when $\mathbf{Y} = \mathbf{Y}_R$. To measure $\mathcal{L}(\mathbf{Y}, \mathbf{Y}_R)$, it is often necessary to consider the longitudinal output of the ABM, as opposed to observation at a single time $\mathbf{y}_T$. If the ABM is *ergodic* and *stationary*, then a single time-step observation may be sufficient, provided enough time has elapsed for the system to arrive at its unique absorbing equilibrium (Gatti et al., 2018). However, our model is not ergodic (simulations are stochastic, and asset price does not converge to a single value as $T \rightarrow \infty$) and it is not stationary (for example, the model exhibits volatility autocorrelation and so a subset of the output from $t_1 : t_2$ is not representative of the entire output). For our model, as in the general case, it is necessary to include the simulated output over several, possibly all, time steps. A well-established approach to making this procedure tractable is to define $\mathcal{L}$ in terms of longitudinal *moments* corresponding to the stylised facts of interest. This approach is known as the *method of simulated moments* (MSM) (McFadden, 1989; Gatti et al., 2018).

Let us represent an ABM output *moment* function as $\mu(\mathbf{Y})$, and the vector of all such moments as $\boldsymbol{\mu}(\mathbf{Y})$. We discuss the specific moments that are of interest to the present chapter in Section 4.3. As mentioned by Gatti et al. (2018), a popular choice for $\mathcal{L}$ is the quadratic loss

$$\mathcal{L}(\mathbf{Y}, \mathbf{Y}_R) = (\boldsymbol{\mu}(\mathbf{Y}) - \boldsymbol{\mu}(\mathbf{Y}_R))^2 \quad (4.8)$$

This choice is justified by the non-negativity of the function's range, and the function's symmetry with respect to values of $\boldsymbol{\mu}(\mathbf{Y})$ that are greater than, or less than, $\boldsymbol{\mu}(\mathbf{Y}_R)$ (Bishop, 2006). By substituting Equation 4.7, Equation 4.8 is usually presented in terms of parameter vector $\boldsymbol{\theta}$ and pseudorandom number seed s as

$$\mathcal{L}(\boldsymbol{\theta}, s) = (\boldsymbol{\mu}(\boldsymbol{\theta}, s) - \boldsymbol{\mu}_R)^\top \mathbf{W} (\boldsymbol{\mu}(\boldsymbol{\theta}, s) - \boldsymbol{\mu}_R), \quad (4.9)$$

where $\mathbf{W}$ is a weight matrix, $\boldsymbol{\mu}(\boldsymbol{\theta}, s)$ is the vector of simulated moments generated by the ABM from parameters $\boldsymbol{\theta}$ and pseudorandom number seed s , and $\boldsymbol{\mu}_R = \boldsymbol{\mu}(\mathbf{Y}_R)$ is the vector of empirical moments. It is typical to use elements from the inverse variance-covariance matrix of empirical moments for $\mathbf{W}$ (Gourieroux et al., 1993).

The effect of this choice for $\mathbf{W}$ is to normalise the moment variances so that each moment is treated with equal importance in the loss function. Initial model conditions $\mathbf{X}_0$ are held fixed for all instantiations of the ABM (otherwise the members of $\mathbf{X}_0$ would more properly be described as parameters) and hence $\mathcal{L}$ is usually not written as a function of the initial conditions.

Due to the stochastic nature of each run of the ABM, it is not possible to uniquely characterise $\mathcal{L}(\boldsymbol{\theta})$ from a single ABM run. We must therefore perform multiple runs of the ABM in order to estimate the loss

$$\hat{\mathcal{L}}(\boldsymbol{\theta}) = \frac{1}{I} \sum_{i=1}^I \mathcal{L}(\boldsymbol{\theta}, s_i), \quad (4.10)$$

where I is a hyperparameter controlling the number of Monte Carlo replicates used to estimate $\hat{\mathcal{L}}(\boldsymbol{\theta})$ and s_i is the i th pseudorandom number seed, which in practice would be unique for all instantiations of the ABM. As pointed out by [Gilli and Winker \(2003, p.311\)](#), increasing I “reduces the variance of the Monte Carlo approximation of the objective function, making it smoother and thus reducing the depth of local minima due to this variance.”

Thus defined, $\hat{\mathcal{L}}(\boldsymbol{\theta})$ lends itself readily to comparison between competing models, a feature which is used extensively in calibration, as described in Section 4.1.4. However, for the purpose of model validation, it is not sufficient to simply specify the form of $\hat{\mathcal{L}}(\boldsymbol{\theta})$. Instead, it is additionally necessary to place a *threshold*, τ , on $\hat{\mathcal{L}}(\boldsymbol{\theta})$, such that we can claim a positive validation for the ABM under parameter vector $\boldsymbol{\theta}$ if $\hat{\mathcal{L}}(\boldsymbol{\theta}) < \tau$. [Franke and Westerhoff \(2012\)](#) offer a principled choice for τ based on the empirical loss distribution, calculated by comparing empirical moments to the estimated empirical moments (i.e., the mean of the empirical moment distributions). Studies using a small number of assets (it is typical to use a single asset, see Table 4.1) such as [Franke and Westerhoff \(2012\)](#) must resort to block bootstrapping to estimate empirical distributions. However, if we have access to N realisations of the rwDGP, for example for multiple assets across multiple dates, as in our case, then it is possible instead to directly estimate

$$\hat{\boldsymbol{\mu}}(\mathbf{Y}_R) = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\mu}(\mathbf{Y}_R^i), \quad (4.11)$$

where $\boldsymbol{\mu}(\mathbf{Y}_R^i)$ is the vector of longitudinal moments generated from real-world output $\mathbf{Y}_R^i, i \in (1, \dots, N)$. Then we can calculate the empirical loss distribution as the set

$$\mathcal{P}_{\mathcal{L}} = \{(\boldsymbol{\mu}(\mathbf{Y}_R^i) - \hat{\boldsymbol{\mu}}(\mathbf{Y}_R))^T \mathbf{W}(\boldsymbol{\mu}(\mathbf{Y}_R^i) - \hat{\boldsymbol{\mu}}(\mathbf{Y}_R))\}, i \in (1, \dots, N). \quad (4.12)$$

Franke and Westerhoff (2012) then define τ as the 95th percentile of the distribution defined by $\mathcal{P}_{\mathcal{L}}$. Parameter vectors $\boldsymbol{\theta}$ giving rise to $\hat{\mathcal{L}}(\boldsymbol{\theta}) < \tau$ cannot be rejected as compatible with the empirical loss values generated by the rwDGP at the 95% (two standard deviation) level of confidence. To accommodate the stochastic nature of the calculation of $\hat{\mathcal{L}}(\boldsymbol{\theta})$, Franke and Westerhoff (2012) perform repeated measurements of $\mathcal{L}(\boldsymbol{\theta}, s)$, varying only the random seed s . Positive validation is then defined as the case where more than 5% of these replicates result in $\mathcal{L}(\boldsymbol{\theta}, s) < \tau$. While principled, this validation condition was found to be very permissive compared to other approaches we trialled, and indeed Franke and Westerhoff (2012) caution against inferring too much from it. However, Franke and Westerhoff (2012) go on to introduce a significantly more stringent test of validation, based on a similar principle — the *moment coverage ratio* (MCR) (Franke and Westerhoff, 2012; Braun-Munzinger et al., 2018; Kårvik et al., 2018).

Consider again the N realisations of the rwDGP introduced above. For each moment μ we can define an empirical distribution of the values of μ for each real-world observation $\mathbf{Y}_R^i, i \in (1, \dots, N)$, thus

$$\mathcal{P}_{\mu} = \{\mu(\mathbf{Y}_R^i)\}, i \in (1, \dots, N), \quad (4.13)$$

and hence we can take the 5th and 95th percentiles of $\mathcal{P}_{\mu}$ to define an empirical confidence interval for μ , let us denote this (μ_-, μ_+) . ABM instantiations giving rise to moments $\mu_- < \mu(\boldsymbol{\theta}) < \mu_+$ cannot be rejected as compatible with the moment's empirical confidence interval. The moment coverage ratio is defined as the fraction of I replications for which $\mu_- < \mu(\boldsymbol{\theta}) < \mu_+$, with a positive validation defined as the case where more than 5% of replicates results in $\mu_- < \mu(\boldsymbol{\theta}) < \mu_+$. In effect, Franke and Westerhoff (2012) define an alternative loss function

$$\mathcal{L}_{\mu}^{MCR}(\boldsymbol{\theta}) = \frac{1}{I} \sum_{i=1}^I \mathbb{I}_{\mu}(\boldsymbol{\theta}), \quad (4.14)$$

where $\mathbb{I}_{\mu}(\boldsymbol{\theta})$ is an indicator variable given by

$$\mathbb{I}_{\mu}(\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } \mu_- < \mu(\boldsymbol{\theta}) < \mu_+, \\ 1 & \text{otherwise.} \end{cases} \quad (4.15)$$

The moment coverage ratio can be evaluated for a single moment, as in Equation 4.14, but a more powerful validation test is provided by the *joint* moment coverage ratio (jMCR) which is defined as the case where all moments are *jointly* within their empirical confidence intervals, thus for *vector* $\boldsymbol{\mu}$

$$\mathcal{L}_{\boldsymbol{\mu}}^{jMCR}(\boldsymbol{\theta}) = \frac{1}{I} \sum_{i=1}^I \mathbb{I}_{\boldsymbol{\mu}}(\boldsymbol{\theta}), \quad (4.16)$$

where

$$\mathbb{I}_{\boldsymbol{\mu}}(\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } \boldsymbol{\mu}_- < \boldsymbol{\mu}(\boldsymbol{\theta}) < \boldsymbol{\mu}_+, \forall \mu \in \boldsymbol{\mu}, \\ 1 & \text{otherwise.} \end{cases} \quad (4.17)$$

Positive validation is then defined by Franke and Westerhoff (2012) as the case where $\mathcal{L}_{\boldsymbol{\mu}}^{jMCR}(\boldsymbol{\theta}) < 1 - \tau'$, with τ' set at 5%. In other words, the ABM output for parameter vector $\boldsymbol{\theta}$ is rejected as compatible with the rwDGP if there is a probability of less than 5% that the simulated moments will jointly fall within all empirical moment confidence intervals².

However, due to the expense of evaluating the ABM for large I , it is necessary to place confidence intervals around our measurement of $\mathcal{L}_{\boldsymbol{\mu}}^{jMCR}(\boldsymbol{\theta})$. We extend the methodology of Franke and Westerhoff (2012) by relaxing their implicit assumption of $I \rightarrow \infty$. In fact, it is straightforward to calculate confidence intervals around the estimate of $\mathcal{L}_{\boldsymbol{\mu}}^{jMCR}(\boldsymbol{\theta})$ for small I , and Brown et al. (2001) recommend the Wilson interval (Wilson, 1927) for such cases. In the remainder of this chapter, our validation condition, based on the joint moment coverage ratio, is that ABM output is rejected as compatible with the real-world data generating process if jMCR is below 5% with 2σ (95%) confidence, calculated using Wilson's correction.

²Franke and Westerhoff (2012) further argue that the joint MCR should be scaled according to the number of independent moments. In our case we have 10 moments (see Section 4.2), but conservatively we might assume that 7 are independent if we treat moments representing longer lags of autocorrelation functions as conditionally dependent. Franke and Westerhoff (2012) point out that, by construction, we would expect $1 - 0.95^7 \approx 30\%$ of Monte Carlo samples for a perfectly calibrated model to fall outside the empirical 95% moment confidence interval.

4.1.4 Calibration

With the preceding definition of model validation established, it is straightforward to define what is meant by *calibration* (Gatti et al., 2018).

Calibration is the process of locating parameters $\hat{\theta}$ in order to minimise the value of the selected loss function. Formally,

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \hat{\mathcal{L}}(\theta), \quad (4.18)$$

where θ is a parameter vector in the D -dimensional parameter space $\Theta \subset \mathbb{R}^D$. Even if $\Theta \subset \mathbb{N}^D$, for computational reasons it is not typically practical to evaluate all possible parameter vectors in the full feasible space Θ , and so calibration is the process by which candidate solutions converge to $\hat{\theta}$.

Two observations immediately follow from this definition. Firstly, *any* ABM can in-principle be calibrated, because there will exist some parameter vector θ that minimises $\hat{\mathcal{L}}(\theta)$ (though it is of course not guaranteed that a discovered minimum is the global minimum, or even that a unique global minimum exists). Computational constraints may (and in general, will) prevent the exhaustive evaluation of every parameter vector. As such, calibration can be viewed as a *bandit problem* (Robbins, 1952) where a trade-off exists between exploration of new regions of Θ and exploitation of previously sampled regions of Θ . Thus calibration is seen to be nothing more than an optimisation problem, and the vast array of methods in the optimiser's tool-kit can potentially be brought to bear, subject to computational and other constraints.

A second observation is that, as previously noted, model validation requires the specification of both a loss function $\hat{\mathcal{L}}(\theta)$ and a threshold τ , below which values of $\hat{\mathcal{L}}(\theta)$ are deemed to be sufficiently low such that model output is acceptably close to real-world observables. Although calibration makes direct use of $\hat{\mathcal{L}}(\theta)$, it does *not* require a threshold τ to be identified, and thus the key difference between calibration and validation emerges. Calibration reduces the distance between simulated and real-world observables, whereas validation determines when this distance is sufficiently small for success to be declared. As a consequence, although any model can be calibrated, it does not necessarily follow that any model can be successfully validated.

It is instructive to give consideration to the *distribution* of parameters obtained through a calibration process. Bayesian approaches to calibration provide such distributions directly, whereas for other methods it may be necessary to repeat the calibration a number of times in order to determine the distribution of parameters.

This is discussed in detail in Section 4.6. If the distribution for a calibrated parameter is indistinguishable from a uniform distribution over the parameter's range, this indicates that the parameter has not been *identified* by the empirical data. It could be that the parameter is irrelevant to the model; that the data are insufficient to meaningfully constrain the parameter; that the parameter's sampled range is insufficiently small (and so the model output does not exhibit sensitivity to the parameter in the sampled range); or that further calibration replications are necessary to establish the parameter distribution with sufficient statistical significance. A one-sided Kolmogorov-Smirnov (K-S) test is a straightforward method by which a calibrated parameter's distribution can be compared to a reference (e.g., uniform) distribution. The consideration of parameter distributions is another manifestation of the progression of ABM onto a more rigorous quantitative basis (Platt, 2019).

The properties that characterise ABMs, including the strong non-linear interactions between agents, agent heterogeneity, emergence, and the presence of tipping points (discontinuous changes between model output from adjacent parameter locations), make calibration of ABMs a challenging prospect (Gualdi et al., 2015). It is not generally possible to represent the ABM's input-output transformation function (Equation 4.7), and hence $\hat{\mathcal{L}}(\theta)$, in closed analytic form. This property precludes the availability of analytic gradients, meaning that powerful techniques for parameter selection such maximum likelihood estimation (MLE) are not generally applicable. Even if gradients were available, the strong non-linear interactions between agents in the model and non-linear dependence on model parameters (which will often take a mixture of continuous and discrete values) means that $\hat{\mathcal{L}}(\theta)$ is very unlikely to be convex. Local minima in the loss function landscape are an important consideration for any ABM calibration attempt. The stochastic nature of ABMs adds further complexity insofar as any evaluation of a parameter vector θ in the ABM must be considered as a sample from the space of possible outcomes. This property can introduce discontinuous differences in $\hat{\mathcal{L}}(\theta)$ for adjacent pairs of parameter vectors, compounding sudden changes in $\hat{\mathcal{L}}(\theta)$ (i.e., tipping points) that ABMs frequently exhibit (Gualdi et al., 2015). Lux and Zwinkels (2018, p.481) offer the following observation on ABM calibration:

“For practical applications, this means that we cannot use the set of convenient gradient-based optimizers that econometricians use for other problems. Otherwise one would almost with certainty end up in some local optimum rather than identifying sensible parameter values. Finding

an appropriate optimisation routine for non-standard problems (such as Nelder-Mead, simulated annealing, genetic algorithms, and others) can become a research topic in itself, and the best choice might be problem-specific.”

In summary, the selection of appropriate calibration methods is constrained by the following properties of $\hat{\mathcal{L}}(\theta)$ and the parameter space. The ideal method:

1. Must support discrete parameters,
2. Cannot assume the availability of analytic gradients for $\hat{\mathcal{L}}(\theta)$,
3. Must tolerate noise in the measurement of $\hat{\mathcal{L}}(\theta)$, and
4. Must not require $\hat{\mathcal{L}}(\theta)$ to be continuous or convex.

4.2 Big Data Approach

As discussed in Chapter 3, our model contributes a novel synthesis of micro-level and macro-level ABMs. We are not aware of any previous attempts to calibrate and validate such a compound model. Due to the diversity of components, each section of the model has its own set of data requirements. Figure 4.1 presents the three different components of the model. These are:

1. The leverage relationships between banks and funds,
2. The asset portfolio holdings of funds,
3. The response of asset markets to trading activity, represented by the microstructure ABM.

In this section, we describe the relationship between each component of the model and the relevant dataset, and how we have accommodated the availability or otherwise of the necessary data into our approach to calibration and validation of the model components.

4.2.1 Macro-Level Component: Fund–Asset Network

Investors in US equities, with more than \$USD 100mn under management, are required under US regulatory law to produce quarterly statements detailing their investments at the asset level. These statements are known as form 13-F filings and are publicly available from the US Securities and Exchange Commission (SEC)’s EDGAR

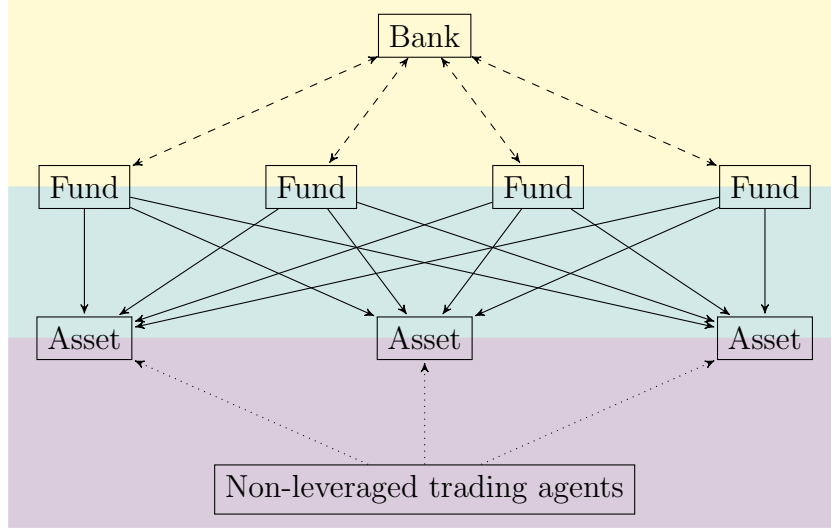


Figure 4.1: Summary of datasets relevant to our model. We enhance Figure 3.1 by adding shading to indicate which dataset is required for the validation of different components of the model. **Top:** Bank–fund leverage data (highly proprietary). **Middle:** Fund–asset network of portfolio holdings. **Bottom:** Market microstructure price and volume data.

portal³. It is possible to access bulk data from EDGAR using web scraping tools (García and Norli, 2012)⁴. We downloaded the full dataset from 2018-04-17, which was the most recent sample available when we commenced research using these data. After cleansing and normalisation⁵, the dataset includes all filing entities ($n_f = 4600$ funds) with investments spanning $n_a = 20100$ unique assets.

We model the 13-F dataset in two ways. Firstly, as a bipartite graph, where an edge between a fund node and an asset node represents an investment (in \$USD) by the fund in the asset (see Section 3.3.3). An important limitation of the 13-F filing process is that only *long* (positive) positions are reported, and so all edge weights in the fund–asset network are positive. Secondly, we can represent the fund–asset network by its weighted adjacency matrix $\mathbf{A} \in \mathbb{R}_+^{n_f \times n_a}$. This representation allows us to make two key measurements. Firstly, the sum of the i th row of $\mathbf{A}$ gives the

³<https://www.sec.gov/edgar/searchedgar/webusers.html>, accessed 2018-04-19. Details on who must file form 13-F are available from <https://www.sec.gov/fast-answers/answers-form13fhtm.html>, accessed 2018-04-19.

⁴We used scraping tools provided by <https://github.com/DZGoldman/13F-Form-Scraper->, accessed 2018-04-18, No license stated.

⁵Before the SEC moved to an XML format in 2015, filing was free text format, which presents difficulties for automated analysis. Examples of problems in the data include inferring whether filings are made in terms of asset quantity or asset value (i.e., the product of quantity and price), and whether the filer reports in \$USD 000’s or \$USD. We created heuristics to tackle these issues as far as possible.

total value π_i of fund i 's portfolio. Collecting the row summations yields a vector $\boldsymbol{\pi}$ of portfolio values for all funds

$$\boldsymbol{\pi} = (\pi_1, \dots, \pi_{n_f}), \text{ where } \pi_i = \sum_{j=1}^{n_a} A_{ij}, \pi_i \in \mathbb{R}_+. \quad (4.19)$$

Secondly, the sum of the elements of the j th column of $\mathbf{A}$ gives the total reported value invested in the j th asset, providing a lower bound for asset j 's market capitalisation. Collecting the column summations yields a vector $\mathbf{v}$ of market cap estimates for all assets

$$\mathbf{v} = (v_1, \dots, v_{n_a}), \text{ where } v_j = \sum_{i=1}^{n_f} A_{ij}, v_j \in \mathbb{R}_+. \quad (4.20)$$

The market cap estimate is subject to certain caveats. As mentioned above, only long (positive) investments are represented in the 13-F data. The dataset also excludes derivative positions such as options, which are often utilised by market participants as a way to obtain exposure to the price movements of an asset without actually owning it (e.g., to avoid tax implications). Finally, the 13-F dataset only contains information on US equity ownership by large institutional managers (with at least \$100mn under management). We follow the approach of [Koijen and Yogo \(2019\)](#) who attribute the difference between $\mathbf{v}$ and the true market cap of assets (obtained from an alternative source) to the “household” sector ([Koijen and Yogo, 2019](#), p.16), who do not engage in sophisticated leveraged portfolio management or algorithmic trading, and so are not relevant to the present investigation. Due to this reason, we do not introduce an extra agent to represent the household sector. Full details regarding the construction of simulated fund portfolios are given in Chapter 5.

The input validation of the fund–asset network has two aspects.

1. Input validation of the system rules and agent behaviours.
2. Fixing initial conditions using empirical data.

The input validation of system rules and agent behaviours for the fund–asset network component of the ABM was presented in detail in Section 3.3. In the present chapter we are able to significantly enhance this input validation by fixing initial conditions using regulatory 13-F data. Since the fund–asset network is fully-specified by the publicly available 13-F data, there are no free parameters associated with this dataset in our model. Calibration of this component is not, therefore, necessary (as there

are no parameters in need of tuning). Input validation is a matter of instantiating entities in our simulation framework to represent the ~ 5000 funds and ~ 20000 assets in the 13-F network. Making a simulation of this scale tractable was a technical, rather than conceptual, challenge (Section 5.1.3). Chapter 5 presents a comprehensive investigation into the properties of the fund–asset network under our simulation framework. However, the behaviour of the full-scale simulation is critically dependent on the response of assets to the distressed selling behaviour by funds. This aspect of the simulation, discussed in Section 4.2.3, requires a complex calibration procedure using fine-grained market data, that results in the quantitative output validation of the microstructure component of our ABM.

It is not meaningful to perform output validation on this component of the ABM in isolation, because the composite ABM dynamics emerge from the activity taking place in nodes of the fund–asset network (i.e., asset price movements and fund distressed selling decisions). The quantitative output validation of the full 13-F scale ABM is also not possible, because a large-scale intraday fire-sale induced deleverage event has yet to be observed in equity markets⁶. By exploring realistic scenarios in Chapter 5 we contribute to the quantification of conditions giving rise to an increased risk of such an event, as well as testing the efficacy of potential regulatory interventions intended to reduce that risk.

4.2.2 Macro-Level Component: Bank–Fund Leverage

Input validation for the bank–fund leverage component of the ABM consists of two aspects, as for the fund–asset component discussed in Section 4.2.1. Input validation of the simulation structure and agent behaviours was presented in Section 3.3. However, unlike the fund–asset component, regulatory filings do not fully specify the initial conditions of the bank–fund leverage component. 13-F filing regulations do not require investors to disclose their leverage and margin arrangements. Since 2013, some US funds are required to report their leverage using SEC form PF⁷ (in particular, private equity funds), however, this form only requires funds to report their aggregate borrowing, and does not allow inference of the full bank–fund network. More significantly, the majority of 13-F reporting funds are not private equity firms, and so are not required to file form PF.

⁶The closest historical example is the August 2007 Quant Quake (Section 2.3.2), but due to the hedged nature of affected funds, the distress did not propagate into the wider economy or trigger a systemic crisis

⁷<https://www.sec.gov/about/forms/formpf.pdf>, accessed 2019-03-28.

By considering the matrix representation of the fund–asset network, encapsulated in the US 13-F regulatory filing dataset, we can infer a lower bound on the total value of each fund’s portfolio, and hence determine for each fund the product $C^0(1 + \lambda^0)$ (see Section 3.3.1). However we do not have access to data that would allow us to disentangle the product.

The unavailability of data covering the bank–fund network component of the ABM gives rise to a large number of free parameters ($\mathcal{O}(n_b \times n_f)$) that could in principle be calibrated. However, calibration requires the specification of a loss function (Section 4.1.4). Since a large-scale fire-sale induced deleverage event has not been observed in equity markets, there are no empirical data available on which to base such a loss function. Thus, calibration and output validation are not relevant to this component of our ABM. Instead, in Chapter 5, we consider alternative scenarios that allow us to investigate the relative influence of capital, leverage, and leverage tolerance hysteresis, at the fund level. Our scenario selection is guided by the findings reported by [Ang et al. \(2011\)](#), who obtained privileged access to leverage information for a wide range of asset management funds.

4.2.3 Micro-Level Component: Market Microstructure

As with the macroscopic components of the ABM, input validation of the asset-level microstructure component consists of two aspects. The structure of this component, and the behaviour of the various trading agents was presented in detail in Section 3.3, with further details about specific trading agent classes provided in Appendix A. The second input validation aspect, the initial conditions, are handled somewhat differently to the macro-level components:

1. All asset start prices are fixed to \$100. Although our data would allow us to specify start prices identical to asset real-world counterparts, analysis is simplified if we re-scale to a uniform price, with no loss of generality⁸.
2. Limit order books are initially empty. Although this convention gives rise to the need to accommodate an *initial transient* phase while the LOB becomes populated (see Section 4.4), we do not have sufficient data to populate the LOB at the level of individual orders (the properties of the available data are discussed below).

⁸This is discussed in more detail in Section 5.1.3.

We also assume that all limit orders placed by trader agents are multiples of a 100 share *round lot size*, the standard size for US equities (Roseman et al., 2018). Sub-lot orders (orders for less than 100 shares) accounted for only 8.31% of market volume for NYSE and NASDAQ listed securities between 9 December 2013 and 31 March 2014 (Roseman et al., 2018). The authors also question the extent to which algorithmic trading is responsible for sub-lot volume. The minimum price increment between orders on the LOB, the *tick size*, is set at \$0.1, which results in a relative price difference of 0.1% (10 *basis points*) per tick for assets with prices near \$100. This is a similar relative price difference to S&P 500 e-mini futures contracts during the 6th May 2010 Flash Crash (25 basis points, Paddrik et al., 2012).⁹

In order to facilitate the *quantitative* output validation of our ABM, we have obtained access to US equity price and volume tick-level data, for 2009-Q1–2018-Q4 inclusive. After cleansing, the dataset comprises an average of 640 equity assets for each of the 2500 trading days in the sample period. The market cap range of the assets is representative of the market cap ranges estimated from the 13-F dataset using Equation 4.20, and our dataset therefore offers a *whole-of-market* perspective.

A survey of existing literature on microstructure ABM calibration and validation (see Table 4.1) reveals that our dataset is orders of magnitude larger in scope than other comparable works. Our dataset is more extensive than most comparable studies in terms of the date range used (10 years), the number of assets represented, and the time resolution of the data. We also make use of more recent data than is typical (including data up to the end of 2018). Figure 4.2 illustrates the scope of our dataset in comparison with previous studies.

In order to reduce this extremely large dataset into a form that is suitable for the quantitative validation of the microstructure component of our ABM, we perform the following operations, discussed in more detail below:

1. Sample 180 ten-second trade price and volume data windows, representing contiguous blocks lasting 30-minutes per asset per day (the justification for this approach is given below).
2. Compute the value of a set of longitudinal moments on each data sample (full details of moments are given in Section 4.3).
3. Partition the assets into a fixed number of bands according to their market capitalisation (the justification for this approach is also given below).

⁹It is standard practice in financial literature and industry to use basis points to specify small relative price differences. One basis point is equal to one-hundredth of one percent <https://www.investopedia.com/terms/b/basispoint.asp>, accessed 2019-03-28.

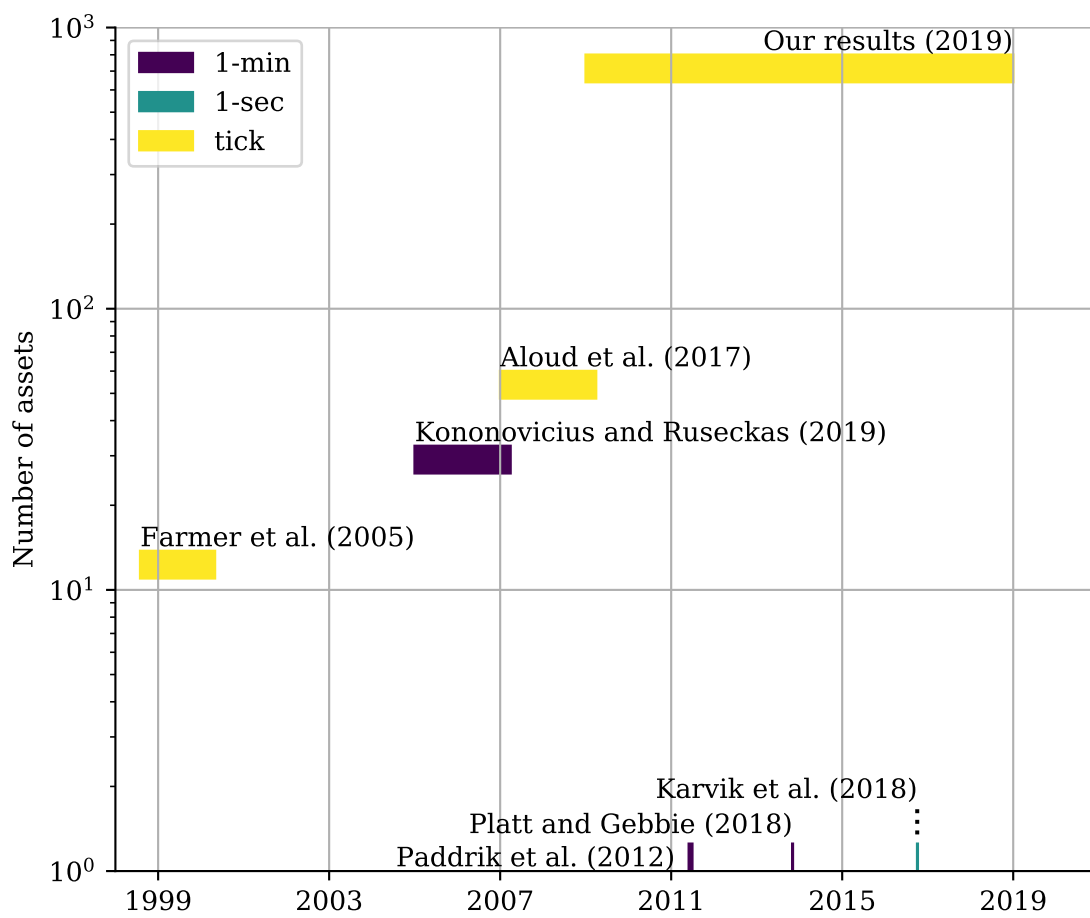


Figure 4.2: Comparative scope of microstructure ABM empirical datasets.

Paper	From	To	n_a	δt	N
Fabretti (2013)	2005-10-14	2007-10-12	1	daily	520
Recchioni et al. (2015)	2011-02-25	2012-02-23	4	daily	980
Braun-Munzinger et al. (2018)	2010-Q4	2015-Q2	1	daily	1500
Platt and Gebbie (2018)	2013-11-01	2013-11-05	1	1-min	2400
Franke and Westerhoff (2012)	1980	2007	1	daily	7300
Paddrik et al. (2012)	2011-06-01	2011-06-30	1	1-min	1.0×10^4
Lux (2018)	1980-01-01	2004-12-31	7	daily	4.5×10^4
Kårvik et al. (2018)	2016-10-03	2016-10-07	1	1-sec	1.4×10^5
Farmer et al. (2005)	1998-08-01	2000-04-30	11	tick	6.0×10^6
Kononovicius and Ruseckas (2019)	2005-01	2007-03	26	1-min	7.0×10^6
Aloud et al. (2017)	2007-01	2009-03	48	tick	1.1×10^8
Our results	2009-01-01	2018-12-31	640	tick	5.1×10^{12}

Table 4.1: Comparison of the scale of datasets used to calibrate and validate microstructure ABMs. *From* and *To* refer to the (inclusive) date range of the dataset, with time resolution denoted δt . *Tick* refers to non-binned datasets with information at the level of individual trades. n_a is the number of assets in the dataset. The table rows are sorted in ascending order by the total number of observations N . For direct comparison with binned ($\delta t \neq tick$) datasets, our derived 10-second resolution dataset comprises $N = 4.8 \times 10^9$ observations.

4. Concatenate the moment data for all assets in a given band on a given day into a single matrix.
5. Concatenate the per-day moment matrices for a given band for all dates in a given quarter, into a single matrix.
6. The result is a set of moment matrices, one for each (band, quarter) pair. Each matrix has a column per longitudinal moment and a row for each 30 minute data window sampled for each asset on each day of the quarter (approximately $N = 8000$ observations in each matrix).
7. Estimate the mean and confidence interval for the empirical moments for each (band, quarter) using Equations 4.11 and 4.13.

Following this procedure, we arrive at a *training set* (Bishop, 2006) consisting of an estimated set of empirical moments $\hat{\mu}(\mathbf{Y}_R)$ for each market cap band and quarter in the microstructure dataset.

Aggregating assets into a small number of market cap groups helps us manage the model’s bias-variance trade-off (Bishop, 2006). Sirignano and Cont (2019, p.12) make a similar argument: “the stock-specific model is more exposed to overfitting due to the smaller dataset, while the universal model is able to generalize by interpolating

across the rich scenario space of the pooled data set, and therefore is less exposed to overfitting.” Each stock produces a single training data point (a vector of empirical moments, see Section 4.1.3). Variance is minimised if we use all stocks to train a single ABM, however such a *one-size-fits-all* model may be highly biased, because stocks with very different trading characteristics would be simulated identically. Bias is minimised if we train a separate model for every stock, however this approach maximises variance because the number of fitted parameters would be extremely high, whereas the number of training data points is limited to one per asset, as mentioned above. We have selected a suitable intermediate level of asset aggregation based on a division of assets into five sets of equal cardinality (hereafter, *bands*) by market capitalisation. Let the five market cap bands be indexed using variable $b \in (0, \dots, 4)$, such that band $b = 0$ contains the smallest 20% of assets by market cap, band $b = 1$ contains assets in the 20th–40th percentiles, and so on. All bands thus contain an approximately equal number of assets.

A further advantage of aggregating assets into a small number of market cap bands is that we remove the effect of asset industry sector, or other idiosyncratic effects that could limit the generality of our results.

Market cap bands are recalculated on a daily basis in order to avoid introducing survivorship bias (Brown et al., 1992), and to allow band boundaries to vary with the market cap of the equities sector. We observe trends towards an increase in overall market cap for 2009–2018, concentrated into a smaller number of securities (following data cleansing), consistent with widely reported post-crisis financial market observations¹⁰. Table 4.2 provides an overview of the date and asset coverage of the microstructure dataset. The breadth of this dataset is far more comprehensive than is typical for previous microstructure ABM calibration research, where data availability is often restricted to a small number of assets, or a small number of dates, or both. As stated by Gould et al. (2013, p.1732) “strong assertions have been made by empirical studies based on single stocks over very short time periods.”

We sample microstructure data in 30-minute windows (i.e., 180 observations at 10-second resolution) in order to align our measurements with the characteristic time scale of major flash crash events (the events of 6th May 2010 occurred over a period of approximately 30 minutes (Menkveld and Yueshen, 2019)). Since equity markets are known to exhibit patterns of intraday seasonality (Admati and Pfleiderer, 1988), we control for this source of bias by selecting a randomised start time for each 30-minute data window. The applicability of our results is, therefore, only constrained to the

¹⁰<https://bloom.bg/2RwuEBE>, accessed 2019-03-07.

intraday time window represented by the earliest random window start time (selected at 10:15, to allow initial market volatility to subside (Admati and Pfleiderer, 1988)) and the latest random window end time (selected at 15:45, to avoid effects related to the market close at 16:00).

Item	Value
Number of market cap bands	5
Assets per market cap band in 2009	134
Assets per market cap band in 2018	121
Market cap range in 2009 / \$USD billions	0.023 – 415
Market cap range in 2018 / \$USD billions	0.034 – 1121
Primary listing exchange for assets in sample	New York Stock Exchange (NYSE)
Primary listing exchange for assets in sample	Nasdaq Stock Exchange
Primary listing exchange for assets in sample	American Stock Exchange (AMEX)
Time resolution of sample bins / seconds	10
Sample period duration / seconds	1800
Number of bins in sample period	180
Intraday sample window time-zone	(UTC-5, UTC-4 during DST) [†]
Equity market open time	09:30
Intraday sample window earliest start time	10:45
Intraday sample window latest end time	14:45
Equity market close time	16:00

Table 4.2: Market microstructure dataset properties for 2009–2018. [†]See ISO 8601 standard for date and time interchange at <https://www.iso.org/standard/70907.html>, accessed 2019-03-20.

The availability of training dataset $\hat{\mu}(\mathbf{Y}_R)$ permits us to define a loss function $\hat{\mathcal{L}}(\boldsymbol{\theta})$ that quantifies the ability of the microstructure ABM to produce output that is consistent with the rwDGP (Section 4.1.3). The definition of the longitudinal moments that constitute the output of our model is given in Section 4.3.

As described in Section 4.1.4, we can also use $\hat{\mathcal{L}}(\boldsymbol{\theta})$ as a way of selecting between competing models generated as part of a calibration process. In the present chapter we calibrate 17 parameters of the microstructure component of the ABM. These parameters, along with their permitted feasible ranges are given in Table 4.3. The parameters cover the number of each of the different classes of trader agent introduced in Section 3.3.4, along with parameters governing the order placement activity of those agents (order size, and spread) as well as risk management parameters related to maximum inventory size and sensitivity to toxic market conditions (Easley et al., 2012). These parameters generalise the approach taken by Paddrik et al. (2012) and

offer the possibility of deploying the microstructure ABM for equity assets, and indeed further asset classes, beyond the scope of [Paddrik et al. \(2012\)](#)’s original S&P 500 e-mini simulation domain. The parameter ranges presented in Table 4.3 were based on the parameters used by [Paddrik et al. \(2012\)](#) to simulate the S&P 500 e-mini futures contract. This financial instrument is much more liquid than any single equity, indeed it is one of the most heavily traded financial instruments in the world ([CFTC and SEC, 2010](#)). In preliminary calibration experiments we verified that the calibrated parameters were not found to be near the upper limit of the selected ranges, and on this basis, we consider the ranges to be sufficiently wide. In order to keep the number of free parameters tractable, all agents within a given agent class (e.g., high frequency traders) share identical order placement and risk management settings. Our approach is consistent with existing ABM methodology. Moving beyond it would require extremely fine-grained data, and would imply progress to Level 3 (the final level) of the taxonomy of [Axtell and Epstein \(1994\)](#). Restricting parameters in this way can be compared to a form of model regularisation ([Bishop, 2006](#)), and our approach is consistent with Machine Learning methods, where regularisation is deployed in order to reduce the risk of model over-fitting, and improve model generality.

Calibrated parameters	Symbol	Range
Number of small noise trader agents	<code>n_noise</code>	$(1, \dots, 500)$
Number of fundamental buyer/seller agents	<code>n_f_bs</code>	$(1, \dots, 500)$
Number of opportunistic trader agents	<code>n_opp</code>	$(1, \dots, 500)$
Number of market maker agents	<code>n_mm</code>	$(1, \dots, 200)$
Number of high frequency trader agents	<code>n_hft</code>	$(1, \dots, 20)$
Market maker maximum spread	<code>mm_spread_max</code>	$(1, \dots, 200)$
HFT maximum spread	<code>hft_spread_max</code>	$(1, \dots, 200)$
Market maker inventory	<code>mm_inventory</code>	$(1, \dots, 200)$
Opportunist inventory	<code>opp_inventory</code>	$(1, \dots, 200)$
HFT inventory	<code>hft_inventory</code>	$(1, \dots, 200)$
Small trader maximum order quantity	<code>noise_order_qty_max</code>	$(1, \dots, 200)$
Fundamental buyer/seller maximum order quantity	<code>fbs_order_qty_max</code>	$(1, \dots, 200)$
Opportunist maximum order quantity	<code>opp_order_qty_max</code>	$(1, \dots, 200)$
Market maker maximum order quantity	<code>mm_order_qty_max</code>	$(1, \dots, 200)$
HFT maximum order quantity	<code>hft_order_qty_max</code>	$(1, \dots, 200)$
Fundamental buyer/seller toxicity threshold ticks/min	<code>mm_toxic</code>	$(1, \dots, 50)$
Market maker toxicity threshold ticks/min	<code>fbs_toxic</code>	$(1, \dots, 50)$

Table 4.3: Model parameters requiring calibration. All parameters are positive natural numbers, though this is not a condition of the calibration methods used in this chapter, and does not reduce the generality of our approach.

Finally, Table 4.4 presents microstructure ABM parameters that we choose not to expose to calibration. The table provides the values selected for these parameters. In addition to the asset initial conditions and properties (price, tick size, round lot size) discussed above, we also choose not to calibrate the timescale of agents in the simulation. We hold these values fixed according to the activity timescales identified by Paddrik et al. (2012), and Hayes et al. (2014). Consistent with the real world, it would not make sense to allow small traders to operate at HFT speeds, or to allow HFT agents to trade very infrequently. The timescale of these agent behaviours is a key component of their behaviour, and, along with the other properties of the agents comprising the microstructure component of our model, derived from Paddrik et al. (2012), were based on extensive interviews with market participants carried out by regulators in the aftermath of the 2010 Flash Crash (CFTC and SEC, 2010) as reported by Kirilenko et al. (2017)¹¹. Based on these interviews, regulators categorised market participants into different groups according to their trading behaviour (full details in Section 3.3.4):

1. Small (noise) traders
2. Fundamental buyers / sellers
3. Opportunistic trend followers
4. Market makers
5. High-frequency traders

Within these classes, regulators probed the specific behaviours observed during the Flash Crash, such as the critical level of price falls, below which certain participants would cease order placement in the market. The regulators determined the number of participants in each trader class and their aggregate inventory capacities, along with behavioural assumptions idiosyncratic to each class. The implementation of these behavioural features for the different classes of agents are presented in Appendix A. The stated aim of Paddrik et al. (2012) was to produce an ABM with an internal structure that is as consistent as possible with observed real-world behaviours. We extended Paddrik et al. (2012)'s approach by introducing realistic intraday volatility auctions into the model and simulation.

The remaining parameters presented in Table 4.4 are HFT passive/aggressive bias, and fundamental buyer/seller private valuation. Available evidence suggests that HFT agents adaptively switch between passive and aggressive strategies during

¹¹Originally published as Kirilenko et al. (2011).

the trading day, so an intermediate aggression probability is justified (Menkveld, 2016; Paddrik et al., 2012). Setting the fundamental buyer/seller private valuation to values near the asset start price of \$100 reflects the longer-term nature of the valuation scheme used by that agent class (i.e., referring to data from previous trading days, rather than responding to intraday price movements).

Imposed parameters	Notation	Value
Asset start price / \$USD	p^0	100
Asset tick size / \$USD		0.1
Asset round lot quantity		100
Activity timescale of small trader agents		7200s
Activity timescale of fundamental buyer/seller agents		60s
Activity timescale of opportunistic trader agents		120s
Activity timescale of market maker agents		20s
Activity timescale of high frequency trader agents		2s
Activity timescale of distressed seller agents	δ	10s
HFT passive/aggressive bias		50%
Fundamental buyer/seller private valuation / p^0		99.9%–100.1%

Table 4.4: Summary of fixed model parameters.

4.2.4 Summary of Data and Validation Approach

Component / dataset	Input validation	Output validation	Calibration
Bank–fund leverage	Structural	n/a	n/a
Fund–asset network	Initial conditions	n/a	n/a
Microstructure	Structural	Quantitative	17 parameters

Table 4.5: Summary of validation and calibration approach for the model

As the preceding sections have revealed, each of the distinct, interconnected components of the ABM has its own data requirements (see Figure 4.1), some of which are satisfied and some which are not. In addition, we have described the different methods appropriate to each component for input validation, output validation and calibration. Table 4.5 summarises the approach we will take to calibration and validation of the ABM, which is represented schematically in Figure 4.3.

Although we have justified the inclusion of all parameters of the ABM based on features and behaviours that are believed to be representative of the real-world, input validation alone cannot determine whether the selected parameters are sufficient to

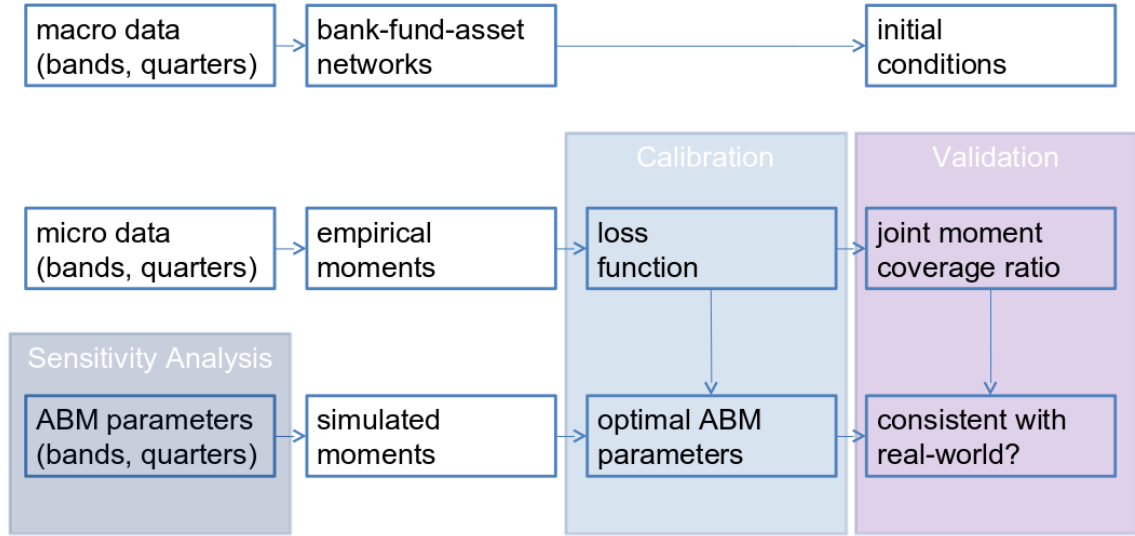


Figure 4.3: Block diagram of our calibration and validation methodology.

allow the model to be calibrated to, and subsequently validated against, real-world data. We address the quantitative output validation of the ABM in Section 4.7. Furthermore, input validation alone cannot determine whether the selected parameters are redundant, and do not materially affect model output (Gatti et al., 2018). The assessment of parameter relevance is known as *sensitivity analysis* and we undertake this procedure in Section 4.5. However, calibration, output validation and sensitivity analysis all require a rigorous definition of model output $\mathbf{Y}$, and the identification of longitudinal moments derived from the model output, $\boldsymbol{\mu}$. We address this matter in the next section.

4.3 Model Output Moments and Stylised Facts

As discussed in Section 4.2, it is the market microstructure component of our ABM that is to be subjected to quantitative calibration and output validation. The aspects of ABM output we will measure and compare with real-world data are thus restricted to longitudinal moments, i.e., stylised facts, based on market microstructure regularities. We aim to use a comprehensive set of moments, following the guidance from Winker et al. (2007, p.132), who state “it is advisable to include a large number of moments, taking into account all relevant characteristics of the actual data in order to increase the probability of identification of model parameters.”

The set of moments $\boldsymbol{\mu}$ that we use in order to compare the output of the ABM with the rwDGP predominantly consists of those that are well-established in the financial

ABM literature:

1. Return standard deviation (volatility) (Cont, 2001)
2. Return skewness (Kim and White, 2004)
3. Return kurtosis (Mandelbrot, 1963)
4. Lagged autocorrelation of returns (lags 10 and 50 seconds) (Cont, 2001)
5. Lagged autocorrelation of volatility (lags 10, 50 and 100 seconds) (Kirman and Teyssiere, 2002)

In addition to these moments, we introduce two further metrics which extend the set of stylised facts beyond those commonly encountered in the ABM literature:

1. Maximum drawdown (Leal and de Melo Mendes, 2005), and
2. Log liquidity (Amihud and Mendelson, 1991)

We give details regarding the reason for selecting these moments, and their calculation, below.

All moments are calculated using price and volume data that are partitioned into bins, each lasting ten simulated seconds. This binning scheme corresponds exactly to the structure of the empirical microstructure dataset described in Section 4.2.3. We thus enforce a direct correspondence between simulated and real time. Let us introduce the following definitions:

Definition 4.1 *Let $\mathcal{W} = (w_i), i \in (1, \dots, n)$ be a sequence of non-overlapping time intervals (bins), such that bin $w_i = (t^i, t^{i+1})$, where $t^i < t^{i+1}$. Let t^i be a time measured to a precision of 10^{-3} seconds for empirical data, and measured to a precision of $\delta_t = 5 \times 10^{-2}$ seconds in simulation (reported as an offset from the start of the simulation, in this case). For convenience, we fix $t^{i+1} - t^i = 10$ seconds $\forall i$. For empirical data, we select $t^0 \sim \text{Uniform}(10:15:00.000, 14:15:00.000)$ (see Table 4.2 and discussion in Section 4.2.3). For simulations, we select $t^0 = 0$, i.e., the start of the simulation.*

Definition 4.2 *Let us define a trade $\mathbf{x} = (t, p, q, a)$ where t is the reported execution time of the trade to a precision of 10^{-3} seconds for empirical data and $\delta_t = 5 \times 10^{-2}$ seconds for simulated data, $p \in \mathbb{R}_+$ is the trade price and $q \in \mathbb{N}_+$ is the quantity of traded shares. The trade is associated with asset $j \in (1, \dots, n_a)$.*

Definition 4.3 Let us further define the time-ordered sequence of trades for the j^{th} asset that occur in bin w_i as $\mathcal{T}_j^i = (\mathbf{x} \mid t^i \leq \mathbf{x}(1) < t^{i+1}, \mathbf{x}(4) = j)$, where $\mathbf{x}(k)$ refers to the k^{th} element of the trade tuple in Definition 4.2 ($k = 1$ corresponds to trade time t , and so forth).

Definition 4.4 Bin volume can now be defined for bin w_i and asset j as $v_j^i = \sum_{\mathbf{x} \in \mathcal{T}_j^i} \mathbf{x}(3)$.

Definition 4.5 Recall from Definition 4.3 that $\mathcal{T}_j^i$ is ordered by the arrival time of its constituent trades. We can thus define the bin price for bin w_i and asset j as $p_j^i = \mathbf{x}^*(2)$, where $\mathbf{x}^* = \mathcal{T}_j^i(|\mathcal{T}_j^i|)$, i.e., the price of the latest trade, $\mathbf{x}^*$, observed in $\mathcal{T}_j^i$.

Definition 4.6 Finally, let us define the time-ordered sequence of bin price returns for asset j as $\mathcal{R}_j = (r_j^i), i \in (1, \dots, n-1)$, where $r_j^i = \frac{p_j^{i+1} - p_j^i}{p_j^i}$ is the price return from bin w_i to bin w_{i+1} for asset j .

4.3.1 Moment Definitions

Having established definitions for the binned price and volume data that constitutes both our simulated and empirical datasets, we can now give definitions of the longitudinal moments (stylised facts) that form the basis of the quantitative loss function $\hat{\mathcal{L}}(\boldsymbol{\theta})$ (Equation 4.10).

Definition 4.7 (Return standard deviation (volatility)) The standard deviation of bin price returns for asset j is given by $\sigma_j = \mu_{\text{std(rets)}} = (\frac{1}{n-1} \sum_{i=1}^{n-1} (r_j^i - \mu_j)^2)^{\frac{1}{2}}$, where $\mu_j = \frac{1}{n-1} \sum_{i=1}^{n-1} r_j^i$ is the mean price return for asset j .

Definition 4.8 (Return skewness and kurtosis) The skewness and kurtosis of bin price returns for asset j are higher order central moments (the third and fourth order, respectively), given by $\mu_{\text{skew(rets)}} = \frac{E[(r_j - \mu_j)^3]}{\sigma_j^3}$ and $\mu_{\text{kurtosis(rets)}} = \frac{E[(r_j - \mu_j)^4]}{\sigma_j^4}$, where $E[\cdot]$ is the expectation operator (Kim and White, 2004).

Definition 4.9 (Lagged autocorrelation of returns) The autocorrelation function for asset j binned price returns with lag $k \in (1, \dots, n-1)$ is defined as $\mu_{\text{autocorr(ret)-lag}k} = \frac{\sum_{i=1}^{n-k} (r_j^i - \mu_j)(r_j^{i+k} - \mu_j)}{\sum_{i=1}^n (r_j^i - \mu_j)^2}$ (Box et al., 2015). We use bin price return autocorrelation lags of $k = 1$ (10 seconds) and $k = 5$ (50 seconds).

Definition 4.10 (Lagged autocorrelation of volatility) Let bin price volatility ν_j^i for asset j in bin w_i be given by $\nu_j^i = |r_j^i|$ (Cont, 2001). Let us further define the mean bin volatility $\zeta_j = \frac{1}{n-1} \sum_{i=1}^{n-1} \nu_j^i$. Then the autocorrelation function for asset j binned price volatility with lag $k \in (1, \dots, n-1)$ is defined as $\mu_{\text{autocorr}(\text{abs}(\text{ret}))-\text{lag}k} = \frac{\sum_{i=1}^{n-k} (\nu_j^i - \zeta_j)(\nu_j^{i+k} - \zeta_j)}{\sum_{i=1}^n (\nu_j^i - \zeta_j)^2}$ (Box et al., 2015). We use bin price volatility autocorrelation lags of $k = 1$ (10 seconds), $k = 5$ (50 seconds), and $k = 10$ (100 seconds).

The inclusion of maximum drawdown allows us to calibrate and validate the scale of temporally-extended price falls encountered among the training data. We define maximum drawdown as the lowest price obtained by an asset relative to its starting price, after correcting for return drift between the prices at the start and end of the simulation window by de-trending. It is necessary to correct for drift because we are concerned with calibrating the behaviour of the model in a general setting, where prices could be trending up, down, or remaining relatively constant over the sampled real-world interval.

Definition 4.11 (Maximum drawdown) Let the time-ordered sequence of geometric bin price returns for asset j be $\mathcal{G}_j = (g_j^i), i \in (1, \dots, n-1)$, where $g_j^i = \ln(1 + r_j^i)$ (Jacquier et al., 2003). Furthermore, let the de-trended sequence of geometric returns for asset j be $\mathcal{G}'_j = (g_j'^i) = (g_j^i - \frac{1}{n-1} \sum_{i=1}^{n-1} g_j^i), i \in (1, \dots, n-1)$. Then $\mu_{\text{drawdown}} = \exp(\min(\mathcal{H}_j)) - 1$, where $\mathcal{H}_j = (h_j^i), i \in (1, \dots, n-1)$ is the sequence defined by a rolling summation such that $h_j^i = \sum_{k=1}^i g_j'^k$, with conversion back to arithmetic space (Jacquier et al., 2003).

The inclusion of log liquidity — which we define as the log sum of the products of price and quantity traded in each time period (Definition 4.12) — allows us to calibrate and validate the volume (quantity of shares)-related aspects of the model. These aspects include agent order quantities and maximum inventory levels. Volume considerations are frequently absent from ABM microstructure calibration experiments, and even sophisticated recent approaches such as the GSL-divergence metric (Fagiolo et al., 2019) focus exclusively on price data. However, similarly to the maximum drawdown moment, the log liquidity calibration is necessary to scale the ABM such that it exhibits volume levels in agreement with real-world observables. Our convention of fixing simulated asset start prices at \$100 (Section 4.2.3) means that we must use liquidity, rather than just volume, for this metric. Liquidity is also robust to corporate actions such as stock splits, and so this decision reinforces the generality of our approach.

Definition 4.12 (Log liquidity) *The log liquidity for asset j from trades occurring during bins $\mathcal{W}$ is given by $\mu_{\text{logliq}} = \ln(\sum_{i=1}^n v_j^i p_j^i)$.*

Equation 4.11 specifies how we combine multiple real-world observations of a moment, for assets within a given market cap band and over a given date range, into an estimate of the true real-world moment value, that is suitable for inclusion in $\hat{\mathcal{L}}(\theta)$.

4.4 Initial Transient Analysis

In order to minimise noise in the measurement of the moments introduced in Section 4.3, it is necessary to exclude from analysis any initial transient phase (Gatti et al., 2018). The limit order books that constitute the simulation environment of the microstructure ABM are initially empty, i.e., they contain no limit orders. It is therefore necessary to exclude from analysis the initial transient period prior to the LOB becoming adequately populated such that statistical properties reach a stable equilibrium state. In this section we use a graphical method to estimate the duration of the initial transient phase, by the end of which the LOB dynamics have converged and it is possible to accurately measure the required longitudinal moments.

As discussed in Section 3.2.1, the microstructure ABM features a realistic (input validated) representation of an opening auction trading phase. This lasts for one minute, and during this time, orders can be added to, and removed from, the LOB, but no trades occur. In real-world asset markets the intention of the opening auction is to concentrate order placement activity and thus facilitate more robust start of day price discovery for the asset. Even with this provision, equity markets exhibit well-known intraday seasonality, with a transient period of high volatility following the opening auction (Admati and Pfleiderer, 1988).

Figure 4.4 presents an analysis based on 60-second moving averages of the output moments μ . We present the ratio of the standard deviation of the moment in each 60-second window to the standard deviation of the moment over the full simulated interval (50 simulated minutes), scaled such that each moment has unit standard deviation, for ease of comparison¹². In order to sample the ABM output, 800 θ parameter vectors were generated from a quasi-random Sobol' process (Sobol', 2001), which ensures that the parameter space is sampled evenly (Saltelli et al., 2019)¹³. The shaded region in Figure 4.4 indicates the 95% confidence interval derived from 800

¹²Due to the small moving average window size, it is necessary to exclude from this analysis moments relating to the autocorrelation of returns and volatility beyond the first lag.

¹³The uniform space sampling properties of Sobol' processes are discussed in detail in Section 4.5.

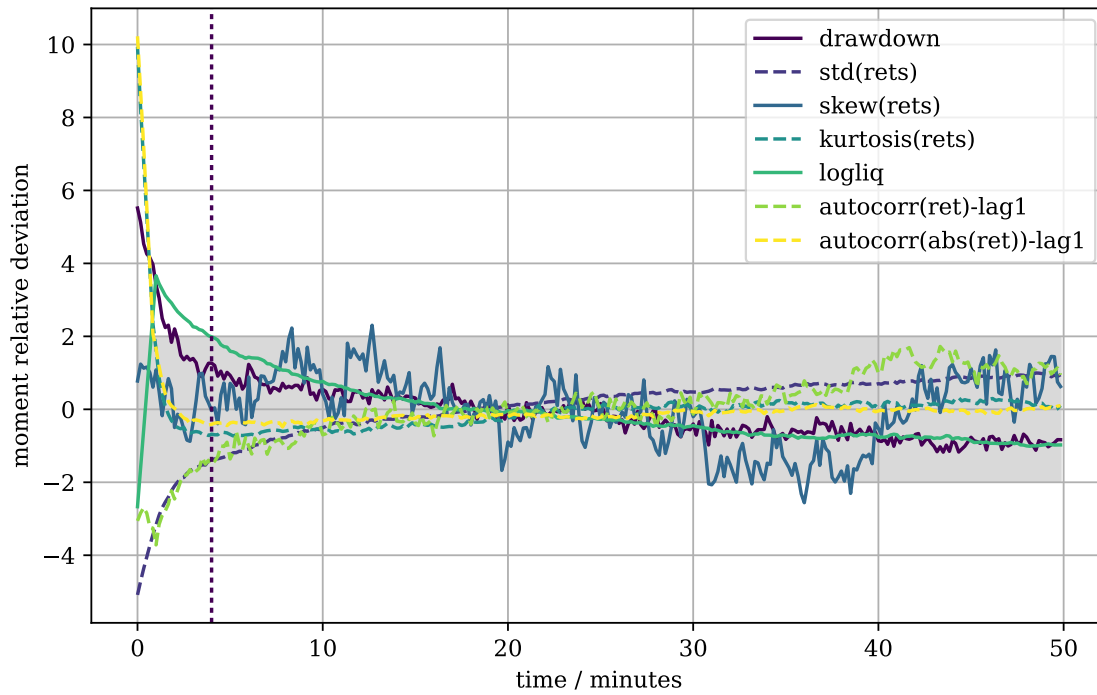


Figure 4.4: The ratio of moment values in a 60-second sliding window to the median of all 60-second moment values over the full simulation. For ease of comparison, the values for each moment are scaled to have unit standard deviation. The dashed vertical line at $t = 4$ minutes indicates the cut-off selected such that initial transients have decayed sufficiently to permit accurate measurement of simulated moments. We present the median and 95% confidence interval from 800 Monte Carlo evaluations of the ABM, with parameter vectors θ sampled from a quasi-random Sobol' process (Sobol', 2001).

simulations. The dashed vertical line at $T = 4$ minutes indicates the time at which the standard deviation of all moments is within each moment's long-term confidence interval. We therefore designate $T < 4$ minutes as the model's initial transient phase, and exclude model output from this phase in all subsequent analyses.

4.5 Global Sensitivity Analysis

The goal of sensitivity analysis is to quantitatively determine how uncertainty in different model inputs contributes to the overall uncertainty of model output (Saltelli et al., 2010). The redundancy of input parameters that do not contribute to output variance can thus be demonstrated. For our purposes, we consider model inputs to be the 17-dimensional parameters θ , and the model output to be the 10-dimensional vector of longitudinal moments $\mu(\theta)$. It is not necessary to incorporate real-world data in a sensitivity analysis, or to define a loss function. Sensitivity analysis instead attributes the model's output variance to variance in inputs parameters or combinations of input parameters. All that is necessary is to be able to sample vectors θ from the parameter space Θ , and to measure $\mu(\theta)$ for these samples.

Sensitivity analysis has received little historical attention in the ABM literature (Gatti et al., 2018), however, as the discipline of ABM moves onto a more quantitative footing, sensitivity analysis is emerging as a growing research direction (Broeke et al., 2016). As stated by Lux and Zwinkels (2018, p.481):

“there is often a danger of over-parametrization of [ABMs], and even a Monte Carlo exercise could easily reveal that a model has redundant parameters that could never be identified with the data one targets.”

There are many approaches to sensitivity analysis, each with its own benefits and drawbacks (Saltelli et al., 2010; Helton and Davis, 2002; Broeke et al., 2016). An intuitive and well-known approach is known as *one-at-a-time* (OAT). OAT sensitivity analysis entails establishing a reference parameter vector θ^* , and examining model output as we vary one parameter at a time, holding the others fixed at their reference values. We applied such a method in Section 3.9.2. The results from an OAT analysis are easy to visualise and understand, however this approach does not scale well to high-dimensional input spaces, and is unable to account for interactions between parameters (Saltelli et al., 2019). As discussed in Section 4.2, the ABM has 17

input parameters and 10 output moments¹⁴. The high dimensionality of these spaces precludes the direct evaluation of all parameter combinations, and so it is necessary to restrict evaluation to a carefully selected subset of parameter vectors.

Global sensitivity analysis attempts to overcome the shortcomings of OAT by using *sampling methods* to ensure that the parameter space is comprehensively explored, even when the dimensionality is high. A sampling method produces a sequence of N parameter vectors $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N) \subset \boldsymbol{\Theta}$, and $\boldsymbol{\mu}(\boldsymbol{\theta})$ is then evaluated for each sample by running the ABM. In general, the number of required samples is large, and is dependent on the dimensionality of the input parameter space (D in our notation). Some common sampling schemes include:

1. *Random*: a pseudorandom number generator, such as the Mersenne Twister algorithm (Matsumoto and Nishimura, 1998) provides points distributed uniformly at random over $\boldsymbol{\Theta}$. This method is computationally inexpensive and simple to estimate, but there is no control over the spacing of points, leading to clustering of points and poorly-sampled voids.
2. *Sobol'*: a quasi-random sequence of points is generated that uniformly fills the D -dimensional parameter space, avoiding the clustering inherent in random sampling, at the expense of somewhat higher computational cost (Sobol', 2001).
3. *Latin hypercube*: a quasi-random sequence with similar properties to Sobol' sampling, with additional guarantees that parameters are distributed as evenly as possible within their ranges, at significantly higher computational cost (Helton and Davis, 2002).

Sampling methods such as the quasi-random Sobol' or Latin hypercube methods are well established in the biological literature, where high-dimensional models are common, and evaluations can be very expensive (Helton and Davis, 2002; Marino et al., 2008). Sample-based sensitivity analysis is beginning to find traction in the ABM literature (Brooke et al., 2016). The output of a global sensitivity analysis is a set of *sensitivity indices* for each input parameter k . The first-order sensitivity $S_{1k} \in [0, 1]$ of model output to a parameter k can be interpreted as the reduction in output variance obtained if k is held fixed. (Saltelli et al., 2010) give the following expression:

$$S_{1k} = \frac{V_{X_k}(E_{\mathbf{X}_{\sim k}}(Y | X_k))}{V(Y)}, \quad (4.21)$$

¹⁴The analysis in Section 3.9.2 explored a smaller set of higher level parameters unrelated to market microstructure.

where (based on Saltelli et al. (2010)'s notation) X_k is the column vector representing the sampled values of the k th input parameter, $\mathbf{X}_{\sim k}$ is the design matrix of all sample parameter vectors *excluding* parameter k , $E(\cdot)$ and $V(\cdot)$ are mean and variance operators, and Y is the scalar model output (corresponding to a single empirical moment, in our case).

It is similarly possible to define the second order sensitivity to k as the reduction in output variance obtained when all pairs of parameters including k are held fixed. The *total-order* sensitivity to k , denoted $S_{Tk} \in [0, 1]$, captures all n th order interactions with k and n other parameters, where $n \in (1, \dots, D-1)$. By construction, $S_{Tk} \geq S_{1k}$, and is given by (Saltelli et al., 2010):

$$S_{Tk} = \frac{E_{\mathbf{X}_{\sim k}}(V_{X_k}(Y \mid \mathbf{X}_{\sim k}))}{V(Y)}. \quad (4.22)$$

In the results that follow, we perform a global sensitivity analysis based on Saltelli et al. (2010)'s extension of Sobol's method (Sobol', 2001), using the implementation provided by the SALib library (Herman and Usher, 2017). We set a computational budget of $N = 100,000$ sampled parameter vectors $\boldsymbol{\theta}$, and evaluate each vector over $I = 8$ independent Monte Carlo replications, where only the random seed s varies between trials. Each trial lasts for $T = 40800$ time steps at $\delta t = 50\text{ms}$ resolution, for a total of 34 simulated minutes. This allows for moments to be measured over 30 minutes, following a 4 minute initial transient period (see Section 4.4). The effect of varying N and I is investigated in Appendix C, where we demonstrate that $N = 100,000$ is sufficient to evaluate the total-order sensitivity of our ABM with acceptable precision.

Figure 4.5 presents the first-order and the total-order sensitivity of each output moment to each parameter, leading to a matrix of results with a row for each of the 17 input parameters, and a column for each of the 10 output moments. In order to make the results easier to interpret, we have applied thresholds to the sensitivity indices to indicate significance, so rather than presenting the raw sensitivity indices, we instead indicate whether the first-order sensitivity of moment m to parameter k exceeds 5% (i.e., explains at least 5% of output moment variance, dark shading), or whether only the total-order sensitivity exceeds 5% (moderate shading). If total-order sensitivity is significantly higher than first-order sensitivity for a parameter, then this indicates that the parameter has strong interactions with other parameters (Saltelli

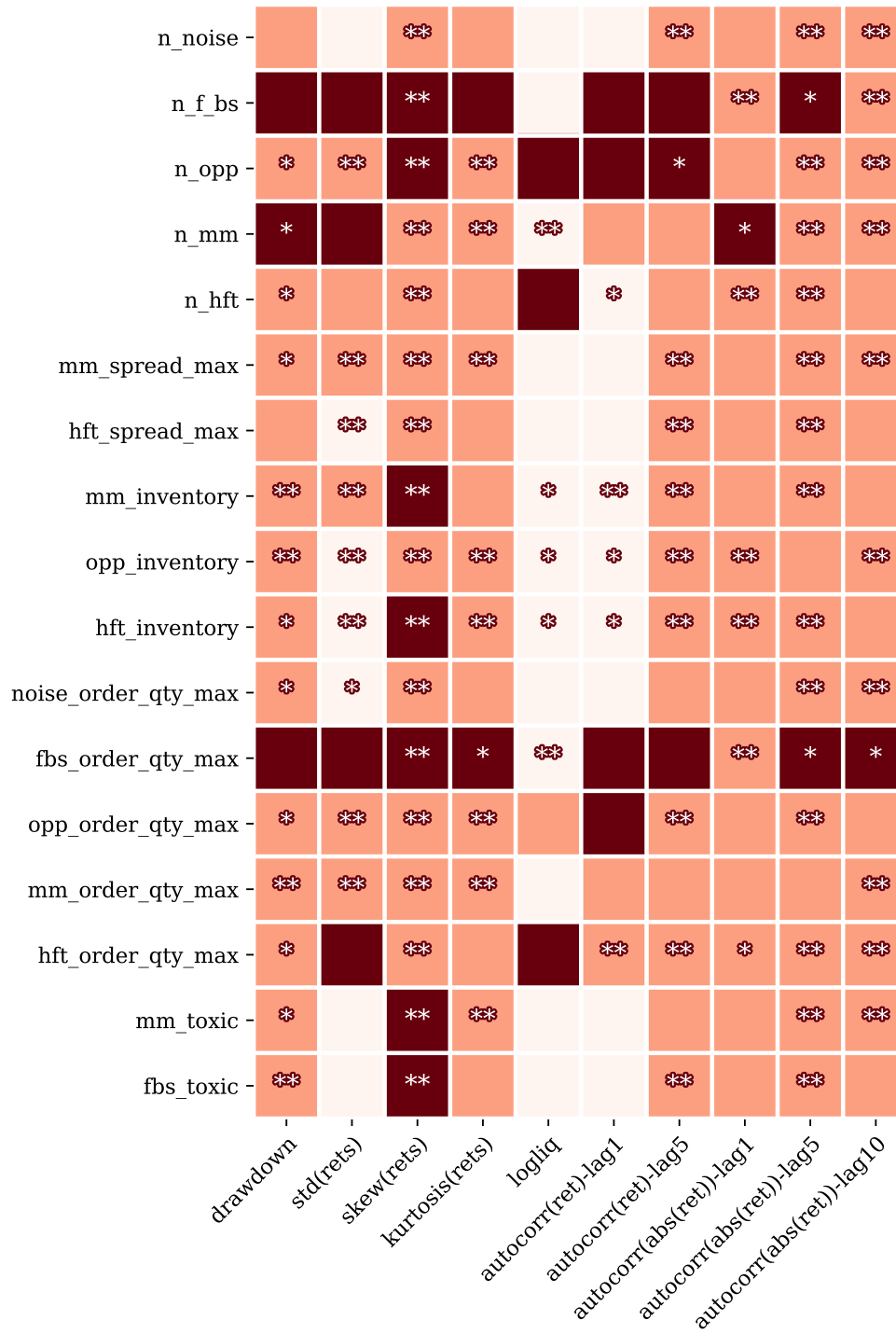


Figure 4.5: Sensitivity of moments to ABM parameters. Dark shading indicates significant first-order sensitivity ($S_{1k} > 5\%$). Moderate shading indicates significant total-order sensitivity ($S_{Tk} > 5\%$). Cases where $S_{Tk} > 5S_{1k}$ are labelled *, and we label with ** cases where $S_{Tk} > 10S_{1k}$.

et al., 2010). Accordingly, we label with * cases where $S_{Tk} > 5S_{1k}$, and we label with ** cases where $S_{Tk} > 10S_{1k}$.

The most important finding to emerge from Figure 4.5 is that every parameter has a significant total-order sensitivity index for at least seven of the ten output moments. This provides evidence that all parameters are relevant to the model, since an irrelevant parameter would not have a significant influence on model output variance (Saltelli et al., 2019). Indeed, some moments (e.g., `drawdown`, `kurtosis(rets)`) have a significant total-order sensitivity to *every* parameter, and many parameters even have a significant first-order sensitivity index for one or more moments. This is an important validation of the way in which the model is parametrised, and can be considered an extension of the model's input validation (Section 4.2).

Figure 4.5 shows that all moments feature significant first-order sensitivity to at least one parameter, and inspection of these first-order sensitivities reveals the critical role played by the number of agents and the order sizes used by those agents in the fundamental buyer/seller, opportunist, market maker and HFT classes, for many output moments. It is perhaps unsurprising that the sensitivity analysis reveals weaker output dependence on the properties of noise traders, as they are much less active in the simulation than other agent classes (they take action much less frequently than other agents and place more widely spaced orders on the limit order book, hence are less likely to participate in trades and directly affect model output). It is interesting that the log liquidity moment (`logliq`) does not have first-order sensitivity to the number (`n_f_bs`) or order size (`fbs_order_qty_max`) of fundamental buyer/seller agents. Instead, log liquidity is revealed to strongly depend on the number of opportunist (`n_opp`) and HFT (`n_hft`) agents, and the order sizes used by those agents (`opp_order_qty_max` and `hft_order_qty_max`, respectively). Opportunist and HFT agents often trade aggressively, by placing orders that match against resting passive orders on the offsetting side of the limit order book. Trades can only occur when an aggressive order meets a passive resting order, and so it is unsurprising that the simulated liquidity depends strongly on the activity of these agent classes. Additional passive resting orders would not, on their own, give rise to increased liquidity because the trading activity would still be limited by the volume of aggressive orders.

Figure 4.5 further reveals that simulated volatility (the standard deviation of price returns, `std(rets)`) depends strongly on the number of agents and order sizes for those agents in the fundamental buyer/seller, opportunist, market maker and HFT classes (`n_f_bs`, `n_opp`, `n_mm`, and `n_hft`, respectively). This result suggests that it is the interplay of trading behaviour between the more active agent classes,

including aggressive order placement in relation to the positioning of passive orders on the limit order book, that controls volatility. Although volatility (`std(rets)`) is not sensitive to inventory parameters (except for market makers, `mm.inventory`) or toxic flow sensitivity parameters, these parameters are nonetheless important for introducing asymmetry into the model output, which manifests as non-zero return skewness (`skew(rets)`). The model generally seems to struggle to produce significant skew, and this may reveal some numerical instability in the sensitivity analysis due to the presence of the output moment variance in the denominator of the expression for S_{Tk} (Saltelli et al., 2010). It does not appear that the sensitivity of the lack of autocorrelation of price returns, or autocorrelated volatility, is dominated by any particular subset of parameters. Rather, all parameters are shown to play a role in influencing these model outputs. The exception is the first lag of price return (`autocorr(ret)-lag1`), the sensitivity of which depends strongly on the same set of parameters as the un-lagged (i.e., raw) price return, `std(rets)`.

Finally, it is necessary to clearly state that while we find that no parameter is redundant in the model, the question of whether or not the model and its parameters are *sufficient* to reproduce empirically observed regularities requires us to compare the model output with real-world data, through calibration and validation, and we present these in the sections that follow.

4.6 Calibration Methodology

The parameters that we wish to calibrate are all discrete in nature, with ranges given in Table 4.3. The exponential number of parameter vectors ($\theta \in \Theta$) renders exhaustive search infeasible (a manifestation of the *curse of dimensionality*). A single parameter set evaluation over thirty simulated minutes requires at least five seconds on a single-core CPU, increasing to tens of seconds for parametrisations featuring large numbers of agents. Even very sparse sampling schemes over Θ are infeasible, and such sampling schemes might additionally imply that valid solutions are skipped over and lost.

Two factors compound the computational cost. Firstly, due to the inherent uncertainty in the measurement of moments, resulting from the stochastic nature of any given simulation run of the ABM, it is necessary to repeat each simulation multiple times with different pseudorandom number seeds in order to arrive at a reliable estimate of $\hat{\mathcal{L}}(\theta)$. Secondly, whereas some optimisation methods provide a direct quantification for the uncertainty in parameter estimates (for example, Markov chain

Monte Carlo), other methods do not. In order to determine confidence intervals for parameter estimates in such cases, it is necessary to independently repeat each optimisation procedure multiple times with different pseudorandom number seeds. Although this repetition implies additional computational cost, due to the independence between optimisation runs, the replications are parallelisable — a feature we exploit in the forthcoming analysis.

In order to objectively compare calibration methods, we adopt the convention of a limited computational budget (Sergeyev et al., 2018). Since individual evaluations of the ABM are relatively costly, it is not feasible for us to invoke the alternative convention of a fixed $\hat{\mathcal{L}}(\boldsymbol{\theta})$ target. The approaches are equivalent. A fixed loss target allows methods to be compared according to how many evaluations they require to obtain the target loss value. A fixed computational budget allows methods to be compared according to the loss value they are able to achieve within a given number of training iterations. The disadvantage of the latter approach is that an acceptable value of the loss function may not be achievable within the available computational budget. In order to adequately characterise the performance of the methods it is necessary to adopt a sufficiently large computational budget. This was felt to be preferable to selecting a target value of $\hat{\mathcal{L}}(\boldsymbol{\theta})$, which some methods may only obtain with computational budgets many orders of magnitude higher than our available resources (Sergeyev et al., 2018).

We define the computation budget as the number of evaluations of the ABM, rather than wall clock duration. Wall clock duration varies as a function of the number of agents present in the simulation (particularly so for HFT agents which are the most active agent class). Fixing the number of ABM evaluations therefore allows us to avoid introducing bias into optimisation results as may be the case for fixed wall clock budgets. A second consequence of this convention is that the CPU cost of non-ABM aspects of the optimisation is ignored. Due to the relatively high cost of ABM evaluations, this approximation is usually not material, with many orders of magnitude separating the two sources of CPU cost. However, non-ABM overheads do become material for Gaussian Process Regression and other surrogate methods, and we discuss this further below.

In the following sections, we describe four candidate calibration methods which we classify according to their strategies for the *exploration* and *exploitation* of the parameter space $\boldsymbol{\Theta}$. Two further considerations contributed to our method selection. Firstly, we have selected methods covering the full spectrum of the exploration-exploitation trade-off, incorporating members of many well-known classes of algorithm (Blum

and Roli, 2003). Secondly, we have endeavoured to make use of the most standard approach to implementing each method, and have avoided layering on additional features. By comparing the efficiency of broad classes of algorithm, we hope to provide useful signposting for ABM researchers to help them narrow down what can seem a bewildering choice of possible optimisation approaches (Sørensen, 2013). By testing straightforward and well-established versions of the algorithms, we hope to gain more general insight into the reasons for the success or failure of an approach, without obfuscating the finding in unnecessary detail. The methods we examine are:

1. Random search,
2. Nelder-Mead simplex search,
3. Genetic algorithms,
4. Markov chain Monte Carlo.

4.6.1 Parameter Recovery Preliminary Exercise

We performed a preliminary parameter recovery exercise in order to guide the selection of the hyperparameters for each candidate calibration method. For this preliminary exercise, a fixed set of ABM parameters $\tilde{\boldsymbol{\theta}}$ were used to generate 10,000 simulated asset time series. Distributions for each of the ten moments were calculated from the synthetic series, allowing us to define a pseudo-real loss function:

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}) = (\boldsymbol{\mu}(\boldsymbol{\theta}) - \boldsymbol{\mu}_S)^\top \mathbf{W} (\boldsymbol{\mu}(\boldsymbol{\theta}) - \boldsymbol{\mu}_S), \quad (4.23)$$

where $\boldsymbol{\mu}_S$ is the vector of synthetic (pseudo-real) moments, i.e., $\boldsymbol{\mu}_S = \boldsymbol{\mu}(\tilde{\boldsymbol{\theta}})$. By construction, $\tilde{\mathcal{L}}(\boldsymbol{\theta}) \rightarrow 0$ as $\boldsymbol{\theta} \rightarrow \tilde{\boldsymbol{\theta}}$, subject to stochastic error in the measurement of $\tilde{\mathcal{L}}(\boldsymbol{\theta})$, similarly to Equation 4.10. A similar method forms the basis of the analysis by Platt (2019).

We found the number of Monte Carlo replicates $I = 8$ to be sufficient during parameter recovery, and to give superior results to lower values. The additional computational cost associated with higher values of I was not found to be justified by induced improvement in $\tilde{\mathcal{L}}(\boldsymbol{\theta})$. We therefore fix $I = 8$ for all methods under consideration. The remaining hyperparameters were constrained by available computational resources. The number of ABM evaluations N was selected to be large enough for the differences between the methods to be fully characterised, since most of the methods exhibit different phases as N increases. The number of repeats of the optimisation

process was fixed at 20, a value that allowed method performance comparison with sufficient statistical fidelity.

Finally, we observed that calibration performance was improved by specifying W as a diagonal matrix with inverse standard deviations, rather than inverse variances, along the main diagonal. We attribute this to numerical stability, since moments relating to price returns (`drawdown`, `std(rets)`) have variances $\approx 10^{-8}$ which are then combined with variances $\approx 10^4$ resulting from the `logliq` moment.

4.6.2 Random Search

We use a form of random search (Solis and Wets, 1981) as a baseline for comparison with other methods. N parameter vectors $(\theta_1, \dots, \theta_N) \subset \Theta$ are sampled uniformly at random. Calibration output is the parameter vector $\hat{\theta}$ giving rise to the lowest encountered value of the loss function $\hat{\mathcal{L}}(\theta)$ (Equation 4.18). This method is appropriate for the optimisation of our model, and ABMs in general, because it does not require the loss function to admit a closed-form gradient, to be convex, or to be smooth. However, this approach does not directly model uncertainty in $\hat{\theta}$, and so repeated optimisation runs are required. A key advantage of random search is that the independence of the sampling process facilitates a high-degree of parallelisation, offering flexibility of implementation for high-performance computing (HPC) environments. A second advantage of the method is that it requires the specification of only a small number of hyperparameters, all of which were selected during the preliminary parameter recovery exercise, and are summarised in Table 4.6.

Hyperparameter	Value
Monte Carlo evaluations I	8
Parameter vector sample budget N	14600
Optimisation repeats	20

Table 4.6: Hyperparameters of the random search algorithm.

In our implementation, random search represents one extreme of the exploration-exploitation trade-off. Random search has only exploration, and no exploitation. Many extensions to random search are possible, incorporating aspects of local structure. However, such approaches are comparable to other well-known classes of stochastic optimisation methods including genetic algorithms and Markov chain Monte Carlo, and we focus on these approaches separately. The performance of random search, in comparison to the other methods, is reported in Section 4.6.8.

4.6.3 Nelder-Mead Simplex

The Nelder-Mead simplex (NMS) method (Nelder and Mead, 1965) prioritises exploitation of local structure over exploration of the parameter space. As Nelder and Mead (1965, p.311) describe: “Our method is highly opportunist, in that the least possible information is used at each stage and no account is kept of past positions.” The method is appropriate for the class of optimisation problems under present consideration because it does not require analytic gradients, however due to the local nature of its update procedure, non-convexity in the loss function can be problematic. Without algorithmic modification, there is no mechanism by which a converged simplex can escape a local minimum. We include it in the present study for two reasons: (1) because it occupies an opposite extreme of the exploration-exploitation trade-off to random search, and (2) because the method has obtained traction in the economic/financial ABM calibration literature (Gilli and Winker, 2003; Fabretti, 2013). In each of these contributions, the method is augmented with a threshold accepting criterion which frees the method from the requirement for strictly monotonic decreases in loss. Consequently, while remaining a local optimisation method, limited tolerance for non-convexity is introduced via a new hyperparameter.

Each instance of the NMS algorithm maintains a simplex comprising $D + 1$ vertices $(\theta_1, \dots, \theta_{D+1}) \subset \Theta$. The initial vertices are sampled uniformly at random from $\Theta \subset \mathbb{R}^D$. An iteration of the algorithm involves the replacement of the vertex with the highest estimated loss with a new vertex generated via an affine geometric transformation (Nelder and Mead, 1965). Specifically, the ‘losing’ vertex is reflected through the hyperplane defined by the remaining D vertices. The initial simplex is the main way in which the algorithm explores the extent of the parameter space, and repeated vertex reflection operations tend to cause the simplex to rapidly converge on a local volume of the parameter space. This global-to-local nature of the search echoes another major class of optimisation algorithms known as *simulated annealing* (SA) (Kirkpatrick et al., 1983), though in SA care is taken to control the rate at which solutions converge; this is not the case with NMS.

Although the NMS algorithm benefits from having few hyperparameters requiring tuning, its suitability for ABM calibration is challenged by the fact that the NMS method is an unconstrained optimisation procedure. Consequently, vertex reflection transformations can easily result in points that are outside the permissible region of the parameter space. This can result, for example, in the selection of negative values for parameters governing the number of agents in the simulation. We deal with this problem by introducing penalty terms in the NMS objective function that are linear

in the absolute distance by which a given parameter lies outside the feasible range. We observed this method to be effective at constraining solutions to the permitted region.

Table 4.7 presents the hyperparameters of the NMS method. The value of the hyperparameters were determined following a preliminary parameter recovery exercise, as described in Section 4.6.1.

Hyperparameter	Value
Monte Carlo evaluations I	8
Parameter vector sample budget N	14600
Optimisation repeats	20

Table 4.7: Hyperparameters of the Nelder-Mead simplex algorithm

4.6.4 Genetic Algorithm

Genetic algorithms (GAs) (Holland, 1992; Goldberg, 2006) employ the natural metaphor of an evolutionary process, whereby a *population* of G candidate parameter vectors $(\theta_1, \dots, \theta_G) \subset \Theta$ (*individuals* in GA terminology) compete for the opportunity to pass their parameters (*genes*) onto later *generations*. GAs, like random search, do not require analytic gradients, and are able to accommodate non-convexity and non-smoothness in the loss function.

Genetic algorithm optimisation is subject to one or more *genetic operators*, typical examples of which are:

- **Selection** — individuals are selected with probability proportional to their *fitness*; fitness corresponds to $-\hat{\mathcal{L}}(\theta)$ in our framework.
- **Recombination** — the parameter vectors of two (or more) selected individuals are combined using *crossover*. Child individuals inherit copies of contiguous blocks of parameters from each parent. This block-based inheritance facilitates exploitation by preserving beneficial parameter constellations.
- **Mutation** — parameter space exploration is facilitated by re-sampling (a typically small fraction of) the child’s parameters uniformly across their full domains, before the child enters the population.

Genetic algorithm hyperparameters are given in Table 4.8. As with the other methods, hyperparameter selection was guided by a preliminary exercise where parameters were recovered from synthetic data generated using the ABM with known parameters

(see Section 4.6.1). The number of ABM evaluations performed per generation is not constant under the present implementation. The selected values of 500 generations with a population of 50 individuals yields a mean of $N = 14600$ ABM evaluations, facilitating unbiased comparison with other calibration methods. Increasing the population size up to 250 individuals did not lead to significant improvements in $\hat{\mathcal{L}}(\theta)$ for a fixed computational budget.

Hyperparameter	Value
Monte Carlo evaluations I	8
Number of generations	500
Population size G	50
Crossover rate	50%
Mutation rate	10%
Optimisation repeats	20

Table 4.8: Hyperparameters of the genetic algorithm.

Each generation of our GA implementation¹⁵ gives rise to a set of independent ABM evaluations, which are performed in parallel. This approach is efficient provided all evaluations take similar wall clock times, however if a subset of evaluations take significantly longer than others, this will result in idle CPU cores until all ABM evaluations in the generation have concluded. Selection of an efficient parallelisation strategy depends on the number of available CPU cores and the number of Monte Carlo samples I used to evaluate each parameter vector.

For the reasons outlined in Section 4.6, the GA implementation evaluated in the present chapter is as standard as possible (a single-point crossover, tournament algorithm (Goldberg, 2006)). An extensive literature is available on genetic algorithm approaches, and it is indeed possible that some alternative variant would outperform the baseline method trialled here.

4.6.5 Markov Chain Monte Carlo

Bayesian methods, including Markov chain Monte Carlo (MCMC), have received recent attention within the ABM community (Grazzini et al., 2017; Lux, 2018). As ABM calibration moves onto a more quantitative footing, it is necessary to account for uncertainty in parameter estimates, and while one way to achieve this is by performing multiple independent optimisation trials (e.g., Platt and Gebbie (2018)), MCMC

¹⁵Based on the DEAP library <https://deap.readthedocs.io/en/master>, accessed 2018-08-13. License: GNU LGPL 3.0.

offers an alternative approach to accounting for parameter uncertainty by directly estimating the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$. From Bayes' theorem (Bishop, 2006)

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}, \quad (4.24)$$

where $p(\mathbf{y} \mid \boldsymbol{\theta})$ is the *likelihood* of the data ($\mathbf{y}$) given $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$ is our *prior* expectation about the value of the parameters, and $p(\mathbf{y})$ is a normalising constant known as the *evidence*. However, for complex, high-dimensional models, such as ABMs, the computation of $p(\mathbf{y})$ becomes intractable. MCMC overcomes this difficulty by approximating the posterior as the limit of a sampling process consisting of a sequence of points $(\boldsymbol{\theta}_t) \subset \boldsymbol{\Theta}, t \in (1, \dots, N)$. The points satisfy the Markov property, i.e., the probability of selecting a new point $\boldsymbol{\theta}_t$ is conditional only on the value of its immediate ancestor: $p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_{t-2}, \dots, \boldsymbol{\theta}_1) = p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1})$. The set of such points thus forms a *Markov chain* in the parameter space. New points $\boldsymbol{\theta}_t$ are selected for evaluation according to a *proposal distribution* which incorporates the prior. Acceptance (i.e., addition to the head of the chain) is a function of the relative likelihoods $p(\mathbf{y} \mid \boldsymbol{\theta}_t)$ and $p(\mathbf{y} \mid \boldsymbol{\theta}_{t-1})$. There are many alternative methods for managing the proposal and acceptance of candidate parameter vectors. One of the best known is the Metropolis-Hastings (MH) algorithm (Robert and Casella, 1999; Grazzini et al., 2017). In the MH algorithm, a new candidate point $\boldsymbol{\theta}_t$ is accepted with probability proportional to $p(\mathbf{y} \mid \boldsymbol{\theta}_t)/p(\mathbf{y} \mid \boldsymbol{\theta}_{t-1})$. The stationary distribution of the sequence of points obtained according to this criterion can be shown to converge to the full posterior distribution $p(\mathbf{y} \mid \boldsymbol{\theta})$. The possibility of accepting a point with lower likelihood allows the Markov chain to avoid becoming trapped in local minima, and suggests that this method may be well-suited to the required properties of support for non-convexity, and non-smoothness in the optimisation objective. The relationship between loss and likelihood is discussed in detail below.

Several factors affect the efficiency of MCMC convergence (Korattikara et al., 2014). A large step size (used to sample from the proposal distribution) may cause the chain to prematurely exit areas of high likelihood, increasing the number of evaluations required for convergence, as a large fraction of candidate points will be rejected. Conversely, with a small step size adequate exploration of the extent of high-likelihood regions may be very slow, leading to incomplete representation of the posterior (for example, this could be particularly troublesome if the posterior is multi-modal, and only one mode is sampled by the algorithm). Furthermore, if the initial location of the chain, $\boldsymbol{\theta}_1$, is far from a region of high likelihood relative to the proposal step size,

then it may be necessary to discard many initial samples as it can take a long time for the chain to arrive at a region of high likelihood, i.e., a region that is representative of the target posterior. This uninformative initial phase is known as *burn-in* (Robert and Casella, 1999). The recommended approach to ensuring full coverage of the posterior is therefore to sample multiple independent Markov chains, with a commensurate increase in computational cost (Korattikara et al., 2014).

In order to use MCMC for ABM calibration, it is necessary to specify several key items:

1. The likelihood function, $p(\mathbf{y} \mid \boldsymbol{\theta})$,
2. The prior, $p(\boldsymbol{\theta})$,
3. The proposal distribution and acceptance criteria.

A major difference between MCMC and the other methods considered in this section is that with MCMC we do not directly compute $\hat{\mathcal{L}}(\boldsymbol{\theta})$. Instead, and equivalently, we represent the empirical moment distributions as a joint multivariate Gaussian distribution. The likelihood of the model output conditional on a parameter vector, $p(\mathbf{y} \mid \boldsymbol{\theta})$, is then trivially found by computing the probability of $\mathbf{y}$ with respect to the multivariate Gaussian. The likelihood $p(\mathbf{y} \mid \boldsymbol{\theta})$ is maximised when the vector of simulated moments $\boldsymbol{\mu}(\mathbf{y})$ coincides with the expected value of the multivariate empirical moment distribution. Thus, by searching for parameter vectors with high likelihood, this construction ensures that, in our implementation, MCMC equivalently searches for parameter vectors with low $\hat{\mathcal{L}}(\boldsymbol{\theta})$.

The ability to specify the *prior*, $p(\boldsymbol{\theta})$, is one of the hallmarks of the Bayesian approach, and provides practitioners with a powerful mechanism for expressing knowledge about the problem domain before the acquisition of data. If no relevant parameter knowledge is available, as in our case, it is necessary to specify *uninformative* priors. In the present usage this corresponds to using uniform distributions over the parameter ranges given in Table 4.3. Informative priors, if available, would be expected to enhance convergence of the MH algorithm, and other MCMC variants.

We adopt the Metropolis-Hastings (MH) MCMC algorithm for our proposal distribution and acceptance criteria. Other simpler Bayesian alternatives have been explored in the ABM literature e.g., rejection sampling and importance sampling (Grazzini et al., 2017). However, MCMC is known to scale well to high-dimensional spaces due to its ability to remain within regions of high likelihood (Bishop, 2006; Betancourt, 2017). More sophisticated MCMC approaches such Hamiltonian Monte

Carlo (HMC) (Betancourt, 2017) are precluded for our purposes by the unavailability of analytic gradients and also by the need to support discrete parameters. Due to these limitations, as well as a general lack of awareness of Bayesian approaches, and the computational cost associated with them (Grazzini et al., 2017), attempts to bring such advanced Bayesian methods to bear on ABM calibration have to date been limited, e.g., Bertschinger et al. (2018).

Table 4.9 presents the hyperparameters of the MH algorithm, selected with reference to the preliminary parameter recovery exercise (Section 4.6.1). Our MH implementation¹⁶ automatically tunes per-parameter proposal step sizes to maintain a reasonable acceptance rate for new parameter vectors. It is necessary to specify the interval for this automated tuning, as well as the initial step size scale which we set to 10% of the parameter range for each parameter. The MH algorithm takes steps for a single parameter at a time, and during our parameter recovery exercise we observed that this resulted in excessive re-evaluations of the most recent accepted point θ_{t-1} . To reduce wasteful repeated computation, we implemented a *caching* function which, upon revisiting a parameter vector, returns the stored likelihood with probability $(1 - \alpha)$ and re-computes the likelihood with probability α . Based on our preliminary exercise we determined $\alpha = 10\%$ to be appropriate for the *cache miss rate*. As $I \rightarrow \infty$ we would expect $\alpha \rightarrow 0$ to be a reasonable approach.

Hyperparameter	Value
Monte Carlo evaluations I	8
Parameter vector sample budget N	14600
Priors	uniform
Step method	Metropolis
Step size scaling interval	100
Step size scale initial	10%
Cache miss rate	10%
Optimisation repeats	20

Table 4.9: Hyperparameters of the Markov chain Monte Carlo algorithm.

4.6.6 Other Alternatives

Gaussian Process Regression has been developed to facilitate expensive black-box optimisation problems, but in our testing the approach did not scale well beyond around 1000 data samples, taking an excessive time to select novel samples for evaluation in

¹⁶Based on PyMC3 <https://docs.pymc.io/>, accessed 2018-07-11, License: Apache v2.

the ABM. Lamperti et al. (2018, p.367) claim that GPR is, in general, unsuitable for ABM calibration: “when the spatial distribution of the data is unknown, as is often the case with ABMs, [GPR] requires specialist knowledge of variography to empirically estimate the spatial dependence of the data. This generally requires a large number of ABM evaluations and extensive exploratory data analysis that increases with the size of the parameter space.” In our testing we observed the expected $\mathcal{O}(n^3)$ dependence on the number of training data points (Bishop, 2006).

Lamperti et al. (2018) propose an alternative method using *boosted gradient trees* as an underlying surrogate model. However, comparison with GPR is not straightforward, as Lamperti et al. (2018)’s goal is to find all regions of the parameter space giving rise to positive calibrations, rather than finding a single optimal parameter vector, as is the case in GPR¹⁷. In extensive testing we were unable to successfully employ Lamperti et al. (2018)’s method. There appeared to be two main reasons for this. Firstly, the reference implementation provided by the authors makes use of GPR for hyperparameter optimisation and so suffers from the same $\mathcal{O}(N^3)$ scaling problems as the pure GPR method. Secondly, it is necessary to initialise Lamperti et al. (2018)’s surrogate model with at least one known positive calibration. Where such parametrisations are unknown (as in our case) this is infeasible without recourse to a preliminary alternative calibration stage. We attempted to overcome this difficulty by initialising the surrogate model with positive calibrations located using the genetic algorithm method. Although this prevented the surrogate model from producing degenerate solutions (e.g., predicting that no positive calibrations exist), we found that (1) the surrogate was unable to improve on the value of $\hat{\mathcal{L}}(\theta)$ obtained by the genetic algorithm method, and (2) the surrogate exceeded available system memory as soon as it subsequently attempted to optimise hyperparameters. In any case, the computational budget, for comparison purposes, had already been consumed by the GA preliminary stage. Lamperti et al. (2018) do not exceed a budget of 2500 evaluations, perhaps reflecting similar difficulties encountered by those authors.

We made one further attempt to use Lamperti et al. (2018)’s surrogate method. We employed a partially trained surrogate model (subject to the constraint on the number of training examples due to the GPR limitation), as a low-cost source of estimates of $\hat{\mathcal{L}}(\theta)$ embedded within the genetic algorithm method. This attempt was unsuccessful in that the final obtained values of $\hat{\mathcal{L}}(\theta)$ were not competitive with

¹⁷Our GPR implementation uses the standard *expected improvement* acquisition function that proposes points for evaluation that minimise expected loss at the candidate location. Alternative acquisition functions that instead maximise information gain, e.g., Bryan et al. (2006) may be more comparable.

the genetic algorithm, even when both methods were allowed the same number of ABM evaluations (i.e., treating queries to the trained surrogate as computationally cost-free). Queries to the surrogate were, however, orders of magnitude faster than evaluations of the ABM — a very attractive feature for ABM practitioners if previously mentioned limitations, i.e., the $\mathcal{O}(N^3)$ dependence and need to pre-populate the algorithm with positive calibrations, could be overcome. It could well be the case that Lamperti et al. (2018)’s method outperforms GPR for ABM calibration with modest computational budgets, but we did not proceed further with either method.

4.6.7 Selection of Loss Function for Calibration

In Section 4.1.3 we presented two alternative loss functions:

1. Quadratic loss, $\hat{\mathcal{L}}(\boldsymbol{\theta})$ (Equation 4.10)
2. Joint moment coverage ratio loss, $\mathcal{L}_{\mu}^{jMCR}(\boldsymbol{\theta})$ (Equation 4.16)

Our approach is to use (1) during calibration, and (2) during validation. The selection of (2) for validation is justified by the principled way in which this function incorporates the statistical significance of the model’s ability to produce output that is consistent with reality (Section 4.1.3). However, the question of which loss function should be selected for use during calibration warrants further discussion. Although Franke and Westerhoff (2012) claim that the jMCR is unfit for this purpose due to the high computational cost involved, Braun-Munzinger et al. (2018) do indeed make use of the jMCR in this way. We would additionally argue that utilising the form of $\hat{\mathcal{L}}(\boldsymbol{\theta})$ given in equation 4.10 avoids introducing artificial plateaux into the objective function landscape: the number of observable levels of jMCR is equal to I , the number of Monte Carlo replicates of each parameter vector $\boldsymbol{\theta}$. As previously discussed, for computational cost reasons, we fix $I = 8$. While this value was found to provide a sufficiently accurate estimate of $\hat{\mathcal{L}}(\boldsymbol{\theta})$, the $\mathcal{L}_{\mu}^{jMCR}(\boldsymbol{\theta})$ landscape would consist of just eight levels, each of which is a perfectly flat isosurface within the parameter space. This feature would hamper the exploitation of local parameter space structure except for regions in the vicinity of a jMCR plateau boundary (which would manifest as a discontinuous jump in $\mathcal{L}_{\mu}^{jMCR}(\boldsymbol{\theta})$). Although optimisation procedures such as genetic algorithms have been shown to successfully perform *neutral search* in such scenarios (Ebner et al., 2001), it seems that without making I infeasibly large, discarding the information provided by the loss function $\hat{\mathcal{L}}(\boldsymbol{\theta})$ would constitute a poor design choice.

An additional justification for measuring jMCR after the conclusion of the calibration process (i.e., exclusively as a validation step), is that the jMCR then represents a type of out-of-sample test, and an extra level of validation for the model and calibration approach. For example, it could help avoid situations where most simulated moments closely match their empirical counterpart but some moments are poorly matched by the model, and lie outside their empirical confidence intervals. Such solutions may be accepted under the optimisation of $\hat{\mathcal{L}}(\boldsymbol{\theta})$ with a threshold τ based on the percentiles of the empirical loss distribution (Equation 4.12), but would go on to fail the jMCR check. Conversely, using jMCR alone could lead to solutions where all simulated moments are ϵ inside their empirical confidence intervals, with no way for the optimisation algorithm to detect, or improve on, this situation. This flattening of the peak of $\mathcal{L}_{\mu}^{jMCR}(\boldsymbol{\theta})$ would add variance to resulting parameter estimates. It would be possible to quantify this variance by performing repeated calibration trials with different values of I , though this would be computationally expensive.

We could also envisage a hybrid loss function that incorporates the Euclidean distance between simulated and empirical moments as in equation 4.10, *as well as* the jMCR. This would require the specification of a hyperparameter, ψ controlling the relative weight of the two sources of loss, thus

$$\mathcal{L}_{\mu}^{hybrid}(\boldsymbol{\theta}) = \mathcal{L}_{\mu}^{jMCR}(\boldsymbol{\theta}) + \psi \hat{\mathcal{L}}(\boldsymbol{\theta}). \quad (4.25)$$

Such a compound loss function could mitigate the problems with plateaux introduced at modest numbers of Monte Carlo replicates I . However, this would be at the expense of introducing a new hyperparameter to tune, and would remove the out-of-sample benefit of a second stage validation procedure provided by a separate application of the jMCR metric. In preliminary calibration experiments we found that $\mathcal{L}_{\mu}^{hybrid}(\boldsymbol{\theta})$ performed slightly better than $\hat{\mathcal{L}}(\boldsymbol{\theta})$, though it was not straightforward to select an appropriate value for ψ . In order to avoid introducing unnecessary complexity into our methodology, in the present work we retain the two-phase approach using $\hat{\mathcal{L}}(\boldsymbol{\theta})$ for calibration and the jMCR exclusively for validation.

4.6.8 Quantitative Calibration Method Comparison

Figure 4.6 presents the results of calibrating the ABM using data from the first quarter of 2017, for assets in the central market cap band (band 2), for each optimisation method. This band and date were chosen because in preliminary experiments this

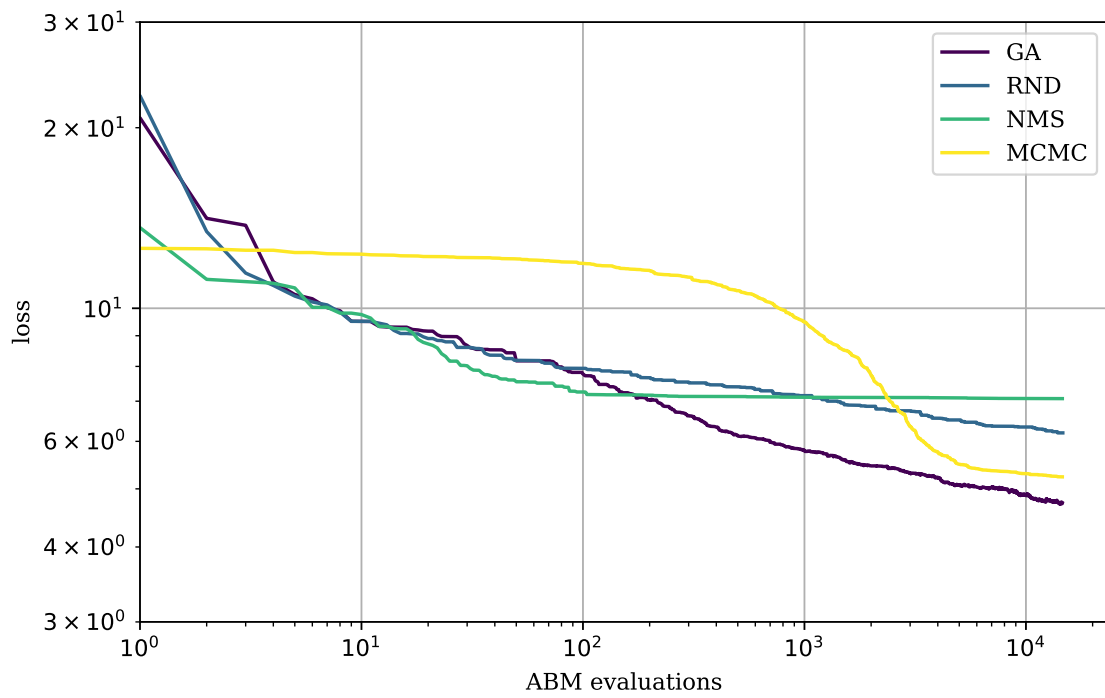


Figure 4.6: Comparison of calibration methods showing the best achieved loss $\hat{\mathcal{L}}(\theta)$ as a function of the number of ABM evaluations, on a log-log scale. GA: genetic algorithm; RND: random search; NMS: Nelder-Mead simplex; MCMC: Markov chain Monte Carlo. We present the mean and 95% confidence interval from 20 independent trails for each method. In each case, the model was calibrated against data from 2017-Q1 in the central market cap band (band 2). This band-date pair was found to be particularly challenging for successful model calibration.

combination proved to be the most challenging for successful calibration (we discuss reasons for this in Section 4.7). The figure presents the mean with 95% confidence interval from 20 independent optimisation trials for each method. The lowest encountered value of $\hat{\mathcal{L}}(\boldsymbol{\theta})$ is plotted as a function of the number of ABM evaluations, $n \in (1, \dots, N)$. The data are shown on a log-log scale in order to enhance readability. For a fixed computational budget of 14600 ABM evaluations, it is clear from Figure 4.6 that genetic algorithms offer the best performance, followed fairly closely by Markov chain Monte Carlo, with random search and finally the Nelder-Mead simplex method providing significantly poorer performance.

The baseline method, random search, exhibits expected monotonic improvement in the lowest encountered value of $\hat{\mathcal{L}}(\boldsymbol{\theta})$ as a function of n in Figure 4.6, though the final value obtained after the computational budget of ABM evaluations has been exhausted (i.e., $n = N$) is not competitive with the genetic algorithm or Markov chain Monte Carlo methods. However, for $n > 1000$ random search outperforms the Nelder-Mead simplex method, and for $n < 2500$, random search outperforms Markov chain Monte Carlo. Both of these observations are connected to the exploration-exploitation trade-off. The dominance of NMS over random search for $30 < n < 1000$ suggests that there is indeed local structure to the parameter space $\boldsymbol{\Theta}$ that is available for exploitation by sampling points in the neighbourhood of previously evaluated candidates. The selection of the initial simplex for the NMS method is identical to random search, and vertex reflection operations for small n result in large changes to the simplex that are again indistinguishable from random search. It is not until $n > 30$ that the performance of NMS can be differentiated from random search, at which point NMS begins to dominate. However, NMS clearly exhibits convergence to a local minimum around $n = 100$, with the value of $\hat{\mathcal{L}}(\boldsymbol{\theta})$ remaining fixed thereafter. By construction, random search must eventually encounter a parameter vector with a value of $\hat{\mathcal{L}}(\boldsymbol{\theta})$ below that obtained by any previously converged NMS trial, and indeed this seems to occur around $n = 1000$. The premature convergence of NMS was anticipated, and is consistent with observations by Platt and Gebbie (2018) who find that even augmenting the NMS method with a threshold-accepting criterion is insufficient to avoid early convergence.

The dominance of random search over Markov chain Monte Carlo for $n < 2500$ can be attributed to the wasteful *burn-in* stage of Markov chain Monte Carlo sampling (see Section 4.6.5). It would appear from Figure 4.6 that it takes hundreds of evaluations before Markov chain Monte Carlo encounters parameters that appreciably improve

on the lowest encountered value of $\hat{\mathcal{L}}(\boldsymbol{\theta})$ ¹⁸. During this early phase of calibration the Markov chain performs a random walk in low-likelihood regions of the parameter space Θ , which constitutes slow exploration. However, there is limited signal available for exploitation in this regime. Eventually, higher-likelihood regions are explored and for $1000 < n < 5000$ the Markov chain Monte Carlo algorithm exhibits by far the fastest improvement in $\hat{\mathcal{L}}(\boldsymbol{\theta})$ of any of the methods under investigation, as local structure in the parameter space is efficiently exploited. This sudden and marked change in the efficacy of the method underscores the need for practitioners to perform preliminary exercises to guide the selection of optimisation methods. At the end of the optimisation run, the Markov chain Monte Carlo method attains a value of $\hat{\mathcal{L}}(\boldsymbol{\theta})$ that would require orders of magnitude more evaluations of the random search algorithm¹⁹.

Genetic algorithms perform effective exploration during the early stages of optimisation. Figure 4.6 indicates that genetic algorithms are indistinguishable from random search for $n < 100$. During early generations of the genetic algorithm (for $n < 110$), the approach is dominated by NMS. This finding suggests (with limited statistical significance) that NMS is able to effectively exploit local structure more efficiently than genetic algorithms when the GA is restricted to a small number of generations (e.g., two or three generations). However, the ability of the genetic algorithm to continue to introduce novel solutions by exploring the parameter space using mutation operations means allows the GA to avoid premature convergence, and for $n > 110$ the genetic algorithm method dominates all others.

It is interesting that MCMC is unable to match the performance of genetic algorithms for our market microstructure ABM. This could be due to the complex interactions between parameters revealed by the sensitivity analysis in Section 4.5. Ensuring adequate *mixing* (i.e., exploration of the high-likelihood regions of Θ) is a known difficulty with MCMC, as discussed in Section 4.6.5. Increasing the number of Markov chains is one approach to mitigating this, however, this would introduce a multiplicative factor into computational costs. Platt and Gebbie (2018) found that introducing the ability for a local method (in their case, NMS) to exit local minima was insufficient to avoid premature convergence. It could be the case that for both Platt and Gebbie (2018) and the MCMC method tested in this chapter, it is more likely that a local search will return to a proximate region of high likelihood

¹⁸Note that for Markov chain Monte Carlo, minimum loss corresponds to a *maximum a posteriori* (MAP) parameter estimate.

¹⁹Observation of Figure 4.6 suggests it would take 500,000–600,000 ABM evaluations for the random search method to attain an equivalent value of $\hat{\mathcal{L}}(\boldsymbol{\theta})$ to that attained by Markov chain Monte Carlo in 14,600 evaluations.

(low $\hat{\mathcal{L}}(\boldsymbol{\theta})$) rather than exploring potentially disconnected and distant regions of parameter space. By continuing to introduce novelty through such exploration, genetic algorithms appear to be more effective at exploring alternative minima in the loss landscape.

The results presented in this section indicate that an intermediate approach to the exploration-exploitation trade-off gives the best quantitative performance for calibration of the microstructure ABM. There are two reasons why this is the case. Firstly, the objective function $\hat{\mathcal{L}}(\boldsymbol{\theta})$ is subject to noise due to its stochastic nature. Noise unavoidably introduces non-convexity and discontinuity into the loss function. Even in the absence of noise (i.e., as $I \rightarrow \infty$, which rapidly becomes computationally intractable), there are no guarantees that $\hat{\mathcal{L}}(\boldsymbol{\theta})$ is convex. Due to the complex, non-linear interactions between agents in the model, local minima are to be expected. For this reason, methods with a strong bias towards exploitation, like the Nelder-Mead simplex method, can readily converge to a suboptimal solution. Other practitioners have augmented NMS with the ability to escape local minima, and our results indicate that this would be necessary for the class of problems that are our present focus. Secondly, our results also indicate that Markov chain Monte Carlo methods struggle due to their preference for exploitation. However, if the low level of exploration employed by MCMC methods locates a region of higher likelihood, this region is rapidly and successfully exploited. This indicates that there is indeed local structure available for exploitation. Genetic algorithms are able to exploit this local structure, whilst continuing to perform effective exploration, and so for microstructure ABM calibration, our results suggest that the exploration-exploitation trade-off employed by genetic algorithms is an effective balance, and one that consistently yields high-quality solutions.

We adopt genetic algorithms as our calibration method for the remainder of this chapter.

4.7 Longitudinal Calibration and Validation Results and Discussion

The method comparison in the preceding section resulted in the identification of genetic algorithms as the most effective approach for the calibration of the market microstructure component of our ABM. In this section, we use genetic algorithms for an extensive calibration experiment, incorporating a comprehensive dataset of asset

trade prices and volumes re-sampled to a time-resolution of 10 seconds, between 2009–2018 inclusive, and for all liquid US equities (i.e., those with a market cap in excess of \$100mn). Section 4.2 gives a full discussion of these data and their preparation for inclusion in our calibration approach.

Due to the extensive scope of the calibration, computational resource management and efficient use of a parallel HPC cluster was of paramount importance. Preliminary experiments indicated that calibration for all five market cap bands, for a single date (e.g., 2010-Q1) would take approximately 50 hours on the 5×56 logical-core machines in our HPC cluster, with the same 20 independent realisations and $I = 8$ Monte Carlo replicates per evaluation as used in the method comparison in Section 4.6 (see Table 4.8 for an overview of the hyperparameters of the genetic algorithm method). We found that 250 generations ($N \approx 7300$) were almost always sufficient to obtain a positive validation, and we discuss this further below. At 50 hours per date, it was not feasible to calibrate all 40 quarters between 2009-Q1–2018-Q4. For this reason we calibrate only the first of each year, i.e., (2009-Q1, 2010-Q1, ...). This selection resulted in approximately 6×10^7 evaluations of the ABM, and took 26 days on our 5×56 core HPC cluster. The decision to select Q1 of each year, rather than Q2–Q4, ensures that we do not attempt to calibrate the model using data from the day of the 6th May 2010 Flash Crash. Thus, any tendency for the model to exhibit flash crashes can be considered a pure out-of-sample test, particularly so for 2009-Q1–2010-Q1 inclusive (i.e., before the Flash Crash of 6th May 2010). We return to this issue in Section 4.10.

Figure 4.7 presents the longitudinal calibration results for all market cap bands, from 2009-Q1–2018-Q1. The solid lines indicate the maximum joint moment coverage (the fraction of 100 realisations of the ABM in which simulation output falls with the empirical confidence interval for all ten moments jointly) attained from 20 independent calibrations for each band-date pair, along with 95% confidence intervals. As discussed in Section 4.1.3, we reject the model output $\mathbf{y}(\boldsymbol{\theta})$ as compatible with the rwDGP ($\mathbf{y}_R$) if the jMCR is less than 5%, with 95% confidence. The horizontal dashed line on Figure 4.7 indicates this 5% validation threshold. Only market cap band 2 (the central band) in 2017-Q1 is below the validation threshold with 95% confidence. A positive validation was subsequently achieved for this band-date pair by increasing the number of generations in the genetic algorithm from 250 to 500 ($N \approx 14600$) and repeating the calibration procedure.

The upper dotted lines in Figure 4.7 show the alternative validation metric based on the empirical loss distribution $\mathcal{P}_{\mathcal{L}}$ (Equation 4.12). These results show that, as

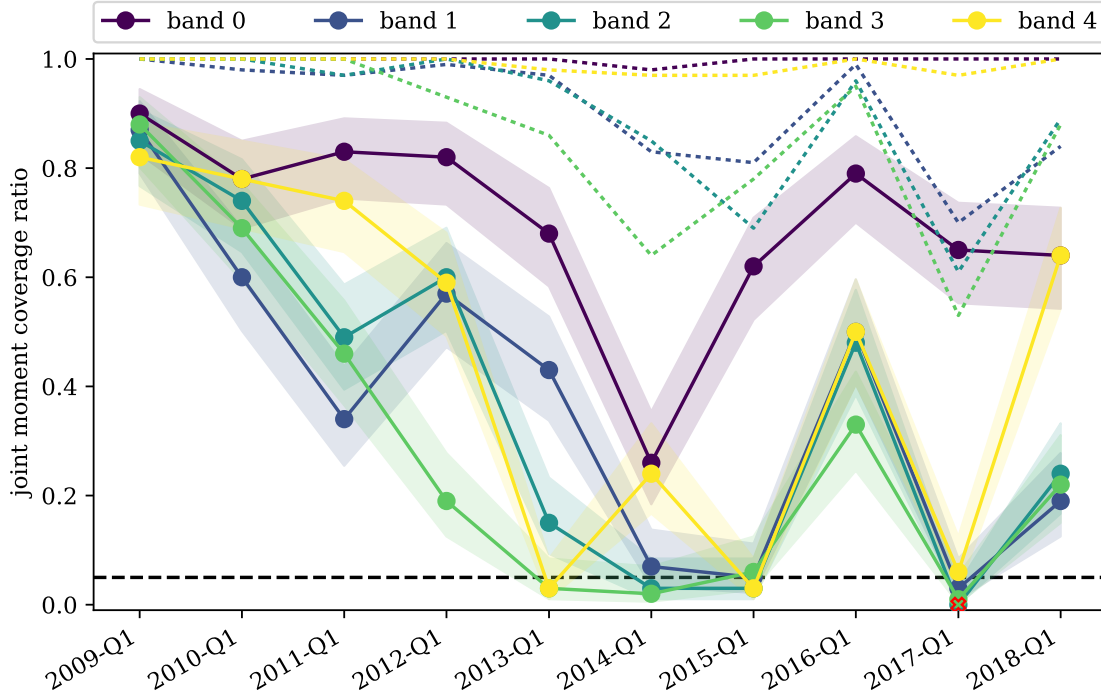


Figure 4.7: **Solid lines:** maximum joint moment coverage ratio (jMCR), with 95% confidence interval, achieved from 20 independent calibration trials using the genetic algorithm method for each market cap band from 2009-Q1 to 2018-Q1. Calibrated models are rejected as compatible with the real-world data generating process if jMCR is below 5% with 2σ (95%) confidence. Such instances are marked with a red cross (band 2 in 2017-Q1). **Dashed line:** 5% jMCR validation threshold. **Dotted lines:** maximal obtained values of the alternative validation metric $\mathcal{P}_{\mathcal{L}}$ (confidence intervals suppressed for clarity).

expected, jMCR is a more stringent test of validation than $\mathcal{P}_{\mathcal{L}}$. Given that jMCR and $\mathcal{P}_{\mathcal{L}}$ are both calculated by performing I Monte Carlo replications of the ABM for parameter vector θ and comparing the output to empirical moments in the case of jMCR, and to the empirical loss distribution in the case of $\mathcal{P}_{\mathcal{L}}$, it would seem to be preferable to use jMCR as a validation criterion, as our results indicate that a successfully validated model according to jMCR will also be considered successful under $\mathcal{P}_{\mathcal{L}}$. We would also argue in favour of a rigorous quantification of uncertainty in jMCR, using an appropriate approximation to the binomial confidence interval given the size of the sample, as discussed in Section 4.1.3.

Several trends are apparent in Figure 4.7. Firstly, the highest jMCR for all bands is achieved in 2009-Q1, with 2010-Q1 in second place, again for all bands. This result is in line with expectations. The structure of the ABM and the selection of the

type of agent classes represented in the model, as well as the properties of the LOB market mechanism, were based on surveys carried out by [CFTC and SEC \(2010\)](#), and originally developed into an ABM by [Paddrik et al. \(2012\)](#). The Flash Crash of 6th May 2010 was a watershed moment in electronic financial markets, and it is not, therefore, surprising that market participants might modify their algorithms post 2010, in light of the newly identified systemic risk. What is much more surprising is that we have shown that it is possible to successfully calibrate and validate the same model in all years since the Flash Crash. Inspection of the parameters obtained in each year may therefore convey information about how the structure of the market has changed since 2010, and we discuss this in detail below.

The second major trend in [Figure 4.7](#) is that 2013–2015 and 2017 seem to present a greater challenge to successful calibration and validation than other years. We have already discussed why model performance might be expected to diminish in the years following 2010. In order to understand the reason for the difficult calibration environment in 2013–2015 and 2017 we inspected the moment coverage ratio on a per-component basis. Recall that the joint moment coverage ratio is the fraction of ABM evaluations during which simulated moments lay jointly within all empirical confidence intervals. Where jMCR falls below the validation threshold, we can decompose the jMCR into per-moment MCR ratios. When we do this we find that validation failure is most often attributable to a mismatch between simulated and empirical log-liquidity, with failures attributable to a lesser extent to mismatched volatility (standard deviation of price returns) and the first lag of price return autocorrelation. We observed a marked asymmetry in cases where validation failure is due to mismatched log liquidity. It was never the case that the simulated log liquidity was too low compared to empirical observations; conversely the simulated log liquidity was only ever observed to lay within, or above, the empirical confidence interval. We attribute this asymmetry to the use of the logarithm function. Log-transforming liquidity before incorporating the moment into $\hat{\mathcal{L}}(\theta)$ induces an asymmetric penalty. For a given relative difference between simulated and empirical liquidity, a positive relative difference (i.e., simulated > empirical liquidity) would contribute a smaller addition to $\hat{\mathcal{L}}(\theta)$, than for a negative relative difference. Our results suggest that avoiding the log transformation may improve the likelihood of a positive validation. However, computational resource constraints prevented us from testing this hypothesis. It could, furthermore, be the case that calibration and validation in the lowest and highest market cap bands (bands 0 and 4, respectively) is made less meaningful

without the log transformation, due to the wide range of market capitalisations represented by the assets in the outer bands. Empirical confidence intervals on non-log liquidity could become extremely large, and subsequently fail to adequately constrain parameter estimates.

The final trend to emerge from Figure 4.7 is that calibration attempts for the lowest market cap band were generally the most successful, compared to the other bands. Low cap stocks are generally known to exhibit higher volatility (standard deviation of price returns) than other stocks (Fama and French, 1993). This could suggest that the model is better suited to representing high-volatility scenarios. This would not be surprising, given, as discussed above, the original version of the model was developed in the aftermath of the 2010 Flash Crash, which itself occurred shortly after the 2007-2009 financial crisis — a high volatility period (Paddrik et al., 2012). The CBOE volatility index (VIX) level,²⁰ the industry standard benchmark for US equity market volatility, fell to its lowest levels since the financial crisis during 2013–2015, and again during 2017, corresponding to the more challenging periods for calibration of higher market cap bands in our model. We stress, however, that positive validations were indeed achieved under low volatility market conditions, and in all market cap bands, albeit with higher computational cost reflecting greater difficulty for the genetic algorithm to locate successful parameter vectors. Another possible explanation is that the trend is a consequence of the way in which assets were assigned to bands. As discussed in Section 4.2, each market cap band (for a given date) is assigned an equal number of assets. If empirical moments exhibit central tendency across the market cap spectrum, then it is to be expected that assets in the outer market cap bands exhibit higher moment variance than assets in inner market cap bands. This would imply that inner market cap bands present a more stringent test of calibration performance. As stated by Winker et al. (2007, p.132) “a smaller variance of the estimates increases the probability to reject the null hypothesis that a moment generated by simulation comes from the same distribution as the actual data.” It is possible that the effects underlying both explanations for the trend in Figure 4.7 are simultaneously present.

²⁰Source: Yahoo Finance. <https://tinyurl.com/yahoo-finance-vix-2009-2019>, accessed 2019-02-19.

4.8 Out-of-Sample Validation Results

The availability of market data over an extended time period (2009-Q1–2018-Q4 inclusive) allows us to test the out-of-sample (i.e., future) performance of the calibrated parameters. Specifically, we are interested in whether the parameter vector $\hat{\theta}$ calibrated using data from a given quarter $\mathcal{Q} \in (0, \dots, 36)$ (labelled such that $\mathcal{Q} = 0$ corresponds to 2009-Q1, and $\mathcal{Q} = 36$ corresponds to 2018-Q1) and giving rise to a positive validation for $\mathcal{Q}$, additionally gives rise to a positive validation when compared with data from $\mathcal{Q}' \in (\mathcal{Q} + 1, \dots, 39)$ (where $\mathcal{Q} = 39$ corresponds to 2018-Q4). The null hypothesis is that calibrated parameters only give rise to a positive validation when compared with the data used in their optimisation (i.e., for in-sample validation such that $\mathcal{Q}' = \mathcal{Q}$). A rejection of this null hypothesis would provide evidence that the model exhibits *predictive* capability, insofar as we would establish with statistical rigour that calibrating the ABM using data available at $\mathcal{Q}$ will allow us to generate simulated market activity that is representative of the real-world data generating process for future quarters $\mathcal{Q}' > \mathcal{Q}$.

Figure 4.8 presents the out-of-sample validation results for models calibrated in all market cap bands, for 2009-Q1–2018-Q1 inclusive (in-sample validation results for the same set of model calibrations are presented in Figure 4.7). We plot the mean with 95% confidence intervals of the joint moment coverage ratio from 100 evaluations of the ABM when comparing the output of models calibrated using data from $\mathcal{Q}$ against data from $\mathcal{Q}' > \mathcal{Q}$. Results are transformed by plotting the joint moment coverage ratio as a function of $\mathcal{Q}' - \mathcal{Q}$. This transformation allows us to investigate the characteristic timescale over which positive validation performance may persist. Because fewer out-of-sample evaluations become possible as $\mathcal{Q}$ increases, the number of evaluations contributing to observations for higher values of $\mathcal{Q}' - \mathcal{Q}$ tends toward zero. In other words, only data calibrated for $\mathcal{Q} = 0$ (2009-Q1) can be evaluated against data from more than 36 quarters into the future — this is because we only have available data up to 2018-Q4 (and we only calibrate the model for Q1 of any year). We thus expect sample variance to increase to the right in Figure 4.8, *ceteris paribus*.

The results presented in Figure 4.8 demonstrate that, with 95% confidence, positive validations are obtained in all market cap bands for some $\mathcal{Q}' > \mathcal{Q}$, and thus we reject the null hypothesis. Several trends are apparent in the figure. Firstly, as might be expected, performance (i.e., the likelihood of a positive validation) decreases as $\mathcal{Q}' - \mathcal{Q}$ increases. This is most apparent in bands 1, 2 and 3, though all bands exhibit

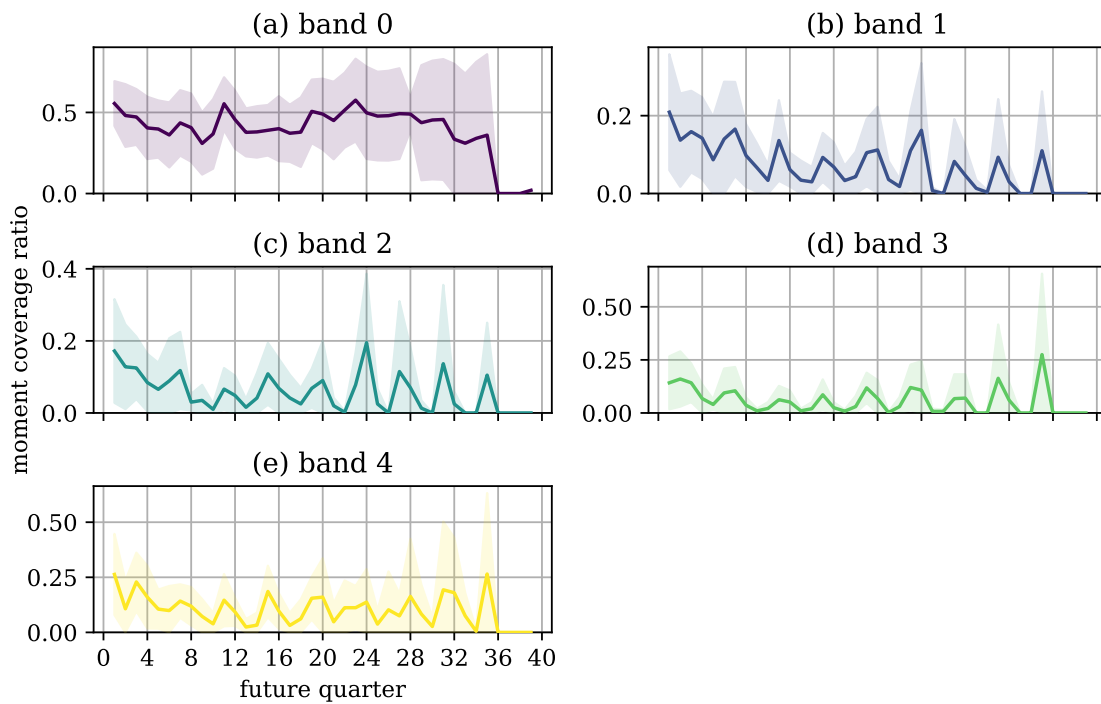


Figure 4.8: Out-of-sample validation of the calibrated ABM showing the joint moment coverage ratio obtained by using parameters calibrated against data from Q1 of each year from 2009–2018 inclusive, but comparing model output to data from future quarters 2009-Q2–2018-Q4 inclusive. The mean and 95% confidence interval is presented for each market cap band with vertical grid lines drawn every four quarters (i.e., annually).

this decreasing performance, and for all bands the maximum of the lower confidence bound is found at $Q' - Q = 1$. The apparent strong performance for $Q' - Q > 24$ (i.e., more than six years into the future) should be treated with caution due to the small sample size problem identified in the preceding paragraph. Indeed, we observe the expected increase in variance to the right of Figure 4.8, and this is apparent for all bands.

A second trend is that, as with the in-sample validation results (Figure 4.7), parameters in the lowest market cap band exhibit the strongest performance, and the strongest persistence in performance. This result may be partly explained by appealing to the argument from Section 4.7 regarding the relative variance of the different market cap bands, i.e., assets in band 0 vary more widely in their properties than do assets in more central bands, and so calibration is consequently subject to tighter constraints in central bands. A second possibility is that the properties of low market cap assets exhibit less longitudinal variance over the sample period compared to assets in other bands. This possibility is explored further in Section 4.9 with reference to longitudinal variation of the calibrated parameters themselves.

A final trend exhibited in Figure 4.8 is the striking four-quarterly (i.e., annual) periodicity that is apparent for all market cap bands, particularly for bands 1–4. Stock markets are well-known to exhibit seasonal (*calendar*) effects (Rozeff and Kinney, 1976; Gultekin and Gultekin, 1983), and these effects are known to be particularly strong in the month of January. In Section 4.7 we explained that our reason for calibrating against data from the first quarter of each year was to avoid using data from the Flash Crash of 6th May 2010, and so our Q1 samples incorporate data from each January during 2009–2018. However, it is not obvious that reported seasonality effects, which are primarily directional in nature (i.e., stock prices tend to rise in January (Rozeff and Kinney, 1976)), would also apply to the range of moments considered in our ABM such that the joint moment coverage ratio would exhibit a corresponding persistent fluctuation. It would be possible, if desired, to directly inspect the data in order to determine the relative degree of seasonality exhibited by each of the output moments. Though such an exercise may increase our understanding of asset markets, it is of peripheral importance to the present study: our objective is to demonstrate methodology sufficient to give rise to positive ABM validations in *any* market conditions, and to advance the calibration methodology. Moreover, the study of calendar effects is a long-established tradition in the econometric literature (Rozeff and Kinney, 1976).

4.9 Analysis of Calibrated Parameters

We now turn our attention to the calibrated parameters that resulted from the optimisation runs reported in Section 4.7. In the results that follow, we include the parameter vectors obtained from all calibration trials, regardless of whether the jMCR validation threshold was surpassed in that experiment. This improves the readability and signal to noise ratio in the following figures, without materially affecting the patterns revealed by the data.

Figure 4.9 presents the mean with 95% confidence intervals for values of the 17 parameters of the ABM obtained through 20 independent replications of the longitudinal calibration procedure introduced in Section 4.7. Please refer to Table 4.3 for a description of the parameter labels. For each parameter, we present the value obtained for each market cap band, and for the first quarter of each year from 2009-Q1–2018-Q1 inclusive.

Red markers on Figure 4.9 represent cases where the set of 20 independently calibrated parameter values for a given band-date pair are not statistically different from a uniform distribution across the parameter’s range (given in Table 4.3). We say that such parameters are not *identified* by the data for a band and date. This assessment is based on the critical value of the Kolmogorov-Smirnov (K-S) test statistic corresponding to $p < 0.05$. In cases where $p \geq 0.05$, indicated by red markers, our results suggest that the calibration procedure did not yield a statistically significant reduction in the feasible range for the parameter, for that band and date. This could be because (1) the number of calibration replicates is not sufficient (due to computational cost, we restricted this number to 20); (2) the parameter has no effect on the simulation output (however, we established via extensive sensitivity analysis in Section 4.5 that model output is in fact sensitive to all parameters); or (3) the empirical data are insufficient to significantly constrain the parameter. As a straightforward example of (3), Figures 4.9(p) and (q) indicate that parameters controlling market maker and fundamental buyer/seller toxic order flow sensitivity are rarely identified. This is consistent with expectations because such parameters only become relevant under distressed market conditions, such as during a Flash Crash. Since we are calibrating our model during non-crisis market conditions, it is unlikely that large one-minute price movements of several percentage points are commonly encountered in the data. Therefore, the toxic order flow parameters are rarely relevant to the output of a simulation, though it would be surprising if calibrations resulted in very low toxicity thresholds, as movements of one or two *ticks* (minimum price increments

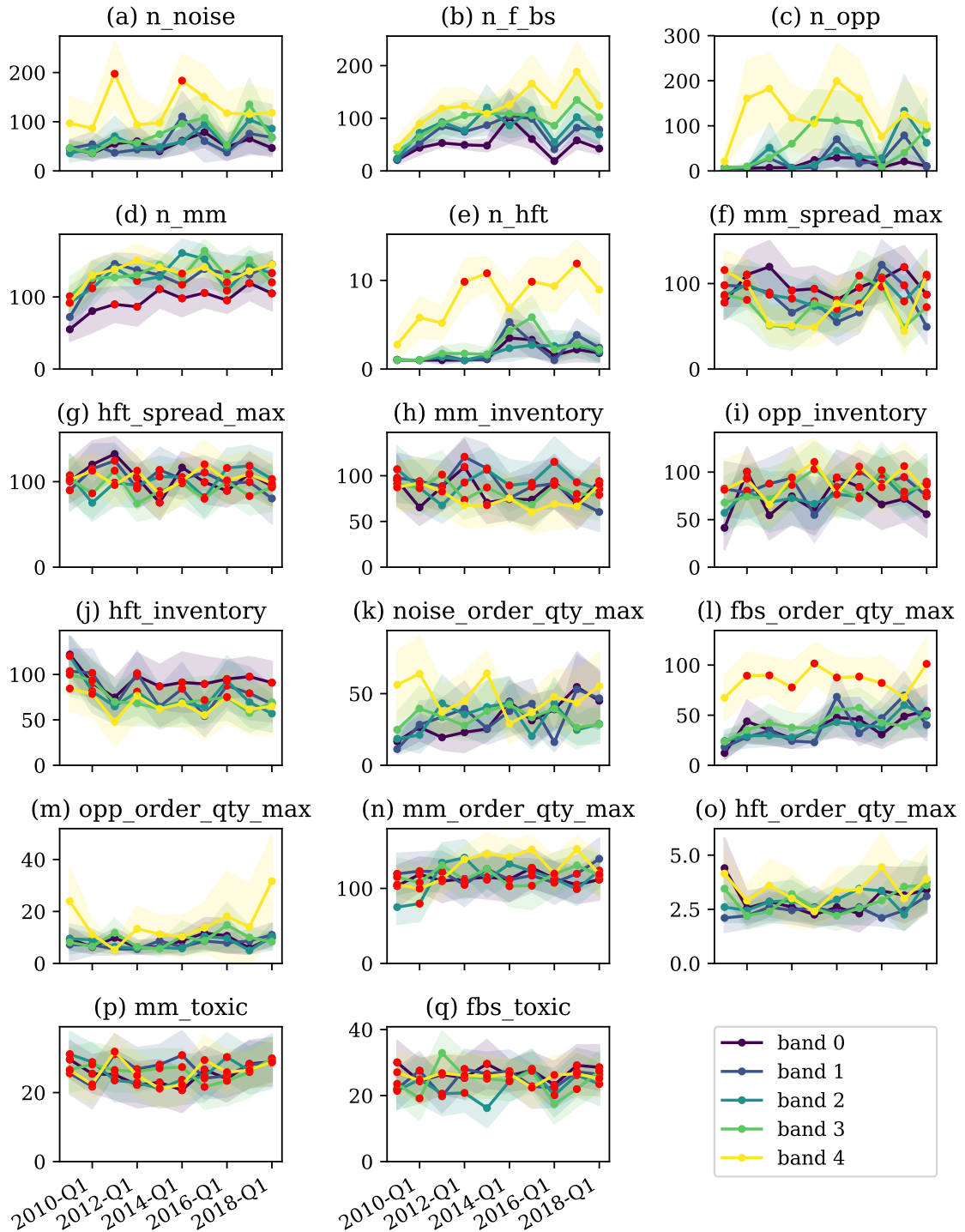


Figure 4.9: Mean and 95% confidence interval of parameter values from 20 independent calibration trials using the genetic algorithm method. Please see Table 4.3 for parameter descriptions. Results are presented for each market cap band from 2009-Q1 to 2018-Q1. Band-date pairs that fail a Kolmogorov-Smirnov test for difference from a uniform distribution at $p = 0.05$ are indicated by a red marker.

on the LOB) are common. Further replications would be required to fully determine whether this is indeed the case.

The K-S test results presented in Figure 4.9 suggest that parameters controlling the number of agents, and the size of orders placed by those agents, tend to be well-identified, i.e., they are meaningfully constrained by the calibration process. An interesting exception are the parameters controlling the number and order size of market maker agents in the lowest market cap band (Figures 4.9(d) and (n)). It appears that market maker activity is less important for low market cap assets than for assets in other market cap bands. A possible explanation is that market makers seek to limit their risk due to adverse selection and therefore avoid volatile low cap assets (Nagel, 2012; Raman et al., 2014). This interpretation is consistent with apparently wider market maker spreads for lower band assets revealed by Figure 4.9 (f), though the results have borderline K-S significance for lower bands.

Figure 4.9 reveals that the number of each class of agent, and the size of the orders placed by the agent class, is generally an increasing function of asset market cap. This trend is clear in Figures 4.9 (a)–(e), the parameters controlling the number of each class of agent in the simulation, and also in Figures 4.9 (k)–(o), the parameters controlling the order sizes placed by each agent class. The highest market cap band (band 4) features the largest number of all agent types, with the exception of the number of market makers (n_{mm} , Figure 4.9(d)) where it is joint highest with bands 1–3. The lowest market cap band (band 0) conversely features the lowest number of fundamental buyer/seller (n_{fbs}) and market maker (n_{mm}) agents (Figures 4.9(b) and 4.9(d)), and the joint lowest number of each other agent class. Correspondingly, Figures 4.9(k)–(o) show that agent order sizes are also well-ordered with respect to market cap, with larger order sizes present for higher market cap assets, though further replications would be beneficial to improve the statistical significance of this trend for the lower market cap bands.

These observations are consistent with expectations on the basis of a liquidity argument. Asset liquidity (i.e., the total traded value of an asset over some horizon, 30 minutes in our case) is known to be highly correlated with market capitalisation (Fama and French, 1993). Since liquidity is directly incorporated within our loss function $\hat{\mathcal{L}}(\theta)$, it is to be expected that calibration against data from a higher market cap band will result in ABM parameters that give rise to higher liquidity during simulations. The sensitivity analysis in Section 4.5 revealed that the parameter sensitivity of the ABM’s simulated liquidity is dominated by the numbers of opportunist and HFT agents, and the size of the orders placed by those agents. These agent classes often

trade aggressively (i.e., they remove liquidity from the order book by placing orders priced to match against offsetting passive resting orders), and, unlike noise traders, trade frequently. Since an aggressive participant is required for any trade to occur, it is to be expected that increasing the number of these classes of agent and their order sizes would increase simulated liquidity. However, Figures 4.9 (a), (b) and (d) also show that higher market cap bands are associated with higher numbers of the remaining agent classes (`n_noise`, `n_f_bs`, `n_mm`), with Figures 4.9 (k), (l) and (n) providing similar evidence that higher market cap bands are associated with larger order sizes for these classes. This result is consistent with the liquidity argument outlined above, however it is surprising that these trends are not also revealed by the sensitivity analysis (Figure 4.5). The sensitivity analysis instead reveals that moments related to price movement (e.g., the standard deviation of price returns) have significant first-order sensitivity to `n_f_bs`, `n_mm` and `fbs_order_qty_max`. A likely explanation is that higher numbers of these agent classes (both of which trade in such a way as to dampen price movements) and their use of larger sized orders are required to offset the extra volatility induced by increasing the number of opportunist and HFT agents for higher market cap bands.

Figure 4.9 also reveals some interesting temporal patterns. In general, the number of agents in each agent class and the size of orders placed by those classes has shown a persistent increase between 2009–2018. These trends are consistent with the increasing equity market liquidity observed in the years since the financial crisis, coinciding with the rapid and persistent growth in equity indices such as the S&P 500 and FTSE 100. This trend, although apparent to each agent class to some extent, is weakest for the market maker agent class. The number of, and order sizes placed by market makers in higher market cap bands seems to have remained stable since 2012 (Figures 4.9(d) and (n)). HFT activity, by contrast, seems to have increased fairly consistently through time across the market cap spectrum (Figures 4.9(e) and (o)). Without obtaining access to market data labelled with participant identifiers (such as the data used by Kirilenko et al. (2017)), it is not possible to independently verify whether the distribution of the different types of participants in equity markets has indeed changed in the way suggested by our results. However, the fact that we are able to successfully validate our model for all market cap bands from 2009–2018 inclusive suggests that the behaviour of market participants has not changed so radically that it is necessary to invoke entirely new classes of agent to explain the observed market dynamics.

Parameters controlling agent spreads and inventories are less well-identified by the data than parameters controlling the number of agents and their order sizes. However, Figure 4.9 reveals some persistent trends that are worthy of further discussion. As discussed above, Figure 4.9 (f) indicates that market maker spreads tend to be lower for higher market cap bands, corresponding to the lower volatility of higher market cap assets. In addition, Figure 4.9 (f) suggests that market maker spreads were lower during the years 2012–2016 across the market cap spectrum, corresponding to historically low market volatility. The dynamics of the ABM parameters with respect to market volatility thus appear to align well with real-world expectations across a wide range of dates and assets. HFT agent inventory (Figure 4.9 (j)) shows a marked decrease following the 2010 Flash Crash. This could reflect a sudden reduction in risk appetite by those market participants following losses incurred during the unprecedented market conditions of 6th May 2010. However, HFT inventory levels recovered to pre Flash Crash levels in 2012 and have remained fairly stable since then, suggesting any extra risk aversion was short lived. Figure 4.9 (j) further suggests that HFT inventories are persistently larger for lower market cap assets than for higher market cap assets. This result stands in contrast to the number of HFT agents and the order size placed by those agents, both of which are increasing functions of asset market cap. A smaller HFT inventory implies that the HFT agents will be more likely to place aggressive orders, and remove liquidity from the order book. Note that they undertake aggressive trading 50% of the time whether their inventory is saturated or not. Smaller HFT inventory for higher market cap bands thus indicates that HFT agents will play a more aggressive role for higher market cap bands, and thus will act to increase simulated market liquidity (defined as the sum of the value of executed trades), in the same way as increasing the number of HFT agents, as described above. Meaningful discussion of HFT spread (Figure 4.9 (g)), and market maker and opportunist inventory limits (Figures 4.9 (h) and (i)) are hampered by low statistical significance. A possible trend towards larger opportunist inventories for higher market cap bands is suggested by Figure 4.9 (i). When opportunist inventories become saturated, they will only place orders that act to reduce their absolute position. Since opportunists follow a herd mentality with a slow time variation, opportunist inventory saturation will likely affect many individual agents simultaneously, and will thus significantly restrict the activity of this class of agents for an extended period, until the stochastic signal (based on a random walk process) changes sign, encouraging opportunists to switch from buying to selling behaviour, or vice versa. It could be the case that fully active (i.e., non-saturated) opportunists are necessary for the

simulation to generate output moments that closely match empirical observations for high market cap bands, giving rise to looser inventory limits for opportunists in these bands. The sensitivity analysis presented in Figure 4.5 indicates that many output moments have a significant total-order sensitivity to opportunist inventory, and indeed on the parameters controlling the inventory of other agent types. Even where there is not significant total-order sensitivity, for example with the log liquidity moment, inventory control parameters are shown to exhibit significant interactions with other parameters. It would thus be interesting to perform further replications of the calibration procedure in order to improve statistical significance for these parameters, as they are expected to be relevant to model output, and there is tentative evidence of temporal and market cap dependence.

4.10 Flash Crashes in the Calibrated Model

As discussed in Section 4.7, the model was calibrated using data from the first quarter of each year in order to avoid including the Flash Crash of 6th May 2010 in the training set. Furthermore, the ABM is not shocked during the calibration process via the introduction of distressed selling agents (or otherwise). As discussed in Section 3.8, the occurrence of flash crashes in the microstructure simulation is strongly dependent on the properties of distressed selling agents. Since these agents are not present during calibration, and since the empirical data used for calibration do not incorporate the Flash Crash, our null hypothesis is consequently that the calibrated model will not exhibit flash crashes.

A rejection of the null hypothesis would be a powerful indicator of the model's generality, since we will have shown that the model produces quantitatively realistic behaviour during non-crisis conditions, as well as qualitatively realistic behaviour during crises. Our microstructure dataset is insufficient for the quantitative validation of crisis behaviour (i.e., during the 6th May 2010 Flash Crash), because (1) we do not have access to data on the level of distressed selling that took place in equity markets on 6th May 2010 (and so are unable to calibrate the size of shock to apply), and (2) there is only a single crisis data point, that of the 6th May 2010 Flash Crash, exposing any calibration procedure to the risk of over-fitting.

In order to test the null hypothesis, we extend the experimental protocol described in Section 3.5, and introduce a distressed seller agent to each simulated realisation of calibrated model parameters for all quarters $Q \in (2009\text{-}Q1, \dots, 2018\text{-}Q1)$ and for all market cap bands $b \in (0, \dots, 4)$. Let us denote the calibrated parameters for a

band and quarter as $\hat{\theta}_{bQ}$. We select $\hat{\theta}_{bQ}$ that maximises the joint moment coverage ratio (Equation 4.16) from the set of 20 repeats of the genetic algorithm calibration procedure for each (band, date) pair, discussed in Section 4.7.

Following the approach of Section 3.8, we investigate the probability of flash crash occurrence as a function 15 settings of the distressed agent order size volume limit $\eta \in [0.005, 0.25]$ and distressed agent activity timescale $\delta \in \{1, 2, 5, 10, 20, 50\}$ seconds. For each combination of parameters (b, Q, η, δ) , the experimental protocol is:

1. *Setup (meta)*: Exogenous selection of ranges of (b, Q, η, δ) parameters, and the number of Monte Carlo replications of each parameter combination. MC instances dispatched to high-performance computing cluster.
2. *Setup (per-trial)*: Construction of agents with stochastic parameters determined via pseudorandom number generation, and using calibrated parameters $\hat{\theta}_{bQ}$, for a single asset.
3. *Intraday model*: Simulate $T = 40,800$ periods of market activity, each step lasting $\delta t = 50$ milliseconds (total of 30 + 4 minutes to allow initial transient decay).
4. *Single period*: Agents are selected for order placement according to a Poisson process. LOBs are updated in industry standard price-time priority.
5. *Shock*: After $T^* = \lfloor 0.2T \rfloor$ periods, to allow for the decay of initial transient effects ($T^* > 4$ minutes), a distressed selling agent is introduced for the asset. We then observe subsequent system dynamics, and measure $\mathbb{I}_{FC}$ (Equation 3.11).
6. *Distributional outputs*: Repeated stochastic MC trials facilitate the collection of distributional results. In this case, the probability of observing a flash crash $P_{FC}(b, Q, \eta, \delta)$.

The distressed selling agent places market orders. Other order types are also possible, and produce qualitatively similar results, as discussed in Section 3.8²¹.

Figure 4.10 presents the probability of flash crash occurrence $P_{FC}(b, Q, \eta, \delta)$ in the calibrated ABM as a function of distressed selling agent properties for each market cap band from 2009-Q1 to 2018-Q1. Each panel presents the longitudinal variation in P_{FC} between 2009–2018, with a line per band (band colour assignments are given in the figure legend). The distressed seller order placement timescale δ is shared

²¹The total number of shares to be sold by the distressed seller is held constant at 2×10^7 , with a total simulated value of $\$2 \times 10^9$, which is a similar scale to the distressed selling activity during the 6th May 2010 Flash Crash reported to total approximately $\$4 \times 10^9$ (Menkveld and Yueshen, 2019).

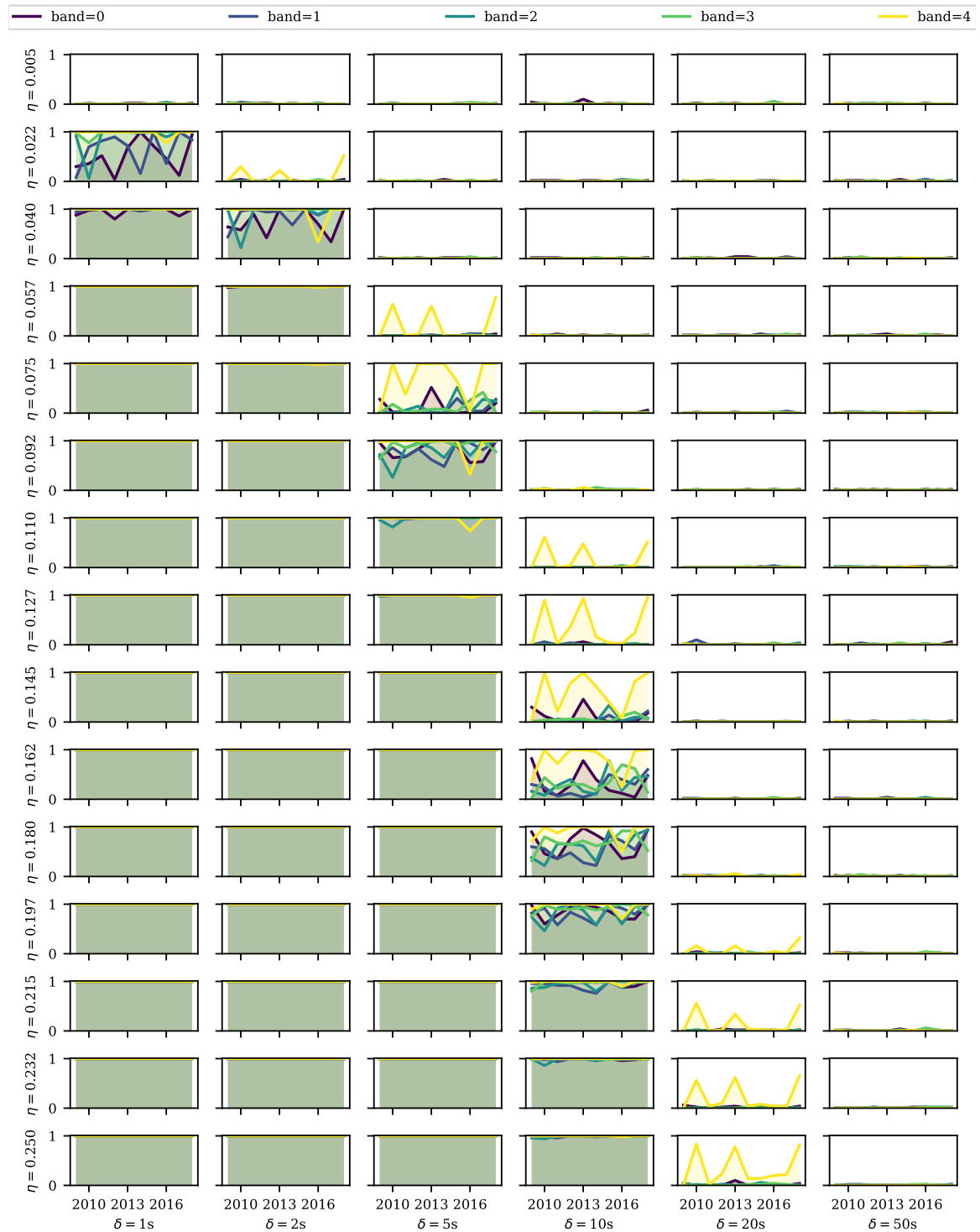


Figure 4.10: Probability of flash crash occurrence $P_{FC}(b, \mathcal{Q}, \eta, \delta)$ in the calibrated ABM as a function of distressed selling agent properties for each market cap band from 2009-Q1 to 2018-Q1. η refers to the order size limit as a fraction of rolling one-minute market volume. δ refers to the mean interval between sell order placements, in seconds. $P_{FC}(b, \mathcal{Q}, \eta, \delta)$ is sampled over 100 Monte Carlo replications. Confidence intervals are suppressed for clarity.

by panels in a single column, and the distressed seller volume fraction η is shared by panels along a single row. For ease of comparison with the results presented in Section 3.8, an alternative way of organising these data is provided in Appendix D, where panels along a row share a common date, with a column of panels for each market cap band.

It is apparent from Figure 4.10 that the model readily exhibits flash crashes. Indeed, in the lower-left panels (i.e., combinations of small timescale δ and high volume limit η), flash crashes are observed with 100% probability for *all* market cap bands in *all* quarters. We thus confidently reject the null hypothesis.

For values of the distressed seller timescale $\delta < 20$ seconds, there exists a critical value of volume limit η above which the probability of flash crash occurrence approaches 100%. Higher values of δ correspond to longer intervals between distressed order placements, and so it is not surprising that the placement of larger orders (i.e., larger η) is a necessary condition for flash crash occurrence as δ increases. At $\delta = 20$ seconds, flash crashes are only observed for the largest order size limits, and even then only for the highest market cap band. We discuss the dependence on market cap band below. Flash crashes are rarely observed for $\delta > 20$ seconds. This dependence on η and δ is qualitatively consistent with results presented in Section 3.8.

Panels along the main diagonal of Figure 4.10 present results that exist along the boundary between stable (upper right) and unstable (lower left) behaviour. Panels in this intermediate regime reveal that flash crashes are more likely to be observed for higher market cap bands. This result is in agreement with empirical observations by Fry and Serbera (2017), but stands in contrast to macro-level models such as Cont and Schaanning (2019) who assume that price impact is inversely related to daily asset liquidity (and hence is inversely related to asset market cap). We attribute this result to the behaviour of liquidity-taking trader agents such as HFTs, and the use of backwards looking volume limits for the distressed seller's orders: Higher market cap assets exhibit consequently higher market volume (Section 4.9), and since the size of orders placed by distressed sellers is proportional to the market volume over the previous minute (with constant of proportionality η), distressed sellers in higher market cap assets can place larger orders than for lower market cap assets, for a given value of η . As also discussed in Section 4.9, higher market cap stocks tend to have proportionately higher numbers of aggressively trading agents such as HFTs. The liquidity-taking behaviour of such agents implies that although trade volumes are higher for high market cap assets, we would not necessarily expect to observe higher volumes of passive, resting orders on the LOB. Indeed, HFT's tendency to inflate

trade volumes without providing genuine market liquidity is a key criticism of the practice (Abrol et al., 2016; Golub et al., 2012; Jain et al., 2016; Raman et al., 2014). The dangerous combination of relatively small resting order volumes on the LOB, with relatively large distressed sell order volumes, means that distressed selling activity can more readily overwhelm the market's ability to offset the distressed volume, leading to increased probability of flash crash occurrence, as observed in Figure 4.10.

The final pattern suggested by Figure 4.10 again relates to the intermediate (η, δ) regime along the main diagonal. In addition to the per-band dependence discussed above, these panels also indicate a degree of longitudinal variation. In particular, the probability of flash crash occurrence for high market cap assets is elevated in 2010, and 2013. Although this observation does not apply to lower bands, it is robust to variations in η and δ . It is perhaps unsurprising that we find the probability of flash crash occurrence to be elevated in 2010, given this is the date of the original Flash Crash of 6th May of that year. We reiterate that our model was not exposed to data from this Flash Crash during training, and so the qualitative similarity to the event is entirely out of sample. Elevated flash crash occurrence probability is exclusively a consequence of the specific ABM parameter vector calibrated for each band, date pair. However, it is not clear why flash crash occurrence also appears to be elevated in 2013 and 2018. The longitudinal variation does not directly correspond to the pattern of market volatility discussed in Section 4.7 — 2013 was a period of low market volatility, whereas volatility returned in 2018. Furthermore, the calibrated parameter values shown in Figure 4.9 do not reveal any obvious features local to 2013 and 2018 for the high band. Our selection of a single parameter vector for each (band, date) pair $\hat{\theta}_{bQ}$ such that jMCR is maximised means that the parameter values could be outliers of the distributions shown in Figure 4.9, but further research would be required to determine the extent to which flash crash occurrence could be attributable to this possibility. Furthermore, we have argued in Section 4.6 that jMCR constitutes a rigorous, state-of-the-art validation criterion, and it does not immediately follow that the parameters selected using this method are inferior, just because they are dissimilar to the mean of parameters giving rise to lower values of jMCR (if indeed such dissimilarity is present). In summary, our results suggest that the probability of flash crash occurrence is elevated for some (band, date) pairs. However, further research would be beneficial to determine the robustness of this observation.

4.11 Conclusions

Our goal in this chapter was to determine whether the model can, under a set of parameters inferred through calibration, give rise to behaviour that reflects that of the real-world with a high degree of statistical confidence (i.e., that the model is valid). We have demonstrated that this is indeed possible, and in so-doing, we have offered a significant contribution to advancing the quantitative rigour of the ABM methodology: We have progressed the validation of our model from Level 1 (*qualitative* agreement with empirical macro-structures) to Level 2 (*quantitative* agreement with empirical macro-structures) in the taxonomy of [Axtell and Epstein \(1994\)](#).

The key results in this chapter include:

1. We have given the first demonstration of the feasibility of performing whole-of-market microstructure calibration, using high-frequency (tick-level) data. This is the first time that a quantitatively validated and calibrated ABM, representative of the full asset market capitalisation spectrum, has been reported in the literature.
2. We have shown that a one-size-fits-all approach to microstructure calibration is insufficient, because asset stylised facts and calibrated parameters exhibit strong variation as a function of market capitalisation. We thus support arguments from [Sirignano and Cont \(2019\)](#) who have performed state-of-the-art work on multi-asset limit order book analysis (in a non-ABM context).
3. We have demonstrated how data from multiple assets can be combined into stylised fact estimates and empirical confidence intervals, thus avoiding the need for bootstrapping methods used in previous studies (e.g., [Franke and Westerhoff, 2012](#)).
4. In order to accommodate larger datasets, more complex models, and hence fewer possible Monte Carlo samples, we extended the joint moment coverage ratio metric from [Franke and Westerhoff \(2012\)](#) by incorporating Wilson-adjusted confidence intervals ([Wilson, 1927](#); [Brown et al., 2001](#)). Our method thus enhances the applicability of the jMCR for principled model validation regardless of model computational cost.
5. We used global sensitivity analysis to demonstrate the relevance of all 17 microstructure ABM parameters, and we advocate for the use of global SA as a key component of any future comprehensive ABM validation experiments. Our approach contributes to the long-standing goal in the ABM discipline of

establishing consistent and sound methodology and terminology (Fagiolo et al., 2019).

6. Although Platt (2019) found Bayesian methods to offer the best ABM calibration performance, our results, based on different data types used (asset trade prices; real data) and different granularity of data (market tick level) indicate that genetic algorithms offer the most effective ABM calibration performance. Further analysis is needed to determine the relative performance of the approaches across the full range of ABMs.
7. The out-of-sample validation of our model — in terms of future date ranges, alternative asset classes, a multi-stage validation procedure, and the model's ability to exhibit flash crashes despite these not forming part of the training set — indicates the generality of the model. This result further supports the relevance of our approach to academic and regulatory ABM practitioners.

Not only have we demonstrated that the model can be successfully calibrated and validated (according to rigorous formal definitions that we have introduced), we have also demonstrated successful calibration and validation at a scope and scale that has not previously been reported in the literature (Section 4.2.3). Our results advance the state-of-the-art for large scale ABM calibration and validation along several key dimensions:

Extensive Range of Dates Previous microstructure ABM calibration approaches that make use of high-frequency (sub-daily) data typically have access to just a few days of market data (e.g., Platt and Gebbie (2018), Kårvik et al. (2018)) with the recent work by Kononovicius and Ruseckas (2019) standing out for its access to two years of market data. In contrast, our study makes use of ten years of high-frequency data, and we demonstrate successful validation for each year of this sample. This result establishes the generality and adaptability of our model. ABM practitioners and policy makers will find our quantitative and qualitative results of interest as they consider appropriate frameworks for modelling systemic risk due to algorithmic trading.

Extensive Range of Assets Previous approaches to microstructure ABM calibration have often been limited to a very small number of assets, most often a single asset. Some daily-timescale models have reported calibration against multiple assets (Recchioni et al., 2015; Lux, 2018), but high-frequency models utilising data from

multiple assets are rare. [Kononovicius and Ruseckas \(2019\)](#) again stands out in this dimension for the authors' access to data for 26 NYSE equities, and [Aloud et al. \(2017\)](#) obtain access to data for 48 currency pairs (e.g., USD/EUR). Our dataset comprises 630–640 assets for the full ten-year date range, allowing us to perform an extensive whole-of-market ABM calibration and validation for the full range of asset market capitalisations, across the full time period of our data sample. Our results indicate for the first time that a one-size-fits-all approach to microstructure ABM calibration is insufficient, as we observed significant differences between parameters calibrated for assets with different market capitalisations. The methodology we have developed in this chapter demonstrates that, and how, time-series from multiple assets can be effectively combined as a function of market capitalisation, in order to avoid the less-satisfactory bootstrapping methods required to estimate empirical moment distributions in previous approaches ([Winker et al., 2007](#); [Franke and Westerhoff, 2012](#); [Fabretti, 2013](#)). Given the lack of previous studies featuring a significant breadth of assets, we hope our results will be relevant to modern market practitioners and policy makers.

High Time Resolution Our access to high-frequency tick-level data represents a significant improvement on most previous approaches. Although financial ABM calibration studies have increasingly reported access to sub-daily data over recent years ([Platt and Gebbie, 2018](#); [Kårvik et al., 2018](#); [Kononovicius and Ruseckas, 2019](#)), it is extremely rare for data with a fidelity of greater than one update per minute to be used. [Aloud et al. \(2017\)](#)'s use of tick-level data is notable, though that dataset was restricted to 48 assets over a period of two years. Our access to tick-level data for a full decade, across the full market capitalisation asset spectrum places our results into an entirely new scale compared with previous studies. We proposed and assessed methodology sufficient to exploit a dataset which exceeds the number of observations in the previous state of the art by four orders of magnitude (Table 4.1). Not only is our methodology of relevance to practitioners and policy makers faced with ever growing high-frequency datasets, we have also used it to comprehensively demonstrate the successful calibration and validation of a hybrid micro-macro ABM of the entire market. The range of dates and assets for which we have demonstrated the model's ability to reproduce high-frequency phenomena that are quantitatively consistent with the real world far exceeds previous approaches, and provides a confirmation of the model's generality.

To further enhance the relevance of our results to financial ABM practitioners, we performed a comparative study of alternative heuristic optimisation methods that may be used to calibrate the ABM. Due to the stochastic and emergent nature of ABMs, computational requirements are often extremely material to ABM practitioners. By comparing a set of adapted optimisation algorithms under a fixed computational budget constraint, our results establish the robust effectiveness, and high relative efficiency, of genetic algorithms for the calibration of market microstructure ABMs. Although recent approaches have promoted Bayesian methods for ABM calibration (Grazzini et al., 2017; Platt, 2019), we find that genetic algorithms offer superior performance for the class of calibration problem represented by our model.

Previous financial ABM research has rarely featured such a comprehensive set of calibration and validation methods as those we have presented here. We advocate the practices of quantitative input validation, output validation, and calibration, and our work helps the financial ABM community advance towards consensus in terminology (Gatti et al., 2018). Furthermore, we have shown that sensitivity analysis can establish with quantitative rigour whether proposed model parameters are redundant.

The calibrated and validated whole-of-market ABM offers policy makers a robust *in silico* experimental environment for the evaluation of regulatory interventions, at a realistic scale and fidelity, that — for the first time — quantitatively connects the complex macroscopic network of funds and assets, to the microscopic price discovery processes that result from the collective activity of electronic trading algorithms.

In Chapter 5 we deploy our calibrated and validated whole-of-market ABM to evaluate the effectiveness of regulatory policies and interventions on systemic risk due to the timely and critical phenomena of flash crash propagation, contagious deleveraging, and fire-sales.

“The proliferation of high-speed algorithmic trading provides another fertile ground for empirical research into circuit breakers ... this prospect too suffers from data availability challenges.”

— Sifat and Mohamad (2019, p.1158)

5

Real-World-Scale Policy Evaluation

IN CHAPTER 4 we proposed, assessed, compared, and deployed a suite of quantitative calibration and validation methods. We showed that the approach that we developed is sufficient to instantiate microstructure ABM components that reproduce a comprehensive set of market dynamical properties (stylised facts) to a level that is not significantly different from reality. In this chapter, we use the ABM parameters determined through calibration to instantiate a version of our ABM that is representative of the real-world fund–asset network. That is, we represent *every* liquid real-world asset, and *every* fund investing in those assets, that features in our empirical dataset (described previously in Section 4.2.1). This chapter represents the first time that a microstructure level simulation has been performed at full real-world scale.

But why would we wish to do such a thing? What insights can a full real-world-scale micro-macro model offer that a small-scale model cannot? The answer to these questions is straightforward. With a quantitatively calibrated real-world-scale ABM to hand, we unlock the full power of ABM as a digital laboratory for meaningful *in silico* regulatory policy analysis.

In systemic risk research we are, by definition, concerned with rarely occurring crisis tail-events. Due to the dearth of empirical examples, and the impossibility of performing controlled experiments with real markets, regulators have expressed a pressing need for models that can offer insights into possible future crises, in order

to mitigate against their occurrence (Haldane, 2016; Haldane and Turrell, 2018). In the context of intraday systemic risk, the relevant policy tools are *price limits* and *circuit breakers* (Sifat and Mohamad, 2019). These instruments are designed to halt market trading activity during periods of excessive volatility, such as during crises. However, the effect of trading halts on systemic risk has, to date, received little attention (Section 2.7). In line with the literature on Flash Crash simulation, previous ABM research into the effectiveness of trading halts has focussed almost exclusively on *single asset* scenarios. Such approaches are therefore unable to bridge the gap to considering the systemic impact of trading halts (Section 2.7). This deficiency is especially acute for market-wide circuit breaker (MWCB) evaluations, as MWCB activations are *only* defined as a function of multiple assets (as represented by a market *index* such as the S&P 500 in the United States)¹. In this chapter, we present the first results that bridge the gap from the micro to the macro in the context of intraday systemic risk and trading halts. Not only do we demonstrate how to represent market-wide regulatory policy interventions in a multi-asset microstructure ABM, we do so using an ABM that is comprehensively calibrated to full real-world scale. Our methodology and results should be of interest to financial regulators and academics seeking to push forward the boundaries of quantitatively realistic large-scale simulation of the financial system, as well as the wider ABM community as an example of a quantitatively calibrated and validated micro-macro model.

This chapter proceeds as follows. In the next section we introduce our methodology for constructing a full real-world-scale ABM of the financial system of the United States, demonstrating how to connect the ABM parameters derived in Chapter 4 with publicly available regulatory datasets of fund asset portfolio ownership. In Section 5.2 we subject the real-world-scale ABM to crisis scenarios, and investigate the extent to which systemic risk in the fifty-fund-fifty-asset model introduced in Chapter 3 is consistent with systemic risk in the full-scale model. We enhance this study by considering the topological properties of the assets receiving initial shocks, and investigate the effect of asset size and connectivity on systemic risk. In Section 5.3 we extend our model with the introduction of regulatory policies designed to halt trading during crises. We incorporate the current US regulatory standard composite Limit-Up-Limit-Down (LULD) and market-wide circuit breaker (MWCB) policy². We believe this is the first time the complete US policy framework has been incorporated into a microstructure ABM. In Section 5.4 we use the full-scale calibrated

¹<https://www.nyse.com/markets/nyse/trading-info>, accessed 2019-10-06.

²<https://www.nyse.com/markets/nyse/trading-info>, accessed 2019-10-06.

ABM to evaluate the effect of current regulatory policies on systemic risk. We further extend this analysis by considering the effect of alternative fund behavioural policies on the activation of trading halts and systemic risk. Our results thus contribute to the ongoing debate regarding the effectiveness of price limits and circuit breakers in financial markets (Subrahmanyam, 2013; Dalko, 2016; Sifat and Mohamad, 2019; Hautsch and Horvath, 2019). Finally, in Section 5.5 we test the effectiveness of an alternative regulatory policy (proposed by BlackRock, 2015), that introduces a complex combination of the existing LULD and MWCBC components of US regulation. This is the first time such a policy has been subjected to quantitative analysis, and our results shed light on the effectiveness of the policy at mitigating systemic risk in comparison with existing regulation.

5.1 Real-World-Scale Agent-Based Methodology

In this section we demonstrate how to combine the ABM parameters determined during calibration (Chapter 4), with regulatory data representing the full real-world-scale fund–asset network. We commence with a statistical analysis of the empirical fund–asset network in Section 5.1.1, and in Section 5.1.2 explain how constraining the parameters of our model using this dataset affects assumptions of agent heterogeneity introduced in Chapter 3. We present our real-world-scale ABM instantiation method in Section 5.1.3, and provide indicative estimates of the computational costs involved in performing simulation experiments with the full-scale model.

5.1.1 Empirical Fund–Asset Network (13-F) Dataset

In Section 4.2.1, we described how we set the initial conditions of the fund–asset network using data from SEC form 13-F regulatory filings. We commenced research using empirical 13-F data on 2018-04-17, and our 13-F sample was downloaded at this date. Due to the cost of downloading and cleansing 13-F data, we use data exclusively from this sample date throughout this chapter. However, as demonstrated in Chapter 4, we successfully calibrated and validated the model for all years 2009–2018 and so we would be free to instantiate the ABM for earlier dates if desired. In Section 4.8 we demonstrated that the ABM exhibits positive out-of-sample performance for up to 9 years into the future (albeit with seasonal effects and an overall (low) diminishing, but still sufficiently representative, performance level, as might be expected), which suggests that we would be justified in instantiating the ABM for future dates, thus

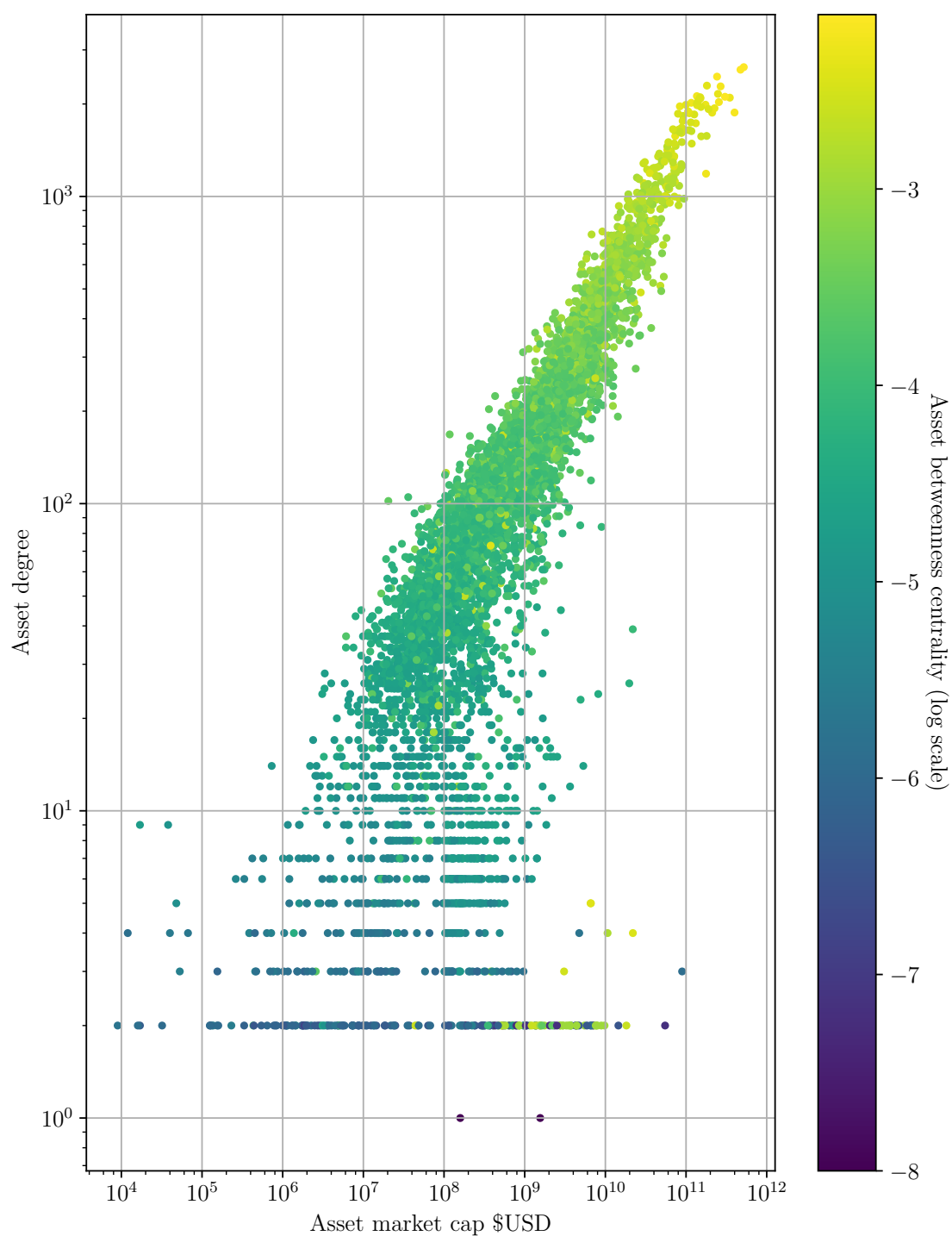


Figure 5.1: Asset weighted degree (equivalent to the market capitalisation) as a function of degree centrality (the number of funds investing in each asset) for real-world SEC 13-F portfolio holdings data from 2018-Q1. Marker colour represents betweenness centrality.

allowing our work to remain of direct quantitative relevance to regulators in 2019 and beyond.

By setting initial ABM conditions using 13-F portfolio data, we are able to represent each of the 4600 filing entities as funds in the ABM, and each of the 20100 reported instruments as assets in the ABM (Section 4.2.1). Although the ABM is capable of simulating 20100 assets simultaneously given reasonable memory and CPU resources, many of the reported assets represent small companies well-below the size necessary to be listed on a major stock exchange such as the NYSE or NASDAQ.³ We therefore filter the assets by removing those with market capitalisations below \$USD 100mn, similar to the minimum NYSE and NASDAQ listing sizes. 6619 assets remain after filtering. The filtering method is described in detail in Section 5.1.3. We access two complementary sources of market cap data⁴, the second of which is used for the \$USD 100mn filtering operation:

1. Summation of a column of the fund–asset network adjacency matrix $\mathbf{v}(\mathbf{A})$ (Equation 4.20).
2. Loading the true market cap directly from a data provider ($\mathbf{v}'$).

Figure 5.1 presents an overview of the properties of the asset nodes in the empirical fund–asset network. We plot the market capitalisation of each asset inferred from summation over columns of the fund–asset network ($\mathbf{v}(\mathbf{A})$, horizontal axis) against its degree (vertical axis). The degree corresponds to the number of funds that have an investment in that asset. We present the results on a logarithmic scale for clarity. It is clear that there is a strong positive correlation between an asset’s size (market cap) and its connectedness (degree). Larger assets attract more investors than smaller assets. Further measures of the importance of an asset in the fund–asset network are given by network *centrality* metrics (Newman, 2003). The colour dimension in Figure 5.1 represents the *betweenness centrality* of the asset. For a graph $\mathcal{G} = (N, E)$ with nodes N and edges $E \subseteq (V \times V)$, the betweenness centrality of node n_i is the fraction of shortest paths between all pairs of nodes $(n_1, n_2) = (N \times N)$ that pass through n_i (Costa et al., 2007). In Figure 5.1, paths are weighted by the size of investments associated with the edges comprising the path. In the context of the present chapter, we might expect a high betweenness centrality to imply that an asset

³NYSE listing requirements: https://www.nyse.com/publicdocs/nyse/listing/NYSE_Initial_Listing_Standards_Summary.pdf, accessed 2019-09-09. NASDAQ listing requirements <https://listingcenter.nasdaq.com/assets/initialguide.pdf>, accessed 2019-09-09.

⁴We acknowledge Wharton Research Data Services for providing market capitalisation data “CRSP Monthly Stock” wrds.wharton.upenn.edu, accessed 2019-01-31

plays an important role in propagating systemic risk between funds. However, since Figure 5.1 shows that betweenness centrality is very highly correlated with market cap and degree, we proceed with consideration of these metrics for the remainder of this chapter.

A further comment is in order regarding the market cap values plotted in Figure 5.1 (estimated by column summation over $\mathbf{A}$), compared with the true market cap data used for the \$USD 100mn filtering. As discussed in Section 4.2.1, we follow the approach of Koijen and Yogo (2019) who attribute the difference between the estimated and true market cap to the “household” sector (Koijen and Yogo, 2019, p.16) who do not engage in sophisticated leveraged portfolio management or algorithmic trading, and so are not relevant to the present investigation. However, it is important that we use the true asset market cap when selecting agent parameters obtained through calibration (Chapter 4). As a reminder, the microstructure ABM was calibrated separately for each of five market cap bands, with band designations based on the true asset market cap (Section 4.2.3). Thus, for an asset to be assigned the correct ABM parameters during full-scale model instantiation, we must utilise the true asset market cap. Full-scale model instantiation is described in detail in Section 5.1.3.

Finally, we note that Figure 5.1 is somewhat noisy, for example featuring a small number of assets that are both fairly large (market cap in excess of \$USD 1bn) and poorly connected (degree less than 5). If snapshots of complex networks are treated as noisy observations of the true underlying network, it is possible to use reconstruction methods such as *maximum entropy reconstruction* to improve the accuracy of the inferred network (Gualdi et al., 2016). Due to the limitations of our data access, and the scope of the present chapter, we proceed with the single 13-F snapshot, and do not attempt to undertake complex reconstruction exercises. Our methodology is general and could be similarly applied on a different, reconstructed dataset.

5.1.2 Differences Between Reduced- and Real-World-Scale Models

The full real-world-scale ABM necessarily features some important differences from the synthetic fifty-fund-fifty-asset networks we generated in Chapter 3. These are summarised in Table 5.1 and discussed in detail below. The main changes relate to the size of the network. Although our work in Chapter 3 significantly extended the state of the art in terms of the size of microstructure ABM simulations from one or two assets to fifty assets, in order to bridge the gap to full real-world scale

in the present chapter we increase the size of the simulation by two further orders of magnitude to over six thousand assets. In addition to this significant change, we relax several assumptions regarding the homogeneity of fund and asset properties such as crowding, diversification, capital and asset liquidity/microstructure. These changes serve to substantially increase the realism of the hybrid micro-macro model. In Section 5.2 we investigate the extent to which the findings from our earlier chapters hold in the full-scale regime.

Feature	Fifty-fund-fifty-asset scale	Full-scale model
Number of funds	50	4593
Number of assets	50	6619
Number of edges	$(50, \dots, 2500)$	1.07×10^6
Fund capital	homogeneous	heterogeneous
Fund diversification	homogeneous	heterogeneous
Portfolio crowding	homogeneous	heterogeneous
Asset microstructure	homogeneous	heterogeneous

Table 5.1: Comparison of features between 50-fund-50-asset models investigated in previous chapters, and the full-scale model.

Fund Diversification and Crowding

In Chapter 3 we presented the network of overlapping asset portfolios owned by funds as a bipartite network with fund diversification and portfolio crowding controlled by experimental parameters ρ and β , respectively (Section 3.3.3). We followed methodology established by Cont and Wagalath (2013) and Caccioli et al. (2014) and assumed that all funds have equal degree, that is, they each invest in the same number of assets, $k_{fund} = \max(\lfloor \rho n_a \rfloor, 1)$, where ρ was held constant across funds in a single experimental scenario, and was varied systematically between scenarios. Homogeneous crowding (our parameter β) was implicit in Cont and Wagalath (2013) and Caccioli et al. (2014). In Chapter 3 we also enforced homogeneous crowding across funds (each fund having the same preference as the others to invest in assets that already have the largest investment), and varied β systematically between scenarios.

By utilising the empirical 13-F dataset of fund portfolios to define the fund–asset network in our model, we remove the degrees of freedom represented by ρ and β . In the empirical network, funds invest in heterogeneous numbers of assets, and the level of overlap between a fund’s portfolio and that of all other funds will be idiosyncratic to the fund. Whilst it is not possible to directly infer β for each fund from data

(because the parameter is part of a generative model that is sensitive to the order in which funds are considered), it is possible to measure ρ . We find that the expected value across funds is $\langle \rho \rangle = 3.52\%$. This relatively low value is consistent with the power law distribution of asset degrees presented in Figure 5.1 where a small number of assets have very high degree and a long tail of assets have low degree, implying funds tend to restrict their investments to a crowded subset of assets. This result is also consistent with previous empirical analysis of 13-F data by Guo et al. (2016), who find $\langle \rho \rangle = 3.29\%$ for data from 2003–2012.

Fund Leverage and Capital

In Chapter 3 we followed the methodology of Caccioli et al. (2014) and constrained funds to be homogeneous in both leverage, and capital. Each parameter was systematically varied between experimental scenarios. The product of capital and leverage in our model defines the total value of a fund’s portfolio, known as the gross market value (GMV). Thus funds had homogeneous GMV in our previous results. However, as explained in Section 4.2.1, it is possible to estimate the (long-only) GMV of a fund’s portfolio by taking the sum of individual asset positions from the 13-F dataset. Thus, in this chapter, funds have heterogeneous GMV, with initial simulation conditions inferred directly from 13-F data.

As we explain in detail in the following section, we do not have sufficient data to decompose the GMV into the leverage and capital components. Use of fund GMV inferred from 13-F data therefore allows us to remove a single degree of freedom from the model, but we are left with a choice as to which parameter remains free to vary — leverage, or capital. In the results that follow we elect to specify leverage (λ^0) as an experimental parameter, and hold leverage constant across funds, in a given scenario. We therefore allow capital (C^0) to vary heterogeneously across funds, such that the product $C^0(1 + \lambda^0)$ is consistent with the GMV inferred from 13-F data for each fund (Equation 3.3). Our approach is thus similar to Banwo et al. (2017) who model a bipartite network of banks investing in assets, where bank sizes are given by a power-law distribution (though their results are not based on empirical data).

Asset Microstructure

In Chapter 4 we proposed and demonstrated a methodology to quantitatively calibrate and validate the micro-level component of our model, for assets across the full market-cap spectrum. Having identified appropriate ABM parameters for all asset market-cap

bands on the date of our 13-F sample⁵, we are able to instantiate heterogeneous asset microstructure model components by looking up the market cap of each asset in the 13-F dataset. In Chapter 3 our results assumed homogeneous asset microstructure, and proposing a valid methodology that enables the relaxation this assumption is a major contribution of this thesis. We are not aware of previous works that incorporate heterogeneous asset microstructure models.

5.1.3 ABM Instantiation

Figure 5.2 provides a schematic overview of the method by which we instantiate the full real-world-scale ABM. The first step is to load the raw 13-F data from the SEC’s EDGAR website⁶ (see Section 4.2.1). These data are then cleansed to remove a suite of filing errors (for example, funds reporting in \$USD whilst claiming to be reporting in \$USD 000s). The resulting set of fund portfolios are compiled into a single fund–asset network adjacency matrix representation $\mathbf{A}'$. We then apply Equation 4.20 in order to infer asset market caps $\mathbf{v}(\mathbf{A}')$. As discussed in Section 5.1, we also load the true asset market cap, $\mathbf{v}'$, from an external data provider. It is often the case that we do not have market cap data available for all assets listed in the portfolios of filing funds, and hence $\mathbf{v}'$ will often be incomplete. In such cases, we use the estimated market cap $\mathbf{v}$ as a proxy for the true value, and produce an error-corrected true market cap vector $\mathbf{v}''$.

We then apply two stages of filtering the the fund–asset network. In the first phase, assets with $\mathbf{v}'' < \$\text{USD } 100\text{mn}$ are removed from the network. As discussed above, this restricts the network to assets that are likely to be listed on major stock exchanges, and hence of interest to sophisticated institutions and algorithmic traders. However, this operation can leave some funds disconnected from the network, for example if all of the assets held by a fund are removed from simulation. We therefore apply a second phase of filtering wherein we extract the largest *connected component* (Newman, 2003) from the network. In the case of the 13-F dataset the largest connected component ($\mathbf{A}$) is a *giant* connected component (GCC), and there is a high probability that a given fund or asset is connected to it. We only remove 0.06% of nodes by extracting the GCC, but the resulting analysis is simplified. For example, we do not need to worry about applying shocks to assets that are not connected to any funds, and so the

⁵We calibrated the ABM using data up to 2018-03-31, which corresponds to the 2018-Q1 13-F regulatory filing window. By extracting 13-F data on 2018-04-17 we allow for the inclusion of late filing data.

⁶<https://www.sec.gov/edgar/searchedgar/webusers.html>, accessed 2018-04-19

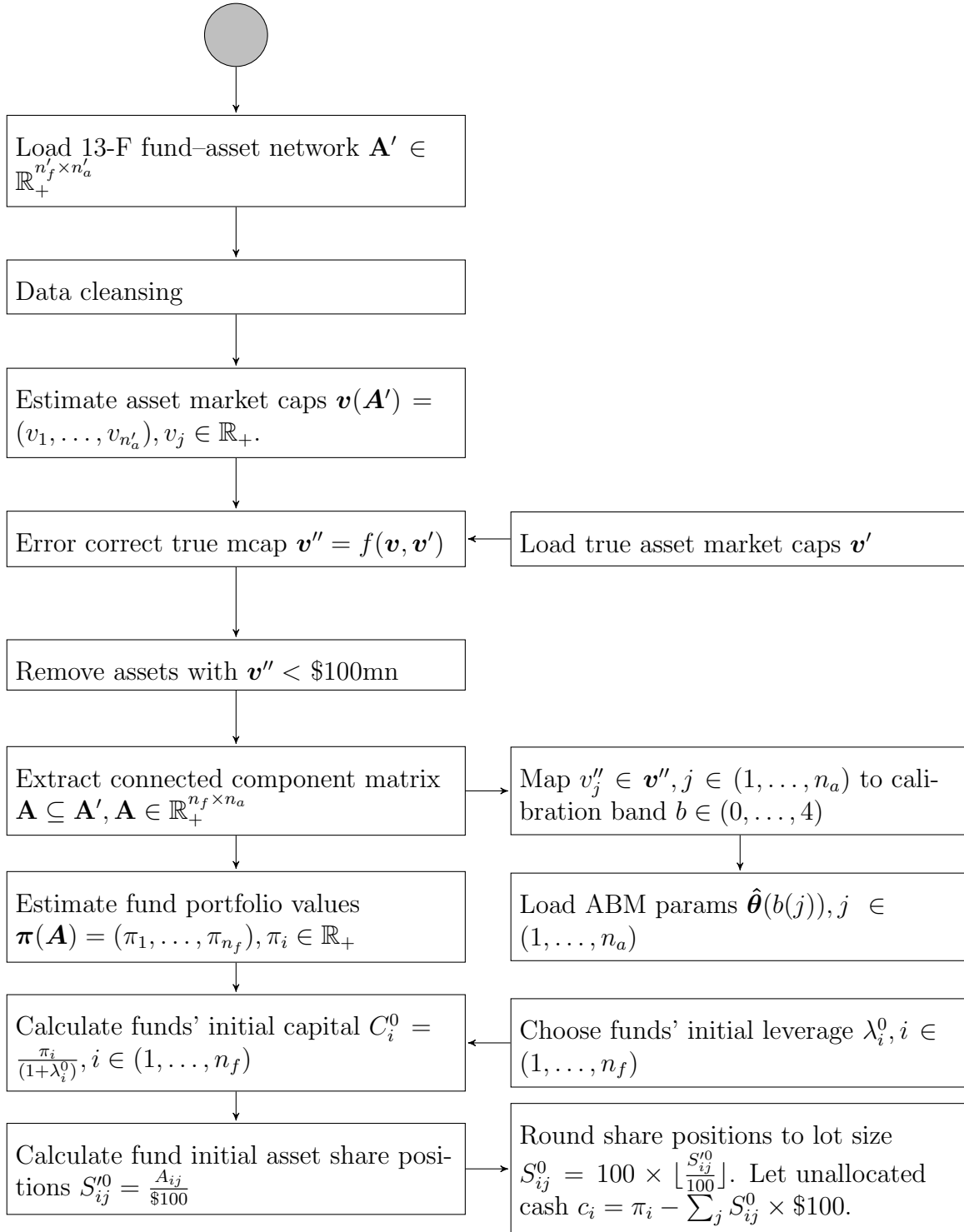


Figure 5.2: The process by which we instantiate the full-scale model using calibrated agent parameters.

checking applied in Section 3.10.1 is no-longer necessary because the set of connected assets $\mathcal{A}$ is isomorphic to the set of all assets ($\mathcal{A} = (1, \dots, n_a)$).

After the filtering phases have been performed, the instantiation process bifurcates with separate branches for the instantiation of the microstructure (asset) and macro-level (bank–fund) simulation components (Figure 5.2).

The key step in the microstructure instantiation is to use the error-corrected market cap for each asset, v'' , to look up the ABM parameters inferred during the calibration process. Let $\hat{\theta}(b)$ be a function that returns the calibrated ABM parameters $\hat{\theta}$ associated with market cap band $b \in (0, \dots, 4)$, and furthermore with a slight abuse of notation let $b(j)$ be a function that returns the error-corrected market cap band for asset $j \in (1, \dots, n_a)$. We are thus able to extract the appropriate micro-level ABM parameters for each asset $j \in (1, \dots, n_a)$.

In addition to their portfolio allocations, further properties of the (macro-level) fund agents in the model are their initial capital C^0 , their leverage λ^0 , and their margin hysteresis tolerance τ (Section 3.3.2). It is also necessary to specify the distressed selling behaviour of funds (we explore the impact of fund behaviour in detail in Section 5.4). We do not have access to data on the distressed selling behaviour or margin tolerance of funds, and so we have to select reasonable values and perform sensitivity analyses to infer the effect of such parameters. Based on the sensitivity analyses in Section 3.9.2 we select a conservative value of $\tau_c = 1.15$, meaning funds allow their leverage to exceed their limit by 15% before taking action to reduce risk. The remaining parameters C_i^0 and λ_i^0 for fund i are constrained such that the product $C_i^0(1 + \lambda^0) = \pi_i$, where π_i is fund i 's total portfolio value (the sum of asset position sizes measured in \$USD). We were unable to source data on leverage usage by funds, and so we cannot disentangle the product $C^0(1 + \lambda^0)$. In the present chapter we therefore choose to specify reasonable (heterogeneous) values for λ^0 across funds, while allowing capital C^0 to vary heterogeneously across funds such that the constraint $C_i^0(1 + \lambda^0) = \pi_i$ is everywhere satisfied. We investigate the effect of λ^0 in the full-scale simulation in the following section.

The final stage in ABM instantiation is to specify the number of shares held by each fund of each asset in its portfolio. As in Chapter 4, we assume an initial price of \$USD 100 for all assets. This convention aids analysis without affecting simulation dynamics; in the real world assets can undergo *splits* and other corporate actions without warning, yet this has a limited effect on investors — it is the value of investment (product of price and quantity) that is material, and we reflect this real-world feature in our model. We also continue to assume a *round lot size* of 100 shares,

meaning that all trades on the limit order book (and by extension, all orders submitted by agents) are multiples of 100 shares. To account for any differences between the total portfolio value π_i inferred from the 13-F data for fund i , and the sum of the rounded asset positions at \$USD 100 per share, we allocate residual amounts (up to \$USD 9999 in value per asset per fund) to a cash pool for each fund c_i , that takes no further part in simulations.

The instantiated full-scale model comprises tens of millions of trader agents, in addition to the thousands of macro-level fund agents and limit order books. The computational burden associated with a simulation at this scale is considerable, and single Monte Carlo replicates take approximately one hour on a 56 core HPC node, using approximately 30GB of RAM. The number of Monte Carlo replicates used in the results reported in this chapter reflects the high time cost of performing full real-world-scale experiments at microstructure (millisecond) fidelity. All simulation runs last for one simulated hour (72000 simulated time-steps) in the results that follow. The number of Monte Carlo replicates varies between the experiments we report here. For some experiments, it was possible to obtain adequate statistical significance with fewer evaluations than for others. We took advantage of this feature where possible, in order to make the most efficient use of our limited computational resources.

5.2 Real-World-Scale Intraday Systemic Risk

In Chapter 3 we generated synthetic fund–asset networks with tunable levels of crowding and diversification, and used these networks to investigate the propagation of systemic risk. In the present chapter we revisit this approach, but in the context of a full real-world-scale ABM, with over a million interconnections between the nodes in the financial network.

In the sections that follow, we subject the fully calibrated ABM to shocks by introducing a distressed seller to a single asset in the same manner that we have established in Chapters 3 and 4. We focus on the probability that a cascade of fund defaults will propagate through the fund–asset network, an eventuality that is quantified by systemic risk metric $P_{contagion}$ (Equation 3.6).

5.2.1 Experimental Scenario Design

As described in Section 5.1, in the full real-world-scale model we relax several assumptions regarding the homogeneity of ABM parameters. In the first experimental

scenario, we seek to understand the extent to which heterogeneity of ABM parameters impacts the dependence of systemic risk on leverage and on the size (market capitalisation) of the asset receiving the initial shock. In Chapter 3 we found that systemic risk is strongly and monotonically dependent on leverage, with increasing leverage leading to increased systemic risk (as measured by $P_{contagion}$). We also found the surprising result that increased portfolio crowding could *reduce* systemic risk in some circumstances (Paulin et al., 2019). We attributed this latter result to the presence of weakly-connected asset nodes in the network. In the present chapter we are at liberty to specify homogeneous fund leverage λ^0 , whilst allowing fund capital C^0 to vary heterogeneously, as discussed in Section 5.1.3. This freedom allows us to directly investigate leverage sensitivity.

However, since network topology is dictated by the empirical 13-F dataset in this chapter, in order to determine the effect of weakly-connected assets, it is necessary to shock assets selected according to their connectivity (rather than constructing networks with the desired connectivity properties). In the second experimental scenario we perform such an analysis, as we seek to disentangle the effect of asset size (market cap) from asset connectivity (measured by degree centrality).

When selecting the asset to shock according to either size or connectivity, we partition the set of all assets into quintiles by market capitalisation, and into an orthogonal quintile partition by asset degree. Let

$$\omega_{\text{mcap}} \in \{[0.2i, 0.2(i+1)), i \in (0, \dots, 4)\} \quad (5.1)$$

be the percentile bins used to partition assets according to market capitalisation, and let

$$\omega_{\text{degree}} \in \{[0.2i, 0.2(i+1)), i \in (0, \dots, 4)\} \quad (5.2)$$

be the percentile bins used to partition assets according to degree. Then each asset is a member of exactly one tuple $(\omega_{\text{mcap}}, \omega_{\text{degree}})$ ⁷. The asset to shock is selected uniformly at random from the desired partition or partition tuple. Note that the specific asset receiving the shock will vary between Monte Carlo replicates regardless of whether the experimental scenario requires an alternative choice of ω_{mcap} or ω_{degree} .

⁷Note that the quintile partitions of ω_{mcap} do not directly correspond to the asset market cap band definitions b . The band boundaries b were defined in Section 4.2.3 as market cap quintiles of the microstructure dataset, whereas ω_{mcap} covers a larger asset universe.

5.2.2 Effect of Leverage and Shocked Asset Size

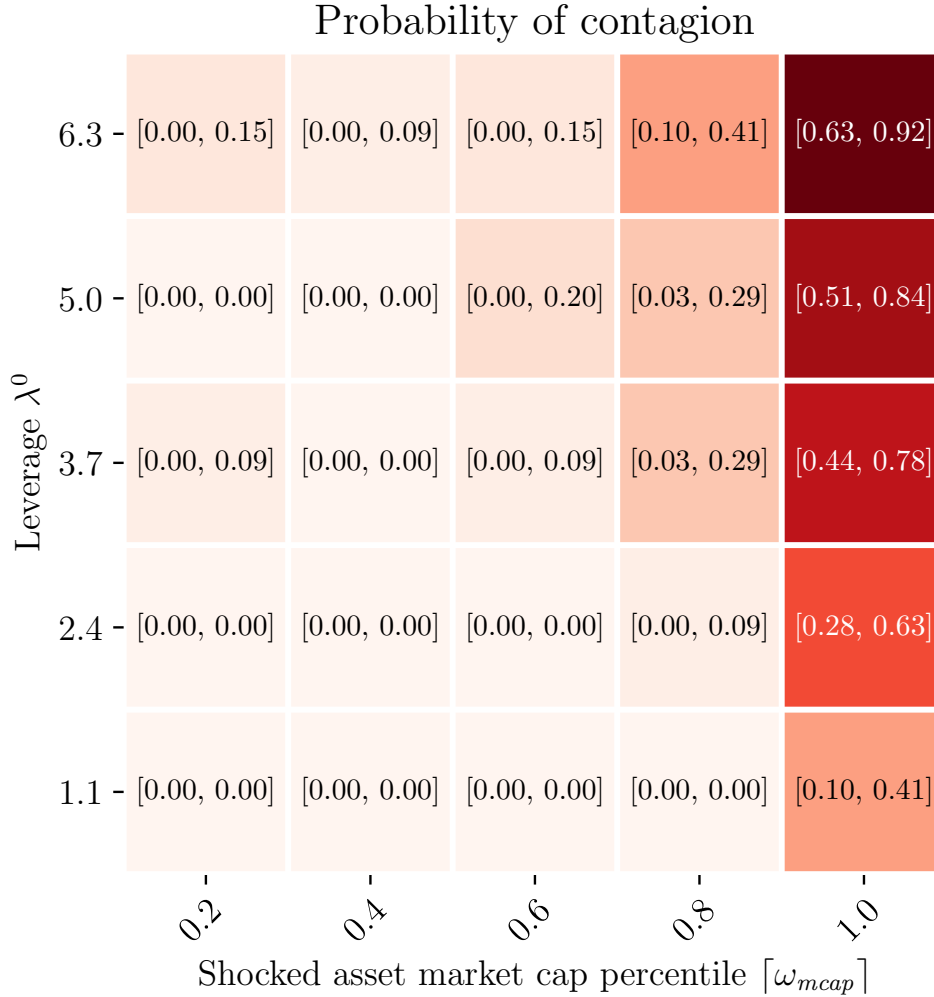


Figure 5.3: Comparing the probability of contagion for leverage $\lambda^0 \in [1.1, 6.3]$ and shocked asset centrality $\omega_{mcap} \in \{[0, 20), [20, 40), [40, 60), [60, 80), [80, 100]\}\%$. Cell shading represents the mean across 31 Monte Carlo replicates. Cells are labelled with 95% confidence intervals.

Figure 5.3 shows a heatmap of the probability of contagion ($P_{contagion}$) as a function of leverage ($\lambda^0 \in \{1.1, 2.4, 3.7, 5.0, 6.3\}$), and the market cap percentile of the shocked asset ω_{mcap} . Horizontal axis labels give the upper bounds of the ω_{mcap} percentile bins. Cells are shaded according to the mean of $P_{contagion}$ from 31 Monte Carlo replications, and are labelled with 95% confidence intervals.

Two trends are apparent from Figure 5.3, and both are consistent with expectations, and our previous results from Chapter 3. Firstly, systemic risk increases

monotonically with leverage. Indeed, it would be extremely surprising if this were not the case. We discussed the reasons for this dependence in Section 3.9. Secondly, we observe that systemic risk increases monotonically with the size of the shocked asset, indicated by ω_{mcap} . This finding is also consistent with our previous work where we observed that systemic collapse is less likely to occur if shocks are delivered to more weakly-connected assets (Section 3.10.4). However, in the previous model asset properties were homogeneous, whereas in the present study assets have different market microstructure properties in accordance with the calibration exercise performed in Chapter 4. During calibration we found that higher market cap assets are more readily prone to experiencing flash crashes under conditions of distressed selling (Section 4.10). This tendency may contribute to enhanced systemic risk at high ω_{mcap} .

The result can be further explained by considering the properties of assets presented in Figure 5.1. The figure demonstrates that asset degree and asset market cap are highly correlated. A shock to a high market cap asset will therefore, on average, impact a higher number of connected funds, increasing the probability that those funds will enter margin call states and be forced to join the distressed fire sale. Furthermore, a given percentage price fall in a high market cap asset will cause a larger total loss to the portfolios of the set of funds compared to an equivalent percentage price fall in a lower market cap asset. This loss must be absorbed by the owning funds, and serves to increase the probability that one or more funds will encounter financial distress.

Although it is not possible to perform a controlled experiment to determine the relative importance of market microstructure and market cap on systemic risk (since market microstructure properties in our model are uniquely determined by market cap and market cap alone), the question remains whether the increase in systemic risk that we observe for high market cap assets is attributable to the larger-size-of-loss, or the larger-number-of-owners effects mentioned above. We address this question in the next section.

We also observe that, for scenarios with a positive probability of contagion ($P_{\text{contagion}}$), the extent of contagion is extremely high ($99(\pm 0)\%$ in all cases).⁸ This result confirms that the full-scale 13-F fund–asset network continues to exhibit *robust-yet-fragile* behaviour (Gai and Kapadia, 2010), whereby the system can be robust to small perturbations, yet fragile in the sense that failure of the financial system is extensive (indeed, almost total) once a default cascade is in effect. This result is consistent

⁸These results are presented in Appendix E (Figure E.1).

with our findings in Chapter 3. Later in the present chapter we extend our analysis to investigate the extent to which robust-yet-fragile behaviour is observed under alternative trading halt policies and fund distressed selling behaviour scenarios.

Finally, we consider whether the full-scale model continues to exhibit flash crashes. We find that the fraction of assets experiencing flash crashes is exactly correlated with probability of contagion, between experimental scenarios. This result is presented in the appendix (Figure E.3).

5.2.3 Effect of the Centrality of the Shocked Asset on Systemic Risk

Figure 5.4 presents the probability of contagion ($P_{contagion}$) as a function of the market cap percentile of the shocked asset ω_{mcap} (rows), and the degree percentile of the shocked asset ω_{degree} (columns). Cells are shaded according to the mean of $P_{contagion}$ from 50 Monte Carlo replications, and are labelled with 95% confidence intervals. As the partitioning into percentiles by market cap and degree is performed independently, some $(\omega_{mcap}, \omega_{degree})$ joint partitions may be empty. Figure 5.4 demonstrates this to be the case for high-degree-low-market-cap cells (see the unshaded cells in the lower-right corner of the Figure 5.4). Specifically, there are no assets that are ranked in the top 40% by degree that are simultaneously ranked in the bottom 20% by market cap. There are also no assets ranked in the bottom 40% by market cap that are simultaneously in the top 20% by degree.

Similarly to the findings presented in the previous section, Figure 5.4 further demonstrates that systemic risk increases monotonically with the market cap of the asset receiving the initial shock. Probability of contagion is higher toward the upper rows of the figure. This result is consistent with expectations established both in the results for the full-scale model presented in this chapter, and also with the results presented for the smaller 50-asset-50-fund models in Chapter 3.

However, Figure 5.4 demonstrates marked non-monotonicity in the dependence of systemic risk on shocked asset degree (ω_{degree}). In almost all cases, systemic risk reaches a maximum when the asset receiving the initial shock is in the 20–40th percentile of degree ranks (the second column, $\lceil \omega_{degree} \rceil = 0.4$, in Figure 5.4). For values of ω_{degree} in the lowest quintile, default cascades are unlikely to occur except for shocks delivered to assets in the largest quintile by market cap. For values of ω_{degree} in the 40–100th percentiles systemic risk is reduced relative to the peak value at $\lceil \omega_{degree} \rceil = 0.4$ in almost all cases. The exception is for the smallest market cap quintile (bottom row of Figure 5.4) where peak systemic risk occurs at $\lceil \omega_{degree} \rceil = 0.6$. However, we

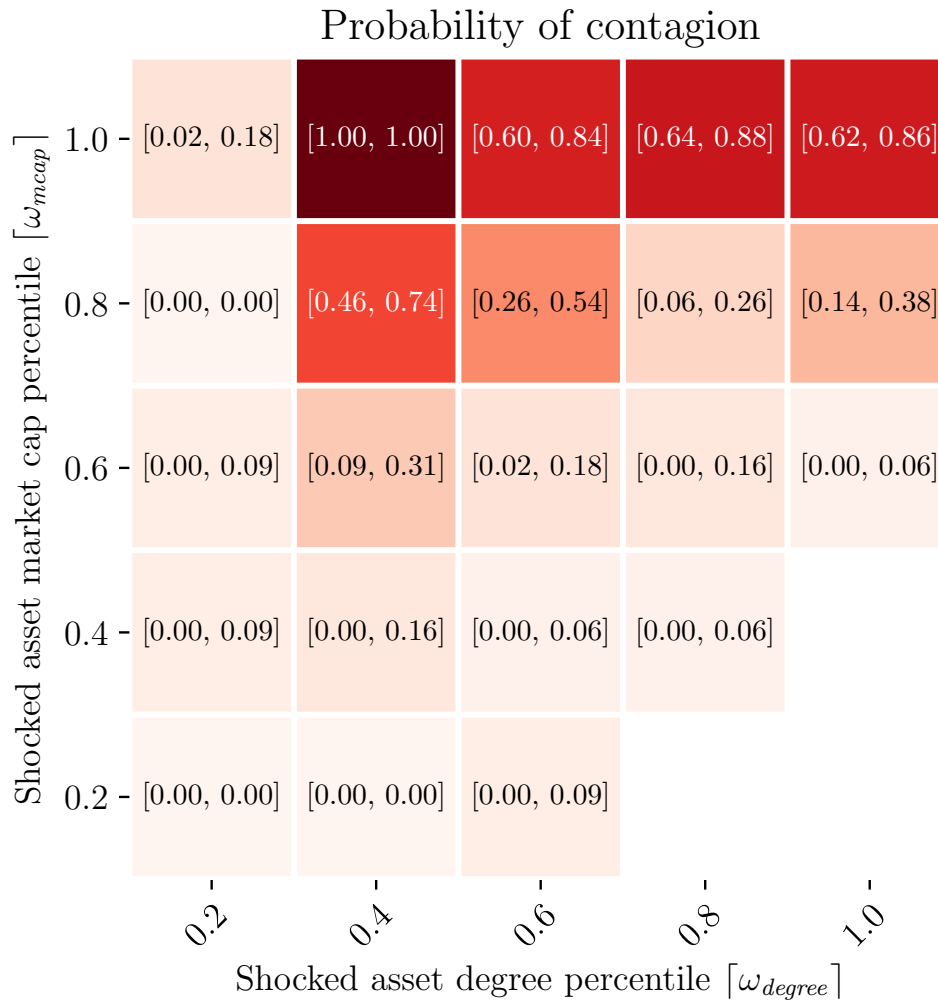


Figure 5.4: Comparing the probability of contagion for shocked asset centrality quintiles of ω_{mcap} and ω_{degree} at leverage $\lambda^0 = 5$. Cell shading represents the mean across 50 Monte Carlo replicates. Cells are labelled with 95% confidence intervals.

caution that these results are based on a small number of Monte Carlo replicates (50) and so some aleatory uncertainty is to be expected.

The complex dependence of systemic risk on the combination of asset size and connectivity (market cap and degree) that we observe in Figure 5.4 can be understood by considering the total losses to be absorbed by the system when an asset is shocked. Let us consider a shock to a very large market cap asset j (e.g., corresponding to the top row of Figure 5.4 where $[\omega_{\text{mcap}}] = 1$). A fall of $r_j\%$ in the price of j will result in a loss to shareholders equal to the product of r_j and the market cap v_j . This dollar-loss is distributed among the shareholders in proportion to the fractional investment of each shareholder in j . In the extreme case where there is a sole investor (asset degree = 1), the loss is borne entirely by that investor, and for a large loss they may well enter a margin call state and commence deleveraging. However, the potential price impact that a single fund in distress can cause on the rest of the financial system is limited — the distressed selling behaviour of a single fund (limited by market volume participation parameters η and δ , see Section 3.8) may not be sufficient to trigger a systemic crisis. However, as the number of affected funds increases (increasing ω_{degree}), we reach a critical point where the set of funds entering distress is sufficient to trigger systemic collapse with high probability. With a relatively small number of funds owning the shocked asset, each must bear a substantial dollar-loss, sufficient to force a substantial fraction of these funds into distress. However, as the number of funds invested in the shocked asset increases further, the average dollar-loss borne by each of them decreases monotonically. Some funds will still suffer a larger loss than others due to the heterogeneity of real-world investment sizes, but the overall trend will be toward a smaller loss burden to each affected fund as asset degree increases. Consequently, funds are less likely to enter margin call states and commence distressed selling for higher asset degree, which explains the reduction in systemic risk as asset degree increases for a given level of shocked asset market cap.

This result can be considered a consequence of the “diversification of assets”, and stands as a counterpoint to the “diversification of funds” results we presented in Section 3.10. In that previous section, we demonstrated that systemic risk reaches a maximum for intermediate levels of fund diversification, i.e., when distressed funds are connected to an intermediate fraction of available assets. We now offer the complementary result that systemic risk reaches a maximum for intermediate levels of *asset* diversification, that is, when shocked assets are owned by an intermediate fraction of the funds in the system. A consequence of this finding is that, for a shock to an asset with a given market capitalisation, other things being equal, systemic risk is

reduced if the shocked asset has either extremely few, or a relatively large number, of investors. Although any large asset will usually be systemically important, a large asset that is owned by a small number of leveraged investors will be particularly so, and we would argue that such assets should be of particular interest to regulatory authorities.

5.2.4 Discussion of Systemic Risk in the Real-World-Scale ABM

The results and methodology that we have presented in this chapter represent the first realisation of a full real-world-scale market microstructure agent based model of the network of overlapping fund–asset portfolios. Our model increases the realism of our earlier work (presented in Chapter 3) along several dimensions. Firstly, and most obviously, in terms of scale which has increased from 50 funds and 50 assets to 4593 funds and 6619 assets. We overcame substantial computational challenges in order to make millisecond-scale simulation feasible in an ABM featuring tens of millions of independent agents. Secondly, our work has increased the realism of multi-asset market microstructure modelling (itself a research area that has received little attention before now) by utilising asset microstructure models that have been quantitatively calibrated against extensive real world data (Chapter 4). Thirdly, we have used regulatory data (13-F portfolio holdings) to instantiate the real-world fund–asset network at a particular recent epoch. This dataset allows us to model the (long) holdings of the largest international managers of US equity assets. As such, our dataset comprehensively represents the liquid holdings of participants in the largest global equity market. Our model could be further enhanced if regulators were to provide researchers with access to idiosyncratic fund leverage declarations of the kind that are now collected via Form PF.⁹ In order to accommodate this limitation in available data we tested the effect of a range of leverage parametrisations on systemic risk in our ABM, establishing quantitative trends that will be of interest to regulators and the academic community who before now have lacked a microstructure-level model that spans the entire US market.

Given these substantial extensions to the model we presented in Chapter 3, it is perhaps surprising that we find that the results we presented with our previous model continue to stand in the full-scale ABM. In particular, we observe that, as expected, systemic risk increases with increasing fund leverage. More strikingly, we not only

⁹<https://www.sec.gov/about/forms/formpf.pdf>, accessed 2019-03-28.

observe that shocks to higher market-cap assets result in increased systemic risk (a result that is consistent with our observations in Section 3.10.4, with our calibration findings in Section 4.10, and empirical research by [Fry and Serbera \(2017\)](#)), but we also observe that, other things being equal, shocks to more highly connected assets result in *reduced* systemic risk. We attribute this latter finding to the effect of “fund diversification”, which complements the “asset diversification” results reported in Section 3.

Our results serve to validate the full-scale model. Not only does the model exhibit features that were predicted in a much smaller scale study in Chapter 3, it also facilitates the investigation of systemic risk with a level of quantitative rigour that would not be possible without the detailed calibration and validation methodology established in Chapter 4. However, the statistical significance of results generated with the ABM is a direct function of the number of Monte Carlo replications it is possible to perform within available computational resources ([Secchi and Seri, 2017](#)). The question of “how many samples are needed?” is not straightforward to answer. As a reminder, the ABM takes approximately one hour to simulate one hour of the market, on a 56 core HPC node. In the presence of such high computational costs it is necessary to target as few scenarios and Monte Carlo replications as possible. The approach we have adopted has been to monitor the progress of Monte Carlo experiments while they are still running, paying particular attention to the overlap or separation of confidence intervals. The number of trials it is necessary to collect is thus a function of the level of difference between scenario output distributions that it would be desirable to detect, given such a difference exists. We reject the null hypothesis that no difference exists between scenario distributions by performing t-tests, robust $\mathcal{X}^2$ tests, or by observing non-overlapping confidence intervals. We will continue to apply this method for the remainder of this chapter.

In this section we have established that the full-scale ABM performs in a manner that is consistent with our earlier 50-fund-50-asset simulations (Chapter 3). We have demonstrated how the full-scale model can be used to quantify systemic risk in the real-world financial system, and we have investigated the dependence of systemic risk on parameters including fund leverage, and the size and connectivity of shocked assets. In the sections that follow, we extend our analysis by relaxing assumptions regarding the uniform deleveraging behaviour of funds, and consider enhanced scenarios whereby funds preferentially target selected subsets of assets to sell. In all further experiments we control for the effect of asset degree established in this section.

5.3 Regulatory Policy: Price Limits and Circuit Breakers

Following episodes of panic selling in 1987 and 1988, the New York Stock Exchange (NYSE), working in collaboration with other major US trading venues and the US securities exchange commission (SEC, the US regulator), instituted a *market-wide circuit breaker* (MWCB) policy known as Rule 80b.¹⁰ The policy operates by suspending trading in *all* US-listed equity and derivative assets in response to falls in the value of the S&P 500 index (the market-cap weighted index of the largest 500 US equity assets) beyond a set of pre-defined trigger levels (relative to the previous day's closing index value). Prior to the Flash Crash of 6th May 2010, the “Level 1” MWCB threshold was -10% . However, the MWCB did not trigger during the Flash Crash, and as a result the MWCB system was amended in 2012 with stricter tolerances on S&P 500 declines. The MWCB pilot was adopted as permanent regulation in October 2019.¹¹ Details of the MWCB mechanism are presented in Section 5.3.2.

A further consequence of the 6th May 2010 Flash Crash was an overhaul of the single-stock circuit breaker (SSCB) rules, which were replaced (on a pilot basis) with the substantially more complex *Limit-Up-Limit-Down* (LULD) policy. The LULD pilot became permanent regulation in December 2018.¹² In contrast to the MWCB policy that suspends the entire market in response to market-level volatility, single-stock rules are designed to prevent excess volatility in a single asset at a time, with no consideration given to the state of other assets or to the wider market. The key difference between the older SSCB rules (as used in Chapter 3) and the current LULD policy is that SSCB rules operated as an *ex post* response to price falls in affected assets, whereas LULD rules specify *ex ante* limits on acceptable future price movements for an asset (Section 2.7). An asset is free to trade within LULD limits (which are updated throughout the trading day, as described in detail in Section 5.3.1). However, if bid and ask prices move such that no executable (passive) liquidity is available on the LOB within the price limits, the asset enters a five-minute trading halt phase.

By their very nature, MWCB and LULD activations are tail events, occurring rarely in comparison to non-crisis market activity. In particular, the MWCB has not activated since 27th October 1997¹³, well before the markets came to be dominated

¹⁰<https://www.nyse.com/markets/nyse/trading-info>, accessed 2019-10-06.

¹¹<https://www.sec.gov/rules/sro/nyse/2019/34-85560.pdf>, accessed 2019-08-26.

¹²<https://www.sec.gov/rules/sro/nyse/2019/34-85560.pdf>, accessed 2019-08-26.

¹³<https://www.sec.gov/files/dera-luld-white-paper.pdf>, accessed 2019-05-21.

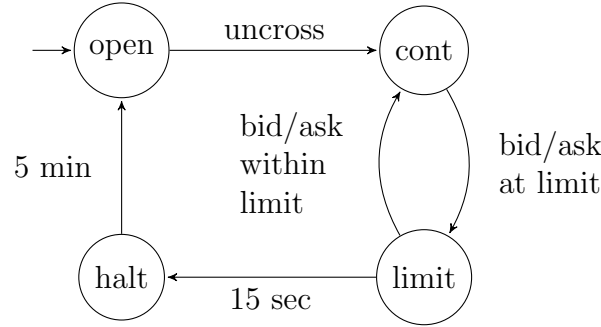


Figure 5.5: Finite state machine representing the operation of the Limit-Up Limit-Down (LULD) per-asset trading halt policy.

by trading algorithms (Buchanan, 2015). This property has hampered academic investigation into the effect of circuit breakers and price limits. Agent based models, on the other hand, offer policy makers the opportunity to explore as many alternative sets of market conditions as they choose. With a calibrated ABM, regulators can perform *in silico* evaluation of policies to model and analyse how they perform in crisis scenarios, and perform inference without the need to observe crisis events in the real-world.

In the sections that follow we describe in detail the LULD and MWCB policies that we have implemented in our ABM. We then proceed to investigate, over *intraday* timescales, the effectiveness of current and prospective policies at mitigating systemic risk. Our work represents the first time that the impact of market-wide circuit breakers and price limits on systemic risk has been studied using a market microstructure level ABM.

5.3.1 Limit-Up Limit-Down Rule

Figure 5.5 presents the operation of the LULD mechanism, as implemented in our ABM, as a finite state machine (FSM). Each asset in the simulation maintains an independent instance of the LULD FSM, which applies at all times except during the opening auction period (Section 3.3.6) or when the asset is halted due to a MWCB (interaction between LULD and MWCB is described in Section 5.3.2). The LULD FSM states are:

Open

The initial state of the LULD FSM is *open*. Trading is suspended during the *open* state, and instructions from buyers and sellers are queued. In reality this state is

fleeting (milliseconds at most), and so the effect on market participants is barely noticeable (delays are of a similar scale to expected delays due to communications latency). However, the *open* state performs the critical function of computing the price at which to open trading in the asset, following a halt period (including volatility halts and scheduled auctions). The opening price is selected as the price that would maximise the volume of executed orders that are resting on the LOB. In cases where the best bid remains below the best ask during the *open* state, no execution is possible and the opening price cannot be determined. In real-world exchanges, this event would typically result in an extension to the preceding halt period. In our ABM, we simplify slightly by falling back to the pre-halt price (this action is also taken by real-world exchanges if a price cannot be determined following an extended halt phase¹⁴). In cases where the best bid is *above* the best ask, the LOB is said to be *crossed*, and the effect of the *open* state is to uncross the LOB with the postcondition that the best bid is below the best ask. Traders are notified regarding any of their orders that were executed during the uncrossing process, and the LULD FSM transitions to the continuous trading (*cont*) state.

Cont(inuous)

During the continuous trading (*cont*) state, the LOB is open to orders and cancellations, and trading can take place in the usual manner. Executions occur when a new sell (buy) order matches against a resting buy (sell) order on the LOB. The operation of the LOB in our ABM during continuous trading is described in Section 3.2.1.

During the *cont* state, the LULD policy defines upper and lower price bands for the asset¹⁵. These bands determine the range of acceptable prices at which the asset can trade, and are specified as a fixed percentage offset from a *reference price*. The reference price is updated every 30 seconds if the arithmetic mean of trade prices observed over the previous five minutes is at least 1% away from the prevailing value. The reference price remains unchanged if no trades are observed in the preceding five minutes, and is initialised to the previous day's closing value (in our ABM we initialise to \$100 without loss of generality, as described in Section 5.1.3). Asset price bands are specified as a function of market cap. The largest 1000 US equities are designated *Tier 1*, with the remainder *Tier 2*. Since our ABM features over 6000 assets, we accommodate both Tiers in our model. Tier 1 and Tier 2 asset price bands

¹⁴<https://www.nyse.com/markets/nyse/trading-info>, accessed 2019-10-06.

¹⁵https://www.nasdaqtrader.com/content/MarketRegulation/LULD_FAQ.pdf, accessed 2019-04-30.

are fixed at reference price $\pm 5\%$ and $\pm 10\%$, respectively. We faithfully implement the reference price and band update semantics in our ABM.

Resting orders outside the LULD price bands are not executable, and so no aggressive (spread crossing) orders are permitted to match against such non-executable passive orders. The effect is that the best bid is prevented from rising above the upper price band, and the best ask is prevented from falling below the lower price band. Although for maximum generality we implement both upper and lower bands, the simulation results in this chapter only result in the best bid falling below the lower price band and the best ask falling to the lower price band. In order to simplify the descriptions that follow we thus focus on interactions with the lower price band, but the reader is requested to keep in mind that corresponding statements are equally applicable to the upper price band dynamics.

If the best ask reduces to the lower price band during continuous trading, the asset enters a *Limit State*, which we represent in our ABM with transition to the *limit* FSM state. We take care to correctly handle different order types in our model, including limit orders, market orders, and pegged orders (defined in Section 3.8).

Limit

During a Limit State¹⁶ (*limit*, in our notation) assets are still able to accept incoming orders to buy and sell within the LULD price bands, but, as with during the *cont* state, incoming orders will not match against resting orders that are outside the LULD bands. This ability to continue trading within price bands is a key difference introduced by the LULD mechanism from the preceding SSCB policy. The Limit State is thus intended to act as an early-warning indicator of volatility to investors.

During the *limit* state, price bands are not updated. The *limit* state can persist for a maximum of 15 seconds. If, during this time, the best ask rises above the lower price limit the *limit* state is cleared, and the LULD FSM transitions back to the *cont* state. However, if the *limit* state does not clear within the regulation 15 seconds, trading is suspended and the asset enters a *halt* state.

Halt

During a *halt* state, the LOB is able to accept new orders to buy and sell at any price (i.e., LULD price bands are not relevant during halts), but no trading occurs. The LOB will typically become *crossed* as it is likely (both in reality, and in our simulation)

¹⁶https://www.nasdaqtrader.com/content/MarketRegulation/LULD_FAQ.pdf, accessed 2019-04-30.

that the LOB will receive, or already hold, some orders to buy that are priced higher than some new or resting orders to sell. In our simulation, as in in reality, LULD halt auctions last for five minutes, at which point the LULD FSM transitions to the *open* state so that the market can be re-opened. The LULD *reference price* is reset to the uncrossing price.

5.3.2 Market-Wide Circuit Breakers

In our ABM, as in reality, the LULD mechanism remains in force at all times during regular trading hours (between the scheduled market open at 09:30 and close at 16:00), unless a MWCB event is in force. The MWCB state overrides any per-asset LULD states. Figure 5.6 represents our implementation of the interaction between the LULD and MWCB policies as a *hierarchical finite state machine* (HFSM), where the MWCB state operates at a strictly superior level to the per-asset LULD mechanism. The compound MWCB-LULD HFSM operates according to the following states:

Pre-Open

During the *pre-open* state, which represents the market opening auction, assets are halted (as per the LULD *halt* state definition, above). In our simulation, this state persists for 2 minutes, a duration selected to satisfy the trade-off between computational cost and adequate volume of orders on the LOB. The pre-open duration is substantially less than the initial transient duration of 4 minutes identified during calibration (Section 4.4) and so plays a more limited role in our simulation than in reality. We refrain from shocking our ABM until 12 simulated minutes have elapsed (representing 20% of total simulation duration, consistent with our systemic risk experiments in Chapter 3).

After two minutes have elapsed, the HFSM advances to the first *Continuous* state, and all assets move in unison to the LULD *open* sub-state.

Continuous

In each of the three *Continuous* states, the market is open for trading in the usual way, and each asset operates according to its own LULD FSM, as described in Section 5.3.1.

The level of the simulated S&P 500 index is calculated every 10 seconds during a *Continuous* state. This update frequency reflects the difficulties faced by regulators attempting to value the S&P 500 index during crises, due to excessive and uncertain

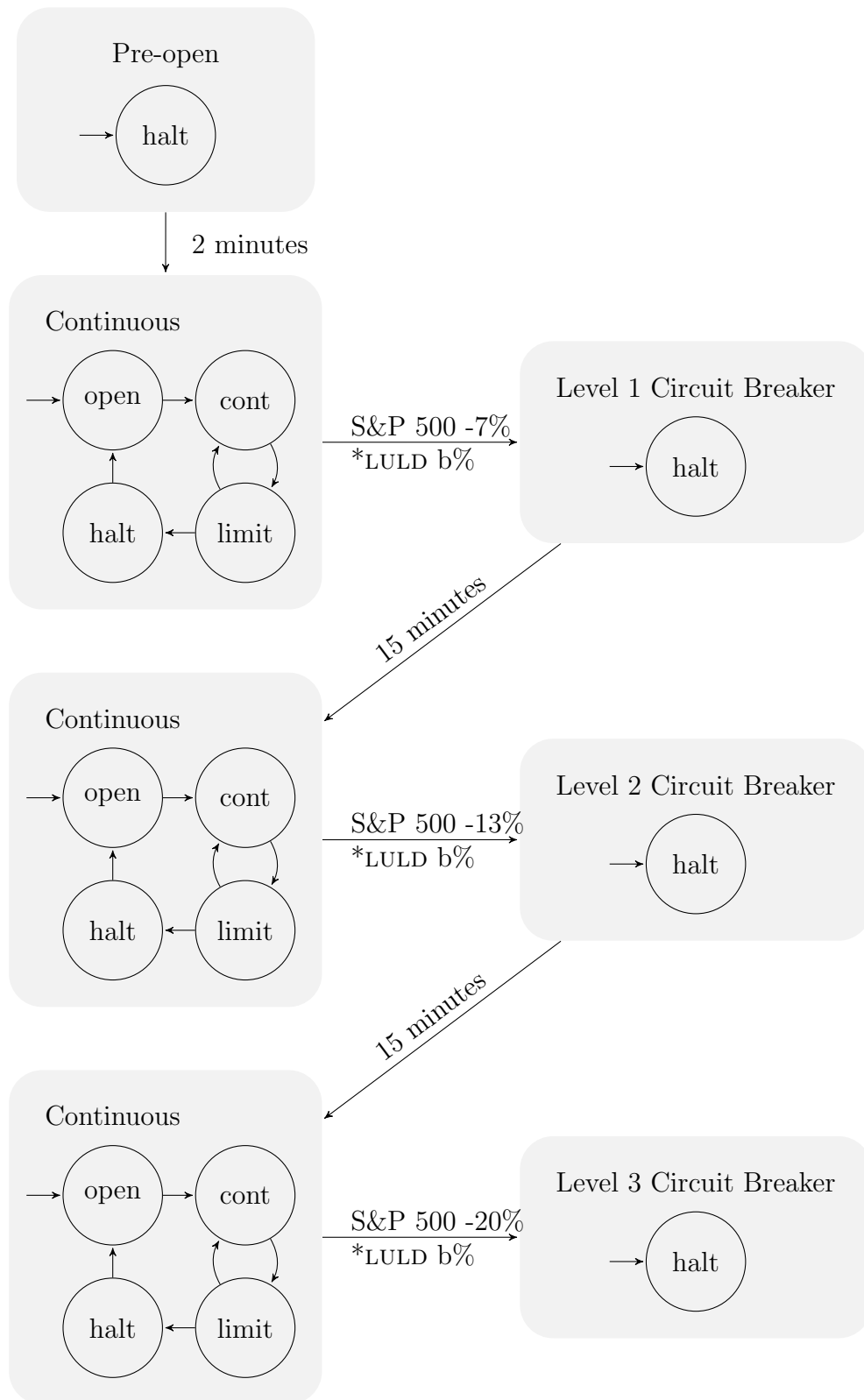


Figure 5.6: Hierarchical finite state machine representing the operation of the Limit-Up Limit-Down per-asset trading halt policy, concurrently with the market-wide circuit-breaker policy. MWCB triggers marked with (*) represent amended trigger conditions due to alternative policy specifications discussed in Section 5.5.

feed latency (CFTC and SEC, 2010) and stale prices due to individual asset halts (BlackRock, 2015). The S&P 500 value is given by

$$I_{\text{SP500}}^t = \frac{1}{\sum_{j \in \mathcal{I}} w_j} \sum_{j \in \mathcal{I}} w_j p_j^t, \quad (5.3)$$

where $\mathcal{I} = \{j \in (1, \dots, n_a) \mid w_j > 0\}$ is the set of assets that are members of the S&P 500 index. The weight of each asset in the index, w_j , is defined in terms of its market cap by

$$w_j = \begin{cases} v_j'' & \text{if } r(j, \mathbf{v}'') \leq 500, \text{ and} \\ 0 & \text{otherwise,} \end{cases} \quad (5.4)$$

where $r(j, \mathbf{v}'') : (\mathbb{N}, \mathbb{R}) \rightarrow \mathbb{N}_+$ is a function returning the rank of asset j in the *descending* ordering defined by asset market caps $\mathbf{v}''$. For convenience, weights w are normalised to sum to unity, and because $p_j^0 = 100 \ \forall j \in (1, \dots, n_a)$, $I_{\text{SP500}}^0 = 100$ also. Let the index return at t be defined as

$$R_{\text{SP500}}^t = \frac{(I_{\text{SP500}}^t - I_{\text{SP500}}^0)}{I_{\text{SP500}}^0}, \quad (5.5)$$

then the MWCB policy mandates that $R_{\text{SP500}}^t < -7\%$ triggers a *Level 1* MWCB. We represent this in our HFSM as a transition to the *Level 1 Circuit Breaker* state. All assets simultaneously enter the *halt* sub-state, overriding any LULD state they held during the *Continuous* phase.

Level 1 Circuit Breaker

During the *Level 1 Circuit Breaker* state all assets are halted for a period of 15 minutes. Any partially completed LULD halts are subsumed by the MWCB state, meaning that after 15 minutes all assets open simultaneously as the HFSM transitions to the second *Continuous* state.

Continuous (second instance)

Market activity during the second *Continuous* state (see Figure 5.6) is identical to that in the first *Continuous* state, with the exception that the level at which the MWCB triggers is reduced from $R_{\text{SP500}}^t < -7\%$ to $R_{\text{SP500}}^t < -13\%$.¹⁷ If R_{SP500}^t falls

¹⁷<https://www.nyse.com/markets/nyse/trading-info>, accessed 2019-10-06.

below -13%, a second MWCB is triggered and the HFSM transitions to the *Level 2 Circuit Breaker* state.

Level 2 Circuit Breaker

This circuit breaker operates identically to the Level 1 Circuit Breaker, and after 15 minutes of halted trading, during which orders to buy and sell are accepted but no orders are matched for execution, the market reopens and for the final time transitions to the *Continuous* state.

Continuous (third instance)

As with the second instance of the *Continuous* state, the MWCB activation trigger is again reduced, this time to $R_{SP500}^t < -20\%$. If the index value falls below -20% the third and final MWCB is activated, and the HFSM transitions to the *Level 3 Circuit Breaker* state.

Level 3 Circuit Breaker

The *Level 3 Circuit Breaker* state is final, and the market remains closed for the remainder of the trading day. There is no closing auction, and all assets conclude the day in the halt state.

In reality, activation of the Level 3 Circuit Breaker would be far from the last substantial action to occur in the financial markets that day, as US investors would find themselves unable to execute in domestic markets, and may be forced to take action in alternative geographical regions in order to hedge (offset) positions and reduce exposures to historically unprecedented levels of financial market distress ([Cliff and Northrop, 2012](#)).

5.4 Simulated Regulatory Interventions

In this section we take our full-scale calibrated micro-macro ABM and investigate the effect on systemic risk of:

1. the composite LULD and MWCB trading halt policy, and
2. alternative fund deleveraging behaviours.

In order to aid the analyses that follow, it is useful to introduce some notation. Let us refer to the full current US regulatory standard composite Limit-Up-Limit-Down

and market-wide circuit breaker policy as the LULD-MWCB policy. Results generated outside the current LULD-MWCB policy (such as those presented in previous chapters and sections in this thesis) will be designated as using the NONE halt policy.

Additionally, in this section we investigate the impact of alternative fund deleverage behaviours on systemic risk. In all results prior to the current section, funds have deleveraged unconditionally. That is, upon entering the distressed margin-call state they commenced selling *every* asset in their portfolio (Section 3.3.5). Let us henceforth refer to this behavioural policy as the *unconditional* policy, denoted $\mathcal{U}$.

Capponi and Larsson (2015), Duarte and Eisenbach (2018), and Cont and Schaanning (2019) each propose that an alternative fund deleverage policy is optimal for systemic risk. Under this policy funds preferentially liquidate the *highest* market-cap assets in their portfolios (this is not to be confused with liquidating the largest *positions*). Let us refer to such a high-market-cap preferential policy as $\mathcal{H}$. In our ABM, under the $\mathcal{H}$ behavioural policy, upon entering the distressed margin-call state, funds commence selling a fraction $\phi \in (0, 1]$ of the assets in their portfolio, yielding the initial set of assets to unwind $\mathcal{A}_i = \{j \in (1, \dots, n_a) \mid r(j, \mathbf{v}'') \leq \phi n_a\}$ for fund $i \in (1, \dots, n_f)$. In our ABM, if an asset $j \in \mathcal{A}_i$ enters a LULD *halt* phase, under the $\mathcal{H}$ policy, the owning fund will respond by commencing trading in the next largest un-halted asset $j' \notin \mathcal{A}_i$. The set $\mathcal{A}_i$ of assets to unwind thus grows in order to maintain the fund's desired level of selling activity ($\mathcal{A}_i \rightarrow \mathcal{A}_i \cup \{j'\}$). Should asset j become tradable once more, the fund is assumed to resume distressed selling in j . This assumption models the unwillingness of funds to repeatedly place and then cancel large meta-orders (Zarinelli et al., 2015). Rather, once a decision to sell out of a particular asset position has been taken, funds are assumed to persist in this decision while the margin-call state is in effect. In our model, funds evaluate the trading halt status of the assets in their portfolios every ten simulated seconds.

In order to test the effect, and to facilitate comparative studies, of the $\mathcal{H}$ policy on intraday systemic risk in the full-scale calibrated micro-macro ABM, we also define two related policies $\mathcal{L}$ and $\mathcal{R}$. The first policy, $\mathcal{L}$, is the exact converse of $\mathcal{H}$ and instead causes funds to preferentially unwind their *lowest* market cap assets. The policy is identical to $\mathcal{H}$ in all regards except the rank ordering function r is reversed such that $r'(j, \mathbf{v}'') : (\mathbb{N}, \mathbb{R}) \rightarrow \mathbb{N}_+$ returns asset market cap ranks ordered from smallest to largest. Finally, as a control, we define a *randomised* behavioural policy $\mathcal{R}$. The $\mathcal{R}$ policy operates identically to the $\mathcal{H}$ and $\mathcal{L}$ policies, except the set of assets to be unwound is independently randomised for each distressed fund. This is also straightforwardly achieved by modifying the rank ordering function r , in this case

letting $r''(n_a) = \lfloor \text{Uniform}(1, n_a) \rfloor$ (with random tie breaking). The purpose of $\mathcal{R}$ is to control for the number of assets that each fund seeks to unwind when distressed. By comparing results under $\mathcal{R}$ with those under $\mathcal{L}$ or $\mathcal{H}$ it is possible to attribute the effect on systemic risk to the properties of the assets being sold, rather than simply to the number of assets being sold, as is the case when comparing results under $\mathcal{U}$ with results under $\mathcal{L}$ or $\mathcal{H}$.

In our implementation, the set of distressed assets is non-decreasing, and so the unconditional policy $\mathcal{U}$ is approached in the asymptotic limit of all the other (conditional) policies. Collectively, let us refer to the set of fund behavioural policies as $\mathcal{F} = \{\mathcal{U}, \mathcal{H}, \mathcal{R}, \mathcal{L}\}$. Since the funds in our model are assigned a particular behaviour (homogeneously across funds, in this chapter), and do not undertake any kind of behaviour selection or optimisation process in the experiments that follow, we continue to classify funds as zero-intelligence agents.

5.4.1 Experimental Scenarios

In the sections that follow, we evaluate systemic risk under the LULD-MWCB policy for experimental scenarios $\mathcal{S} = \mathcal{F} \times \{\omega_{\text{mcap}}^{\text{high}}, \omega_{\text{mcap}}^{\text{low}}\}$, where $\omega_{\text{mcap}}^{\text{high}} \equiv \lceil \omega_{\text{mcap}} \rceil = 1$, and $\omega_{\text{mcap}}^{\text{low}} \equiv \lceil \omega_{\text{mcap}} \rceil = 0.6$. We thus define $|\mathcal{S}| = 8$ experimental scenarios.

For each scenarios $s \in \mathcal{S}$ we assume:

1. Leverage is constant across trials, and homogeneous across funds, at $\lambda^0 = 5$. Following the analysis under the NONE policy in Section 5.2.2, this value was found to be both realistic, and sufficient to give rise to significant systemic risk, for $\omega_{\text{mcap}} \geq 0.6$. Capital varies heterogeneously across funds as described in Section 5.1.3.
2. Fund margin tolerance hysteresis is homogeneous across funds at $\tau_c = 1.15$, consistent with previous sections, and fund distressed selling volume fraction η and timescale δ also remain homogeneous across funds and fixed across trials at $\eta = 0.09$ and $\delta = 10$ seconds, as in previous sections.
3. The asset to which we apply the initial shock is selected uniformly at random from the set of assets from the market cap quintile corresponding to the value of ω_{mcap} for the given trial $s \in \mathcal{S}$. We perform median tests in order to confirm that there is no significant difference in the distribution of shocked asset degrees between scenarios (see Section 5.4.2).
4. Under behavioural policies $\mathcal{H}$, $\mathcal{R}$, and $\mathcal{L}$, the fraction ϕ of assets to be unwound by distressed funds is held constant across trials, and is homogeneous across

funds at $\phi = 10\%$. Results in Chapter 3 indicate that network topology and asset diversification have significant effects on systemic risk. A value of ϕ that is too small would act to dampen the systemic effect of distressed funds to a degree that could render experiments meaningless. A value of ϕ that is too large would cause each conditional policy to asymptotically approach $\mathcal{U}$ as $\phi \rightarrow 100\%$. In preliminary experiments, $\phi = 10\%$ was found to offer a suitable, and empirically plausible, parameter setting. We would argue that $\phi = 10\%$ is a somewhat conservative setting, and real-world behaviour during distressed deleveraging spirals could certainly give rise to the observation of higher realised values of ϕ .

Computational resource requirements present significant limitations to the number of experimental scenarios it is feasible to test in this section, and the number of Monte Carlo evaluations possible for each scenario. The set of scenarios we present thus represents the minimal set that we believe is necessary in order to characterise systemic risk in the full-scale ABM under realistic trading halt policies, a set of widely differing yet plausible fund behaviours, and alternative initial asset shock profiles.

A quantitative calibration of fund behaviour, in the formal terms introduced in Chapter 4, would be an extremely challenging undertaking, indeed it would represent the final evolution of regulatory ABM sophistication to Level 3 of Axtell and Epstein (1994)'s taxonomy. The results that follow offer arguments in justification of why such an extensive undertaking could be warranted.

5.4.2 Effect of Fund Behaviour on Probability of Contagion in Presence of Trading Halts

Figure 5.7 (top) presents the probability of contagion ($P_{contagion}$, Equation 3.6), for each of the eight experimental scenarios $s \in \mathcal{S}$. Probability of contagion is the fraction of Monte Carlo replicates of the ABM simulation in which at least $\gamma = 5\%$ of funds enter default (collapse into bankruptcy). Results are based on 318 Monte Carlo replicates (ω_{mcap}^{high}) and 50 Monte Carlo replicates (ω_{mcap}^{low}) with cell shading indicating $P_{contagion}$ and labelling indicating the 95% confidence interval.

Figure 5.7 (top) shows that, under all fund behaviour policies, $P_{contagion}$ increases with increasing ω_{mcap} . This result is consistent with the previous studies we have reported in this chapter. It is possible to make a direct comparison with the results presented in Figure 5.3 by recalling that we hold leverage fixed at $\lambda^0 = 5$ for the current experiment, and that all funds use the unconditional ($\mathcal{U}$) policy in Figure 5.3 (and indeed for all results presented before the current section). We see that the

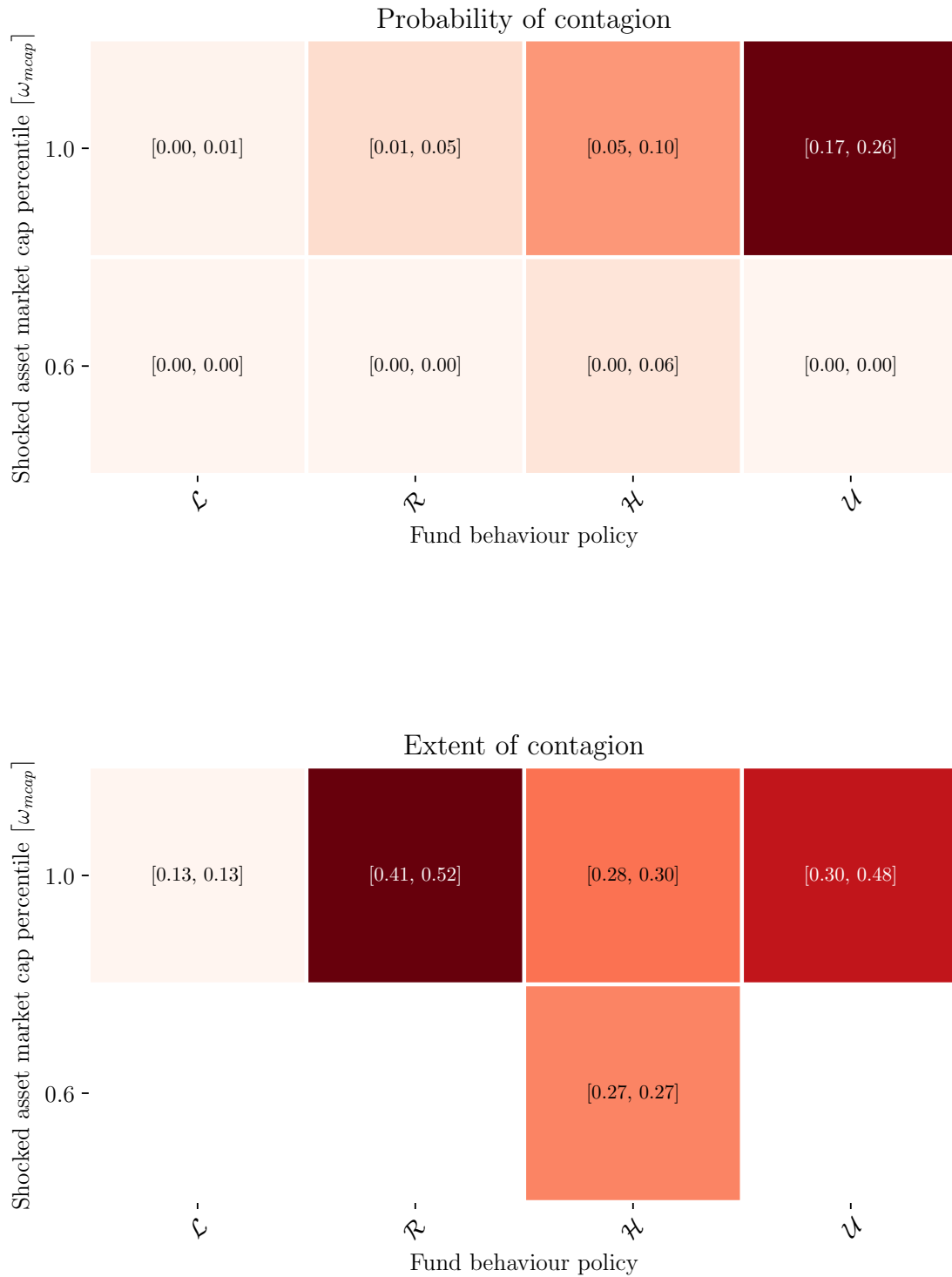


Figure 5.7: Comparing (top) the probability of contagion, and (bottom) the extent of contagion for shocked asset centrality quintiles of ω_{mcap} and fund behavioural policies $\mathcal{F}$ at leverage $\lambda^0 = 5$, under the current US regulatory standard LULD-MWCB trading halt policy. Results are based on 318 Monte Carlo replicates ($\omega_{mcap}^{\text{high}}$) and 50 Monte Carlo replicates ($\omega_{mcap}^{\text{low}}$), with cell shading indicating the mean, and labelling indicating the 95% confidence interval.

effect of replacing the NONE halt policy with the LULD-MWCB halt policy is to reduce $P_{contagion}$ from $10(\pm 10)\%$ to $0(\pm 0)\%$ for ω_{mcap}^{low} and from $67.5(\pm 16.5)\%$ to $21.5(\pm 4.5)\%$ for ω_{mcap}^{high} . Thus our simulation results establish the significant effect of price limit and circuit breaker safety mechanisms in the market. These results are consistent with the hypothesis that more stringent price control policies result in reduced systemic risk. We devote the remainder of this chapter to investigating the detailed mechanisms by which effective price control is maintained, and quantitatively considering possible alternative specifications. Figure 5.7 (top) clearly establishes the need to consider alternative specifications because although we observe reduced systemic risk compared with the NONE policy, we do continue to observe the catastrophic price-mediated failure of the financial system under a quantitatively realistic calibrated model, and so the current real-world policy LULD-MWCB is clearly insufficient in some circumstances.

Although we have by now firmly established the effect of ω_{mcap} on systemic risk, Figure 5.7 (top) also reveals that systemic risk is strongly dependent on the distressed unwinding behaviour adopted by funds. Before discussing this finding in detail, it is necessary to correct for the effect of ω_{degree} between experimental scenarios in our analysis. Since we established in Section 5.2.3 that systemic risk is strongly dependent on both ω_{mcap} and ω_{degree} , if the value of ω_{degree} varies systematically between experimental scenarios then we would not be able to tell if any observed differences between scenarios were due to the scenario treatments, or due to the uncontrolled effect of ω_{degree} . In order to assure that this is not the case, we performed robust χ^2 tests¹⁸ for the distribution of ω_{degree} across Monte Carlo replicates of each pair of experimental scenarios. At the 5% level the only significant difference in the distribution of ω_{degree} was between the behavioural policies $\mathcal{H}$ and $\mathcal{U}$ under ω_{mcap}^{low} . We therefore refrain from making direct comparisons of results between these two scenarios. However, this is a minor inconvenience that does not significantly affect our conclusions. We also applied the robust χ^2 test to the distribution of ω_{degree} between the results in Figures 5.3 and 5.7 (top). No significant difference was observed, and so the reduction in systemic risk between the two sets of scenarios cannot be due to the uncontrolled effect of ω_{degree} . For completeness, we also applied the robust χ^2 test between scenarios with equal ω_{mcap} (but differing λ^0) in Figure 5.3. The only significant differences were for $[\omega_{mcap}] = 0.4$, for the pairs $(\lambda^0 = 3.7, \lambda^0 = 5)$ and $(\lambda^0 = 3.7, \lambda^0 = 6.3)$. Our results are unaffected by this observation.

¹⁸We apply Mood's χ^2 , which tests the null hypothesis that the medians of two empirical distributions are the same. This test was found to be robust to outliers in ω_{degree} and is consistent with our use of percentile binning for ω_{degree} and ω_{mcap} .

With these precautions in place, we can now confidently state that systemic risk is minimised under the behaviour policy that mandates preferential sale of lower-liquidity assets in the funds' portfolios ($\mathcal{L}$). Default cascades are only observed in $0.5(\pm 0.5)\%$ of cases under $\omega_{\text{mcap}}^{\text{high}}$, and not at all under $\omega_{\text{mcap}}^{\text{low}}$. Let us focus firstly on $\omega_{\text{mcap}}^{\text{high}}$. Here, $P_{\text{contagion}}$ increases to $3(\pm 2)\%$ under behaviour policy $\mathcal{R}$ whereby funds unwind a subset of their portfolio selected uniformly at random. $P_{\text{contagion}}$ increases further to $7.5(\pm 2.5)\%$ under the high-liquidity targeting behaviour policy $\mathcal{H}$, and is maximised under the unconditional policy $\mathcal{U}$ with a value of $21.5(\pm 4.5)\%$ as mentioned above. The pattern is similar for $\omega_{\text{mcap}}^{\text{low}}$, where $P_{\text{contagion}}$ is everywhere 0 except under the high-liquidity preferential policy $\mathcal{H}$ where it attains a maximal value of $3(\pm 3)\%$. We thus offer evidence that, under realistic price formation assumptions and market regulatory frameworks, systemic risk *increases* as funds preferentially unwind the high-liquidity components of their portfolios. This result is in direct contrast to the theoretical findings by [Capponi and Larsson \(2015\)](#) and [Duarte and Eisenbach \(2018\)](#) who find that distress contagion is maximised under a policy equivalent to $\mathcal{L}$. Our results also fail to support [Cont and Schaanning \(2019\)](#) who claim the optimal liquidation policy is equivalent to $\mathcal{H}$, though the authors acknowledge that ([Cont and Schaanning, 2019, p.97](#)) “it is not clear that the most liquid asset ex-ante should still turn out to be the most liquid ex-post. Including such considerations requires a dynamic model for the market depth, which is outside the scope of this paper”. We claim that the model and analysis we are presenting in the current chapter provides precisely the kind of dynamic model of market depth referred to by [Cont and Schaanning \(2019\)](#). Indeed, it is by considering the dynamics of liquidity supply under distressed conditions that we are able to make sense of our results, as we will now describe.

The results we have presented in Sections 5.2.2 and previously in Section 4.10 are both consistent with the notion that, *ceteris paribus*, higher market cap assets are more prone than lower market cap assets to exhibiting flash crashes under conditions of distressed selling. It is therefore to be expected that fund behavioural policies that specifically focus distressed selling activity on higher market cap assets (policy $\mathcal{H}$) will be more likely to trigger flash crashes in those assets compared with policies that target lower market-cap subsets ($\mathcal{R}$ or $\mathcal{L}$). Thus, not only will price impact be more extreme under $\mathcal{H}$ compared to $\mathcal{R}$ or $\mathcal{L}$, the consequent total dollar loss to be absorbed by owning funds will be larger for failure of higher market cap assets, and furthermore, the distress will be felt by a larger number of funds because higher market cap assets have more owning funds than lower market cap assets (Section

5.1.1). Our results therefore suggest that modelling the dynamic liquidity provision process (1) facilitates a better understanding of the propagation of systemic risk, and (2) has the potential to *invert* findings generated by macro-level models with simplified price impact assumptions. This observation provides a key argument in support of our hybrid micro-macro agent-based approach to understanding systemic risk in the financial system.

Figure 5.7 (top) reveals one further pattern that warrants discussion. Although the statistical significance is limited for $\omega_{\text{mcap}}^{\text{low}}$, the unconditional policy $\mathcal{U}$ gives rise to lower $P_{\text{contagion}}$ than policy $\mathcal{H}$, whereas for $\omega_{\text{mcap}}^{\text{high}}$, $P_{\text{contagion}}$ is maximised under $\mathcal{U}$. This result could be a consequence of the failed robust χ^2 test between $\mathcal{U}$ and $\mathcal{H}$ under $\omega_{\text{mcap}}^{\text{low}}$, and so we are cautious about claiming too much for this result. With that caveat in mind, a possible explanation is that, similarly to observations presented in Section 3.10.2, and presented by previous researchers (Acemoglu et al., 2015a; Gai and Kapadia, 2010), the present results indicate a phase change in the effect of diversification as shock magnitude increases. Under lower shock regimes enhanced diversification is beneficial to systemic risk because funds are more able to meet their margin tolerance requirement by selling all of their assets unconditionally, thus allowing funds to readily exit the distressed state (and cease distressed selling). For high shock scenarios, higher asset diversification simply becomes a channel for enhanced distress propagation as funds are not able to sell sufficient portions of their portfolio to exit the margin call state without triggering catastrophic price impact in the form of flash crashes. For the reasons discussed, this result is tentative, and more data would help establish whether it is a genuine effect, or an artefact of the low number of Monte Carlo replicates (50) at $\omega_{\text{mcap}}^{\text{low}}$.

5.4.3 Effect of Fund Behaviour on Extent of Contagion in Presence of Trading Halts

The lower plot of Figure 5.7 presents the extent of contagion ($\Omega_{\text{contagion}}$, Equation 3.10) for each experimental scenario. The extent of contagion is the fraction of funds that end up in default, conditional on the occurrence of a default cascade (i.e., when at least $\gamma = 5\%$ of funds enter default). Thus $\Omega_{\text{contagion}}$ is only defined for scenarios with non-zero probability of contagion ($P_{\text{contagion}}$). For almost all pairs of experimental scenarios, relative systemic risk according to $\Omega_{\text{contagion}}$ is consistent with relative systemic risk according to $P_{\text{contagion}}$ in the sense that the member of the pair of scenarios with higher $P_{\text{contagion}}$ is also found to have higher $\Omega_{\text{contagion}}$. There are two exceptions and we discuss these below. Before doing so, we also make a general point

that applies to all experimental scenarios. In the absence of price limits and circuit breakers, and under an unconditional fund behaviour policy, our previous results demonstrated the *robust-yet-fragile* (Gai and Kapadia, 2010) nature of the fund–asset network (Section 3.10.2). Our results confirmed previous findings from macro-level models (Caccioli et al., 2014) that even where systemic collapse is rare ($P_{contagion}$ is small), when a collapse occurs, distress contagion is extensive, with almost all funds ending up in default ($\Omega_{contagion}$ is large). However, Figure 5.7 (bottom) reveals that in the fully-calibrated micro-macro model, $\Omega_{contagion}$ remains relatively modest, reaching a maximum of $46.5(\pm 5.5)\%$ under scenario $(\mathcal{R}, \omega_{mcap}^{\text{high}})$. While a catastrophically high fraction of funds end up in default in this scenario, the value falls well short of the near 100% conditional failure rates observed in Chapter 3. In order to determine whether this improvement is due to the introduction of price limits and circuit breakers, we can compare $\Omega_{contagion}$ under the fund behavioural policy $\mathcal{U}$ and trading halt policy LULD-MWCB with the results obtained in Section 5.2.2, where the implicit fund behaviour policy was $\mathcal{U}$ and the implicit halt policy was NONE. As Figure E.1 makes readily apparent, the expected extent of contagion is significantly reduced under the LULD-MWCB policy, compared to the NONE policy where $\Omega_{contagion}$ was observed to be $99(\pm 0)\%$ in all scenarios with positive $P_{contagion}$. We thus establish a strong connection between the activity of price limits and circuit breakers, and the probability and extent of financial contagion over intraday timescales.

Figure 5.7 (bottom) also demonstrates that it is not possible to ignore the effect of fund behaviour when assessing $\Omega_{contagion}$. Behavioural policies that limit distressed selling to particular subsets of assets ($\mathcal{L}$, $\mathcal{H}$) give rise to lower extents of contagion than policies that spread distress uniformly throughout the network ($\mathcal{R}$, $\mathcal{U}$). To make sense of this observation, consider the $\mathcal{L}$ policy. Funds that enter distress will preferentially sell their lowest market cap assets. Such assets will also likely be among the lowest market cap subset of the portfolio of other funds, and so will also be sold by those other funds, should they enter distress. Thus the set of assets under distress will grow slowly, whilst being restricted to those held by funds that have meaningful investments in the subset of distressed assets. Therefore, slow growth of the set of distressed assets implies slow growth in the set of distressed funds, and hence low $\Omega_{contagion}$. This argument applies equally to the $\mathcal{L}$ and $\mathcal{H}$ policies. For the $\mathcal{R}$ policy each new fund that enters default will target a different subset of assets, which are not expected to have high overlap with the set of assets already under distress, and thus the size of the set of distressed assets will grow, which will trigger distress at a correspondingly growing set of funds, leading to high $\Omega_{contagion}$.

Although there is some overlap of the confidence intervals of the $\mathcal{R}$ and $\mathcal{U}$ policies under $\omega_{\text{mcap}}^{\text{high}}$ in Figure 5.7 (bottom), it is interesting that $\Omega_{\text{contagion}}$ under $\mathcal{R}$ is at least as extensive as under $\mathcal{U}$. This observation may be explained by considering that funds can more readily exit their margin call state under $\mathcal{U}$ than $\mathcal{R}$. This is because the dollar value of a fund's portfolio will decrease more rapidly if more positions are closed concurrently, allowing the fund to more readily cease distressed selling before giving rise to overwhelming price impact. Our model realistically assumes that a fund that is not in distress plays no role in furthering the spread of systemic risk.

The final observation we wish to make for Figure 5.7 (bottom) is that, conditional on the occurrence of a default cascade (the failure of a significant fraction of the financial system), we would expect that the location of the initial shock delivered to the system is of limited relevance. Once a truly systemic event is in progress, the emergent dynamics of the resulting margin spiral (Brunnermeier and Pedersen, 2009) imply that the distress of many funds will not be due to their direct exposure to the initially shocked asset (so-called *first-round* effects), rather, fund distress will be due to overlap with the portfolios of other distressed funds (*second-round* effects). Figure 5.7 (bottom) demonstrates that this is indeed the case, as we observe no difference between the extent of contagion between the $(\mathcal{H}, \omega_{\text{mcap}}^{\text{high}})$ and $(\mathcal{H}, \omega_{\text{mcap}}^{\text{low}})$ experimental scenarios. This result confirms the findings reported by Caccioli et al. (2014, p.239) “the conditional extent of contagion [is] the same for both types of shock. The reason is simple: while the initial conditions of these two processes are different, their dynamics are the same. Once a cascade has begun, it doesn't matter what kind of shock began it”.

5.4.4 Action of Trading Halts Under Alternative Fund Behaviour

In order to understand how the composite Limit-Up-Limit-Down and Market-Wide Circuit-Breaker LULD-MWCB policy reduces systemic risk (as measured by $P_{\text{contagion}}$ and $\Omega_{\text{contagion}}$), it is instructive to consider whether the effect of the policy is (a) to prevent funds entering distress in response to the triggering asset price shock or (b) to dampen the spreading of systemic risk conditional on one or more funds entering distressed states, or some combination of these effects. We found that, for $\omega_{\text{mcap}}^{\text{high}}$, the probability that at least one fund enters distress is approximately 93%, decreasing to 64% for $\omega_{\text{mcap}}^{\text{low}}$ (Figure E.4). Note that the unwind behaviour policy of funds cannot affect the probability that the first fund enters default, which by construction can only be a function of the shock to the initial asset.

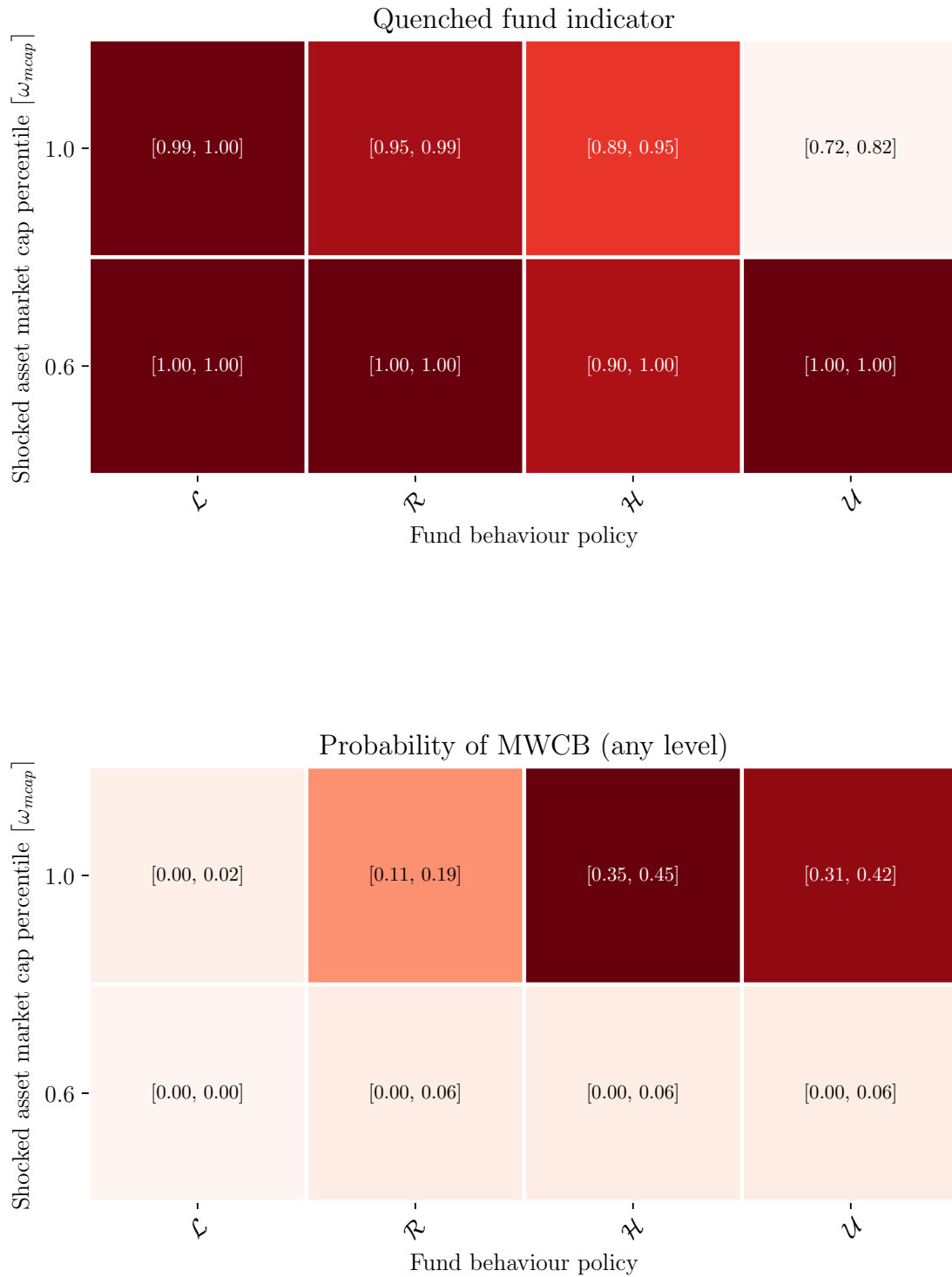


Figure 5.8: Comparing (top) fund distress cascade quenching, and (bottom) market-wide circuit-breaker activation for shocked asset centrality quintiles of ω_{mcap} and fund behavioural policies $\mathcal{F}$ at leverage $\lambda^0 = 5$, under the current US regulatory standard LULD-MWCB trading halt policy. Results are based on 318 Monte Carlo replicates ($\omega_{mcap}^{\text{high}}$) and 50 Monte Carlo replicates ($\omega_{mcap}^{\text{low}}$), with cell shading indicating the mean, and labelling indicating the 95% confidence interval.

In order to determine whether effect (a) is present (i.e., whether LULD-MWCB helps prevent funds entering distressed states) we can compare the probability that at least one fund will enter distress under the LULD-MWCB policy (for unconditional fund behaviour policy $\mathcal{U}$) and under the NONE policy (presented in Appendix E, Figures E.4 and E.2) where the implicit fund behaviour policy was $\mathcal{U}$. We find that under $(\mathcal{U}, \text{NONE})$ the probability that at least one fund will enter the margin call state is $95.5(\pm 4.5)\%$ for $\omega_{\text{mcap}}^{\text{high}}$ and $81(\pm 14)\%$ for $\omega_{\text{mcap}}^{\text{low}}$. The reduction that we observe under the $(\mathcal{U}, \text{LULD-MWCB})$ policy indicates that effect (a) is somewhat present, and that LULD-MWCB acts to reduce the probability that one or more funds will enter distress in response to the falling price of the initially-shocked asset. We further observe that this increased systemic resilience must be due to the effect of the LULD component rather than the MWCB component. This is because under LULD-MWCB the MWCB triggers for falls in the value of the index of the largest 500 assets of at least 7% (under current US regulations, and as simulated in the current chapter). The index is weighted according to the market capitalisation of the constituent assets, but the largest weighting of any asset is 3.8% for the 13-F network on the date we consider (2018-04-17). So a fall of even 100% in the value of the initial shocked asset is not sufficient to trigger the MWCB. This implies that the MWCB cannot trigger until further assets experience significant price falls, and this in turn cannot occur until *after* one or more funds commence distressed selling, thus any reduction in the probability that one or more funds enter distressed selling cannot be due to the MWCB and must instead be due to the LULD component.

However, since the fraction of trials in which a fund enters distress is much higher than the fraction of trials resulting in a default cascade ($P_{\text{contagion}}$), it must be the case that effect (b), above, is *also* present. That is, funds often enter distress, but the dynamics of systemic risk propagation under the LULD-MWCB policy are such that a catastrophic collapse is often averted. Let us define the term *quenched* to describe such extinguished cascades. Informally, a quenched cascade refers to a Monte Carlo replicate wherein at least one fund enters a distressed state, and where the eventual fraction of funds in default remains below the contagion threshold ($\gamma=5\%$). Formally, let us define the probability of quenching to be

$$P_{\text{quench}} = \frac{1}{M} \sum_{m=1}^M \mathbb{I}_{\text{quench}}(m), \quad (5.6)$$

where M is the number of Monte-Carlo replications of a single experimental scenario. Here, $\mathbb{I}_{quench}(m)$ is an indicator variable representing the occurrence of a quenched cascade during trial $m \in (1, \dots, M)$, and is defined as

$$\mathbb{I}_{quench}(m) = \begin{cases} 1 & \text{if } f_{default}(m) \leq \gamma \text{ and } |\mathcal{Z}(m)| \geq 1, \\ 0 & \text{otherwise,} \end{cases} \quad (5.7)$$

where $f_{default}(m)$, the fraction of funds entering default during Monte Carlo trial m , is given by Equation 3.8, and $\mathcal{Z}(m)$ is the set of funds entering a margin call state during trial m , formally

$$\mathcal{Z}(m) = \{i \in (1, \dots, n_f) \mid \exists t \in (0, \dots, T) \text{ s.t. } [\lambda_i^t(m) > \tau_c \lambda^0]\}, \quad (5.8)$$

where λ^0 and τ_c , initial fund leverage and margin tolerance hysteresis respectively, are constant across funds and for all Monte Carlo replicates at a given experimental scenario, and $\lambda_i^t(m)$ is the implied leverage for fund $i \in (1, \dots, n_f)$ at time $t \in (0, \dots, T)$ for Monte Carlo replicate m of a given experimental scenario (please see Section 3.4.1 for further details of leverage calculations in the model).

For completeness, we present P_{quench} with 95% confidence intervals for each of the experimental scenarios under consideration in Figure 5.8 (top). As expected, we observe that under the LULD-MWCB policy quenching is a very likely outcome for most combinations of ω_{mcap} and fund behaviour policy, but that quenching is less likely for fund behaviour policies targeting higher market cap asset subsets under ω_{mcap}^{high} , corresponding to higher $P_{contagion}$ in those cases (as shown in Figure 5.7 (top)).

Since we have established that it is often the case that funds enter a distressed state without an ensuing cascade of fund defaults, let us consider which aspect of the LULD-MWCB policy (the limit-up-limit-down or market-wide circuit-breaker component) is responsible for quenching the default cascade. Figure 5.8 (bottom) presents the probability that a market-wide circuit-breaker will trigger for each experimental scenario. We observe that MWCB trigger events are very rare for ω_{mcap}^{low} at $3(\pm 3)\%$ for $\mathcal{R}$, $\mathcal{H}$ and $\mathcal{U}$ and zero for $\mathcal{L}$. Since 64% of trials at ω_{mcap}^{low} exhibit distressed selling behaviour by at least one fund (Figure E.4), of which almost all cases do not result in a default cascade (P_{quench} is close to 100%), we claim that the LULD component of the composite LULD-MWCB policy is typically sufficient to prevent systemic collapse under ω_{mcap}^{low} shock regimes for the real-world 13-F fund-asset network against which our model is calibrated. For the $\mathcal{U}$ policy we can make this claim more concrete by

noting that the observed reduction of $P_{contagion}$ from $10(\pm 10)\%$ to $0(\pm 0)\%$ for ω_{mcap}^{low} (Figures 5.3 vs 5.7 (top)) cannot be fully explained by the action of the market-wide circuit breaker, as this only triggered in $3(\pm 3)\%$ of trials for $\mathcal{U}$ under ω_{mcap}^{low} . Although we claim that LULD is sufficient to prevent systemic collapse under ω_{mcap}^{low} , partly due to the small sample sizes involved (50 Monte Carlo replicates), and partly because it is not possible to exhaustively test alternative approaches, we refrain from making the stronger claim that LULD is *necessary* to prevent systemic collapse at ω_{mcap}^{low} . We are, however, justified in claiming that LULD alone is *not* sufficient to neutralise systemic risk under ω_{mcap}^{high} for behavioural policies $\mathcal{R}$, $\mathcal{H}$, and $\mathcal{U}$ where market-wide circuit breakers are readily triggered, implying a severe shock to the highest liquidity 500 assets in the system (components of the S&P 500 index). In all scenarios, the fraction of asset-trading-time that is lost due to LULD asset halts is negligible, indicating that the LULD mechanism fulfils its purpose as a low-impact brake on excessive market volatility. However, under high shock scenarios, we find that LULD is not sufficient to prevent severe price falls and systemic risk, and market-wide circuit-breakers are necessary to prevent further asset price falls.

5.4.5 Effectiveness of Market-Wide Circuit-Breakers

As discussed in Section 5.3.2, the current US market-wide circuit-breaker rules specify three levels of market decline at which circuit breakers take effect, 7% (Level 1), 13% (Level 2), and 20% (Level 3). In this final part of our LULD-MWCB analysis we consider the appropriateness of the current parametrisation. In order to assess the impact of the different tiers of circuit breaker, we consider the level of systemic risk in terms of the extent of contagion ($\Omega_{contagion}$) that is attained immediately prior to the triggering of each tier of market-wide circuit breaker. Figures 5.9 (top) and (bottom) show the extent of contagion immediately prior to the triggering of Level 2 and Level 3 circuit breakers, respectively. For completeness, pre-Level 1 MWCB extent of contagion is presented in Appendix E, but in this case the results simply show that no funds are expected to have defaulted at the time of a Level 1 circuit breaker, except in the case of ω_{mcap}^{high} under the $\mathcal{L}$ policy, where $0.5(\pm 0.5)\%$ of funds are expected to enter default prior to the Level 1 MWCB, which is well below the $\gamma = 5\%$ level at which a default cascade is declared (we discuss the case of the $\mathcal{L}$ policy further, below). Figure 5.9 (top) demonstrates that, in seven of the eight experimental scenarios, the Level 2 circuit breaker triggers prior to the level of systemic risk at which a default cascade is declared. However, under the $(\omega_{mcap}^{high}, \mathcal{U})$ scenario $3.5(\pm 0.5)\%$ of funds are expected to default prior to the triggering of a Level 2 circuit breaker. Unlike the other metrics

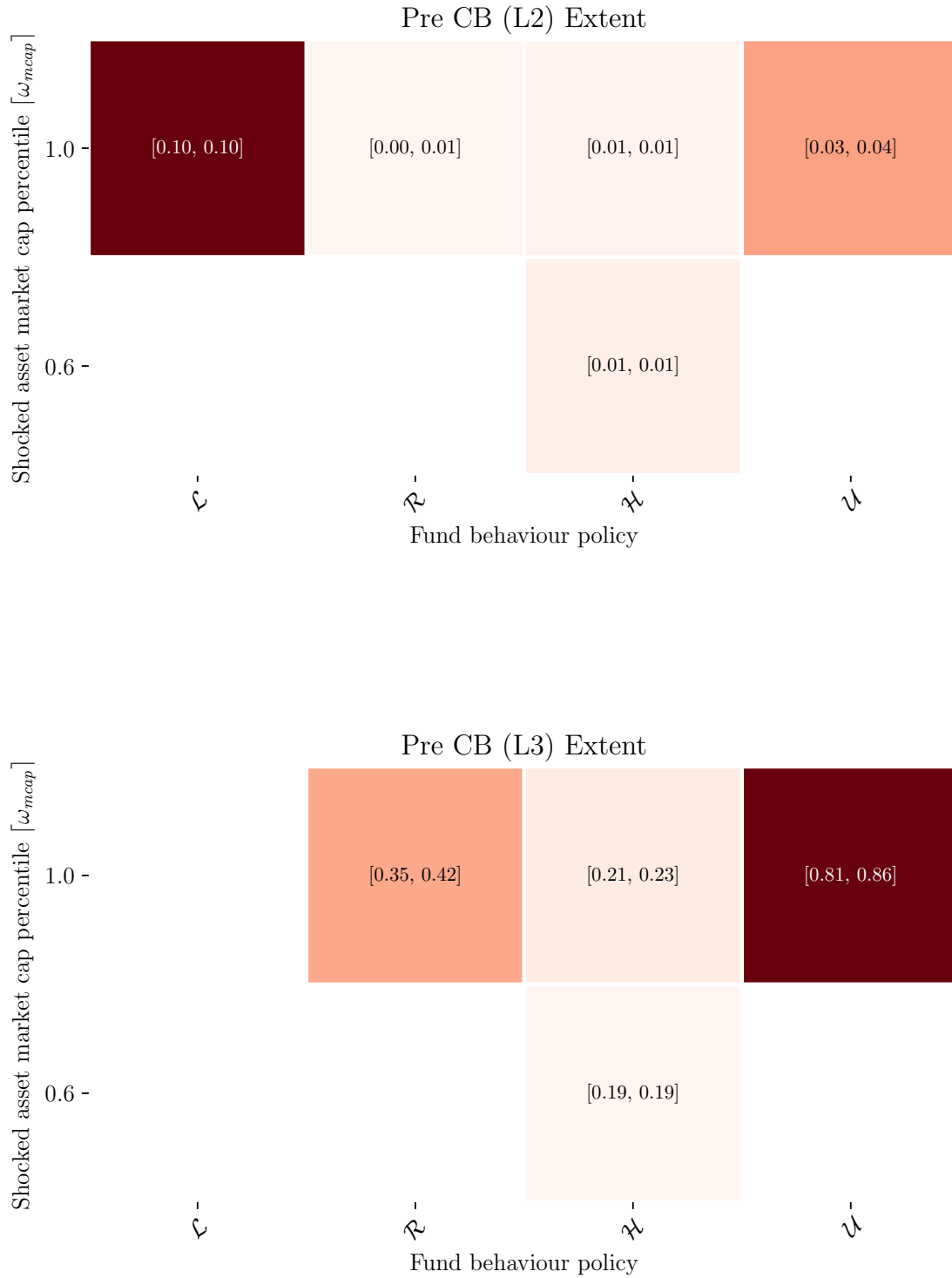


Figure 5.9: Comparing pre-MWCB contagion extent for (top) L2, and (bottom) L3 MWCB activation, for shocked asset centrality quintiles of ω_{mcap} and fund behavioural policies $\mathcal{F}$ at leverage $\lambda^0 = 5$, under the current US LULD-MWCB trading halt policy. Results are based on 318 Monte Carlo replicates ($\omega_{mcap}^{\text{high}}$) and 50 Monte Carlo replicates ($\omega_{mcap}^{\text{low}}$), with cell shading indicating the mean, and labelling indicating the 95% confidence interval.

we have considered in the present chapter, systemic risk is maximised under the $\mathcal{L}$ policy, where, at the time of a Level 2 circuit breaker, $10(\pm 0)\%$ of funds enter default before the market is halted. In order to understand why this is the case, recall that market-wide circuit breakers trigger in response to falls in value of the 500 *highest* market cap assets. Under the $\mathcal{L}$ policy, distressed selling is channelled to the *lowest* market cap assets. Consistent with arguments presented in Section 5.4.3, the set of assets under distress is expected to grow slowly under the $\mathcal{L}$ policy. As assets are individually halted due to LULD price limits, funds will target less preferred assets with increasingly large market caps, and thus distress contagion migrates away from the lower, toward the higher, market cap limit of the asset size spectrum. By the time a substantial portion of the highest market cap assets experience price impact sufficient to trigger the market-wide circuit breaker under $\mathcal{L}$, it is likely that assets at all lower levels of market capitalisation are already in distress, implying that a substantial fraction of funds can be expected to be in distress by such a time. We thus offer further evidence that the effect of fund behaviour cannot be disregarded when considering policy decisions. A consequence of our results is that systemic risk (as measured by $P_{contagion}$) is minimised under the $\mathcal{L}$ policy. However, if funds adopt the $\mathcal{L}$ policy, then the magnitude of index fall required to trigger the Level 2 circuit breaker is inappropriately high, and the Level 2 circuit breaker is not sufficient to prevent a systemic collapse.

Figure 5.9 (bottom) presents the extent of contagion ($\Omega_{contagion}$) prior to the triggering of the Level 3 circuit breaker (which occurs when the index of the 500 largest market cap assets falls by 20%). In all experimental scenarios where the Level 3 circuit breaker was observed, the expected fraction of funds in default before the market was halted was well in advance of the level required to declare a systemic collapse. The lowest value of $\Omega_{contagion}$ observed at the time of a Level 3 circuit breaker, attained for the $(\omega_{mcap}^{low}, \mathcal{H})$ scenario, is $19(\pm 0)\%$. These results suggest that the Level 3 circuit breaker is not effective at extinguishing systemic risk at all, and is not appropriately parametrised for the representative set of fund behaviour, market shock and leverage scenarios we investigate here.

5.4.6 Discussion of Effectiveness of Current Regulation

The results we have presented in this section extend the state of the art of ABM regulatory policy investigation. We offer the first calibrated ABM that connects market microstructure level LOB dynamics due to algorithmic traders, with regulatory interventions designed to prevent market-level systemic risk — price limits and circuit

breakers — at full real-world scale. In so doing, we are able to offer results that connect the trading decisions of investors during crisis scenarios to the contagion of flash crashes and systemic risk throughout the real-world financial system of the United States, arguably the most important financial system in the world.

Our results establish that the current regulatory policy of single asset limit-up-limit-down halts, coupled with a three-tier market-wide circuit breaker (LULD-MWCB, in our notation), is effective at reducing systemic risk compared with the null model featuring no such regulatory interventions (“NONE”). However, our results further establish that fund behaviour makes a critical difference to systemic risk, and does so in ways which are not consistent with results found from simpler non-crisis price impact models such as [Capponi and Larsson \(2015\)](#), [Duarte and Eisenbach \(2018\)](#) and [Cont and Schaanning \(2019\)](#). In particular, we find that a behavioural policy whereby funds preferentially unwind their highest market-cap assets ($\mathcal{H}$) gives rise to a much greater level of systemic risk than alternative policies whereby funds preferentially unwind their lowest market cap assets ($\mathcal{L}$) or a randomised policy ($\mathcal{R}$). These results are consistent with our earlier findings in Section 4.10, and also with empirical research by [Fry and Serbera \(2017\)](#) who find that higher market cap assets are correspondingly more likely to suffer flash crashes compared to lower market cap assets.

Although we do not model the variable collateral value of assets, [Lillo and Pirino \(2015\)](#) and [Cont and Schaanning \(2019\)](#) note that lower liquidity (and hence lower market cap) assets carry higher risk weights due to their typically higher volatility. This feature of real-world margin funding arrangements creates an incentive for market participants to preferentially unwind low market cap assets when facing margin calls, and this raises the encouraging prospect that, in a real-world crisis, distressed funds would select a policy close to $\mathcal{L}$.

We find that both aspects of the current regulation (individual LULD price limits and market-wide circuit breakers) play a role in mitigating systemic risk, and so claim that LULD alone is not sufficient to cover all considered scenarios. Although we find that market-wide circuit breakers are necessary, we also offer evidence that, with their current specification, they fail to defend the financial system against very extensive levels of collapse, under reasonable assumptions of leverage, fund behaviour and asset shock profiles. In particular, the levels of S&P 500 decline needed to trigger the Level 3 circuit breaker are so extreme that the failure of large fractions of the set of funds is extremely likely before the Level 3 circuit breaker is activated. We would caution that these results are expected to depend on the leverage used by funds, and this dependence is not investigated in the current section, though evidence from [Ang](#)

et al. (2011) and quarterly Form PF reports¹⁹ suggests that $\lambda^0 = 5$ is fully consistent with current market practice.

Our model assumes that the value of the S&P 500 is updated every ten simulated seconds. During a real intraday systemic crisis this assumption may even be too generous. As BlackRock explain in their 2015 policy bulletin: (BlackRock, 2015, p.10) “Since the current MWCB thresholds depend on the price of the S&P 500, the MWCB is susceptible to disruptions in the index calculation. If a substantial number of the underlying stocks in the S&P 500 are halted or subject to delayed opens, the index level may be using stale prices. This situation may inadvertently reduce the likelihood that a MWCB will be triggered in stressed markets since the S&P 500 index would not reflect real-time market conditions.” In response, BlackRock propose an alternative policy. Given the problems that the results from our full-scale microstructure calibrated ABM have identified with current regulatory practice, in the following section we exercise the full power of the ABM to quantitatively assess the effect of BlackRock’s alternative proposal on intraday systemic risk and flash crash contagion.

5.5 *In Silico* Policy Evaluation

In their 2015 white paper on US equity market structure, BlackRock (one of the world’s largest asset managers, and a member of the 13-F dataset used in this chapter) proposed a modification to the (still-current) LULD-MWCB regulatory policy (BlackRock, 2015). BlackRock’s twin claim is that (1) the LULD mechanism alone is insufficient to mitigate systemic risk, and (2) the LULD mechanism may affect the timely calculation of the S&P 500 index and thus threaten the appropriate triggering of the market-wide circuit breaker under stressed market conditions. As BlackRock (2015, p.10) explain “a market-wide halt should be triggered when normal market mechanisms are not working, otherwise trading at anomalous prices can destabilize the market. On August 24 [2015], market mechanisms were not working properly. Securities experienced sharp price dislocations and nearly 1,300 LULD halts were triggered as the market lacked transparency and clarity, demonstrating a limitation of the LULD mechanism during market-wide events.” In order to overcome these deficiencies BlackRock (2015, p.10) propose:

¹⁹<https://www.sec.gov/divisions/investment/private-funds-statistics/private-funds-statistics-2018-q1.pdf>, accessed 2019-10-16.

“BlackRock believes policy makers should consider establishing a MWCB that is based on the number of individual securities halted or disrupted. . . . Further analysis by policy makers will be required to determine what percentage of securities would be an appropriate MWCB trigger.”

The results we contribute in the present chapter offer the first quantitative insights into the effectiveness of BlackRock’s proposed policy (which we hereafter refer to as LULD-MWCB-BKRK), and offer clarity into the effect of the policy on systemic risk at different threshold trigger levels, in exactly the sense proposed by BlackRock in the above quotation.

5.5.1 BlackRock Policy Modification and Experimental Scenarios

The modification proposed by BlackRock introduces a new connection between the LULD and MWCB components of the existing LULD-MWCB policy. Although the specification presented in [BlackRock \(2015\)](#) is slightly ambiguous, in the present chapter we interpret the BlackRock’s “percentage of securities” to refer to assets that are members of the S&P 500 index (i.e., the set of 500 largest assets). We thus define a BlackRock trigger threshold fraction $b \in (1/500, 1]$. As indicated with (*) in [Figure 5.6](#), under LULD-MWCB-BKRK, circuit breakers of any level (1, 2, or 3) are triggered if the fraction of S&P 500 assets halted due to LULD price limits exceeds $b\%$. b is held constant across Level 1, 2, and 3 circuit breaker trigger specifications in the current chapter for reasons of computational cost and to avoid overcomplicating our analysis.

In the following section we investigate systemic risk under 20 experimental scenarios $s \in \mathcal{F} \times \{b_1, b_2, b_5, b_{10}, b_{20}\}$, where b_p indicates the BlackRock trigger threshold of p percent (i.e., $b_{20} \Rightarrow b = 0.2$). In order to obtain a reasonable number of Monte Carlo replications within our computational budget, it was necessary to hold fixed fund leverage $\lambda^0 = 5$, as well as the shocked asset market cap quintile $\omega_{\text{mcap}} = \omega_{\text{mcap}}^{\text{high}}$. We perform 150 Monte Carlo replicates for each experimental scenario, and present the mean with 90% and 95% significance indicators in the results that follow.

As with the previous experiment, we sample the shocked asset uniformly from the distribution of asset degrees $\omega_{\text{degree}} \sim \text{Uniform}(0,1)$, and perform robust $\mathcal{X}^2$ tests (Mood’s median test) to control for the effect of shocked asset degree (as introduced in [Section 5.4.2](#)). In the present experiment we compare the distribution of shocked

asset degrees for a given experimental scenario to the distribution under the LULD-MWCB (non-BlackRock) policy for the matching fund behaviour policy. We observe no significant differences at $p < 0.05$. The χ^2 test results are presented in Figure E.6.

5.5.2 Effect of Percentage Threshold on BlackRock’s Policy

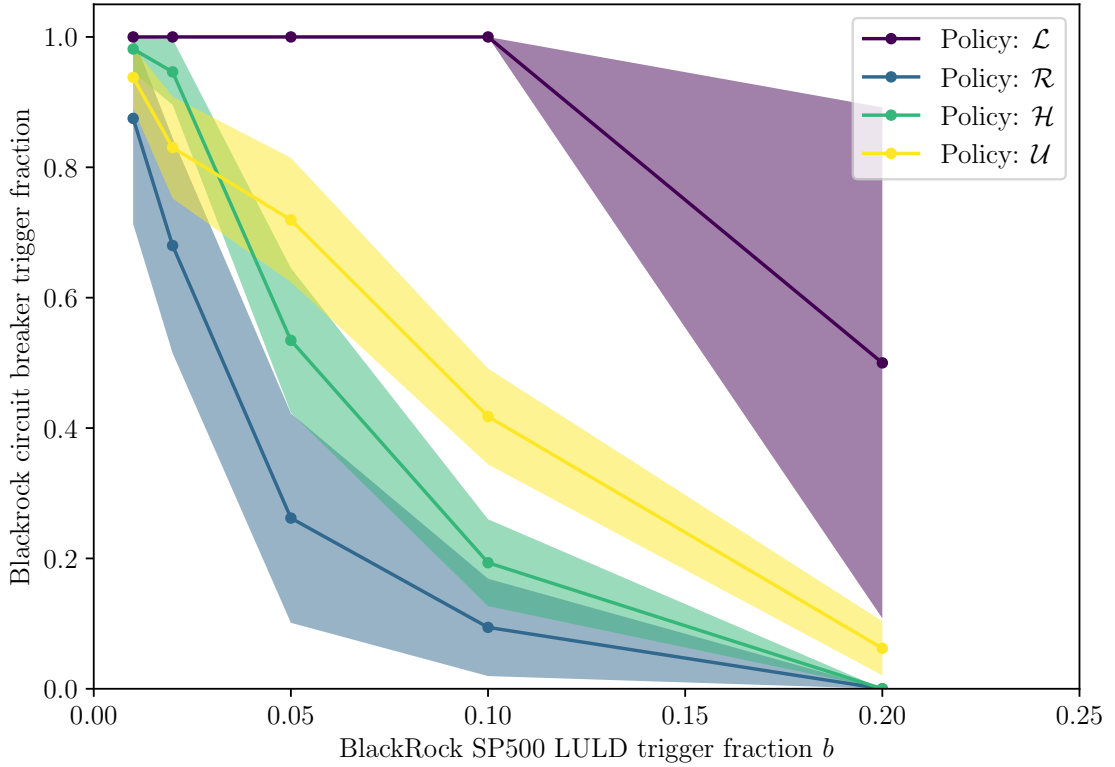


Figure 5.10: The fraction of MWCB events that triggered as a result of BlackRock’s policy amendment, as a function of the trigger threshold b . We present the mean and 95% confidence interval from 150 Monte Carlo trials for fund behavioural policies $\mathcal{F}$.

Figure 5.10 presents the fraction of simulated market-wide circuit breaker events where the MWCB triggered due to BlackRock’s modification, for each experimental scenario. We refer to this metric as the “BlackRock circuit-breaker trigger fraction”. A value of 1.0 indicates that all MWCB events observed for the given experimental scenario²⁰ were triggered as a result of more than $n_a b$ assets concurrently entering LULD halts. A value of 0.0 indicates that MWCB activation was due to the current regulatory standard price fall trigger. By construction, $b \rightarrow \infty \Rightarrow \text{LULD-MWCB-BKRC} \rightarrow \text{LULD-MWCB}$. We therefore expect to see the BlackRock circuit-breaker

²⁰For a given Monte Carlo trial it is possible to observe zero, one, two, or three MWCB events corresponding to the different MWCB levels.

trigger fraction tend towards zero with increasing b . For $b \leq 1/500$ the MWCB would trigger when a single S&P 500 asset enters LULD halt, and so we further expect the BlackRock circuit-breaker trigger fraction to tend towards 1.0 as $b \rightarrow 0$. Both of these expected features are apparent in Figure 5.10. For fund behavioural policies $\mathcal{U}$, $\mathcal{H}$, and $\mathcal{R}$ we observe that when $b \geq 0.2$ there is no difference between LULD-MWCB-BKRK and LULD-MWCB. This allows us to estimate an upper bound on the effective range of BlackRock's threshold percentage of 20% for a wide range of fund behavioural policies. The effective range of b is larger for the $\mathcal{L}$ policy. The reason for this relates to the slow growth of the set of distressed assets under the $\mathcal{L}$ policy, described in Section 5.4.3. Under the $\mathcal{L}$ policy, assets with higher market cap are likely to experience distress *after* assets with lower market caps. This reasoning implies that the lowest-market cap members of the S&P 500 index will experience distress (and hence LULD halts) before higher market cap S&P 500 assets. However, since the S&P 500 is a market-cap weighted index, price movements among lower market cap assets have less effect on the index value than price movements among higher market cap assets. It is thus likely that a large number of the lowest market-cap assets in the S&P 500 will experience LULD halts before the overall value of the index falls sufficiently to trigger the MWCB. This explanation is consistent with the observation in Figure 5.10 that under $\mathcal{L}$ the BlackRock circuit-breaker trigger fraction is higher than under other fund behavioural scenarios for a given value of b .

These results again indicate the difficulty faced by policy-makers when selecting appropriate parametrisations for regulatory interventions²¹. We find that the crisis behaviour of market participants, represented by $\mathcal{U}$, $\mathcal{H}$, $\mathcal{R}$, and $\mathcal{L}$ in our model, has a significant influence on the activation of BlackRock's policy, and so we would argue that any consideration of the adoption of LULD-MWCB-BKRK by policy-makers must consider likely behavioural scenarios as a key aspect of any cost-benefit analysis.

5.5.3 Effect of BlackRock's Policy on Systemic Risk

Figure 5.11 shows the probability of contagion ($P_{contagion}$, top) and extent of contagion ($\Omega_{contagion}$, bottom) in each experimental scenario $s \in \mathcal{F} \times \{b_1, b_2, b_5, b_{10}, b_{20}\}$. Solid lines represent the mean from 150 Monte Carlo trials under LULD-MWCB-BKRK, and horizontal dashed lines represent the mean from 318 Monte Carlo trials under LULD-MWCB (the experimental control). In order to enhance readability, statistical

²¹BlackRock refer to selection of policy parameters as *calibration*, but we refrain from using the term in this context in order to avoid introducing possible confusion with our discussion of calibration and validation in Chapter 4.

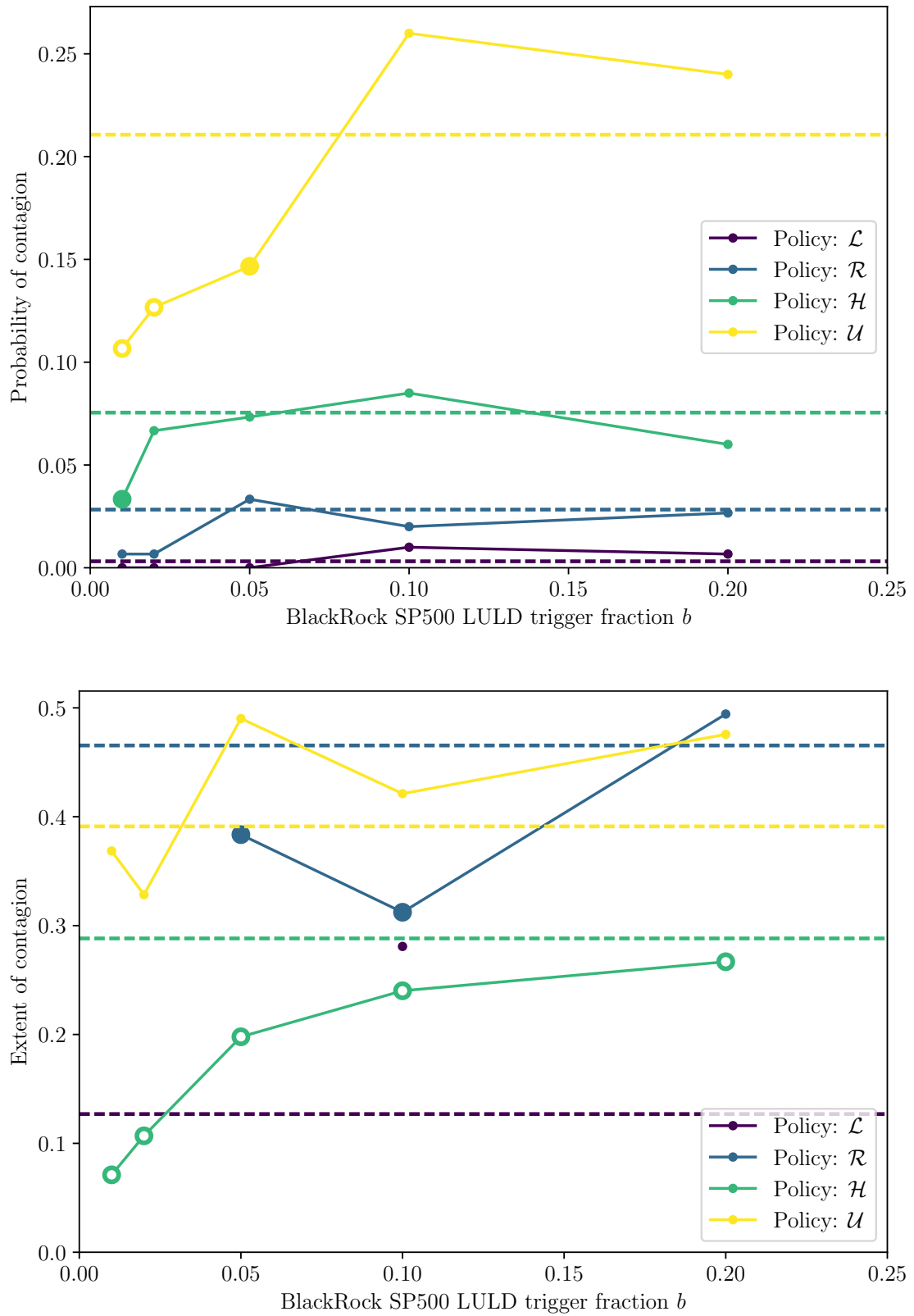


Figure 5.11: Probability (top) and extent (bottom) of contagion as a function of BlackRock threshold parameter b (solid lines) and under the real-world LULD-MWCB policy (dashed lines), for fund behavioural policies $\mathcal{F}$. Statistically significant differences between the LULD-MWCB-BKRK and LULD-MWCB policies from 150 Monte Carlo replicates at $p < 0.05$ are indicated with large hollow circles, and at $p < 0.1$ with large filled circles.

significance is indicated using glyphs. A large filled circle represents a LULD-MWCB-BKRK result that is significantly different from the LULD-MWCB control at $p < 0.1$, and a large circle with an empty centre represents significance at $p < 0.05$.

Figure 5.11 (top) shows that BlackRock’s policy has no significant effect on the probability of contagion for $b \geq 0.1$. However, as b is reduced (and the threshold for MWCB activation is lowered), BlackRock’s policy reduces the probability of contagion for all fund behavioural policies under consideration. This effect is statistically significant for $\mathcal{U}$ and $\mathcal{H}$, and the trend is also apparent for $\mathcal{R}$ and $\mathcal{L}$, albeit with limited statistical power in these cases. These results are consistent with the MWCB activation dependence on b observed in Figure 5.10 and the notion that systemic risk is reduced if the market is halted for smaller price falls. Indeed, it is trivially obvious that systemic risk can be reduced to zero by simply forbidding trading ($b \rightarrow 0$), however this solution would of course carry an unacceptable cost as the market’s function as an orderly price discovery mechanism would be nullified.

The limited effect that LULD-MWCB-BKRK has on $P_{contagion}$ can in part be explained by recalling that $P_{contagion}$ is defined as the fraction of Monte Carlo trials where more than $\gamma = 5\%$ of funds default (Section 3.4). If the effect of LULD-MWCB-BKRK is to reduce the fraction of defaulting funds, this would only impact $P_{contagion}$ if the resulting default fraction were to be below the γ threshold. A decrease in the (unconditional)²² extent of contagion from, for example, 20% to 10%, while of significant benefit to the financial system, would not result in a change to $P_{contagion}$. It is thus instructive to also consider the extent of contagion in order to characterise the effect of LULD-MWCB-BKRK on systemic risk.

Figure 5.11 (bottom) presents the extent of contagion ($\Omega_{contagion}$) for each experimental scenario, and here the effect of LULD-MWCB-BKRK is readily apparent. It is indeed the case that LULD-MWCB-BKRK reduces the extent of contagion as b is reduced. This trend is apparent for all fund behavioural policies, and is particularly significant for $\mathcal{H}$ and $\mathcal{R}$. Figure 5.7 (bottom) shows that the 95% confidence interval for $\Omega_{contagion}$ under the $\mathcal{U}$ policy is wide (relative to the range of $\Omega_{contagion}$ as a function of b in Figure 5.11 (bottom)), at 0.3–0.48, resulting in limited statistical significance of the effect of b under the $\mathcal{U}$ policy in the present section. Further Monte Carlo samples would be beneficial to clarify these results. In particular, we would not expect the extent of contagion to be worse under LULD-MWCB-BKRK than under

²²The unconditional extent of contagion is the fraction of funds in default for a given Monte Carlo replicate. It is not to be confused with $\Omega_{contagion}$ which is the *conditional* extent of contagion, i.e., the extent of contagion conditional on the occurrence of a default cascade. See Section 3.4 for details.

LULD-MWCB because the MWCB triggers strictly earlier under LULD-MWCB-BKRK (before assets have had time to fall so far in price as under LULD-MWCB).

Figures 5.11 (top) and (bottom) also indicate that LULD-MWCB-BKRK does not induce changes in the rank ordering of systemic risk for the fund behavioural policies, either in the case of $P_{contagion}$ or $\Omega_{contagion}$, for any value of b under test. This feature, whilst not surprising, at least serves to avoid further complicating the regulatory selection of an appropriate value of b .

Taken together, Figures 5.10 and 5.11 demonstrate that BlackRock's policy can potentially have a significantly positive impact on systemic risk as measured by the probability and extent of financial contagion²³. In the following section we look in more detail at how the LULD-MWCB-BKRK policy achieves this objective.

5.5.4 Effectiveness of MWCBs under BlackRock's Policy

As a sanity check, we measured the probability that at least one fund will enter the margin call state under the LULD-MWCB-BKRK policy. We found that the probability is no different under LULD-MWCB-BKRK compared to LULD-MWCB, and also that there is no variation as a function of fund behavioural policy or BlackRock threshold b . These results are consistent with our observations in Section 5.4.4 (with further details in Appendix E) that the probability that at least one fund enters the margin call state is, by construction, independent of fund unwind behaviour and MWCB activation (because these events can only occur *after* at least one fund has entered the distressed state). By extension, we would not expect the BlackRock threshold b to affect the probability that at least one fund enters the distressed state, and indeed our results are consistent with this hypothesis.

Given BlackRock's policy amendment cannot (and in our simulation, does not) affect the probability that at least one fund commences the process of distressed deleveraging, how exactly does BlackRock's amendment bring about a reduction in systemic risk?

In order to address this question, in Figure 5.12 we present the (unconditional) extent of contagion immediately prior to the activation of the Level 2 and Level 3 circuit breakers under the LULD-MWCB-BKRK policy. At the time of Level 1 circuit breaker activation, the extent of contagion is not found to be significantly different to

²³We also measured systemic risk in terms of a number of alternative metrics. The conclusions are the same in each case and so for the sake of brevity we do not include these detailed results here. The metrics included (1) the probability of occurrence of flash crashes, (2) the minimum observed value of the S&P 500 index, and (3) the speed of flash crash propagation. These metrics were found to be highly correlated with the metrics presented in the main text.

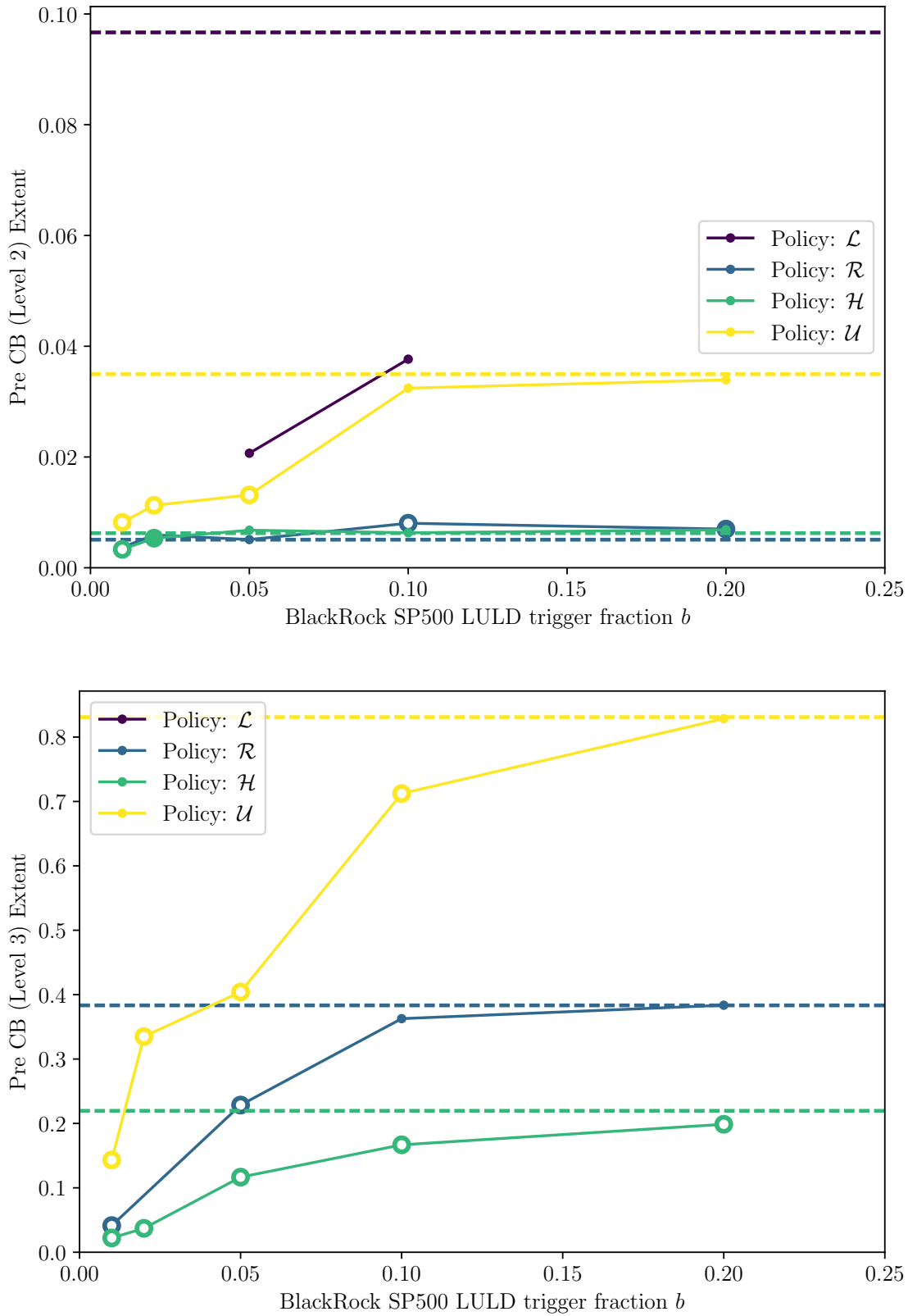


Figure 5.12: Pre Level 2 circuit breaker (top) and Level 3 circuit breaker (bottom) extent of contagion as a function of BlackRock threshold parameter b (solid lines) and under the real-world LULD-MWCB policy (dashed lines), for fund behavioural policies $\mathcal{F}$. Statistically significant differences between the LULD-MWCB-BKRK and LULD-MWCB policies from 150 Monte Carlo replicates at $p < 0.05$ are indicated with large hollow circles, and at $p < 0.1$ with large filled circles.

the control LULD-MWCB policy for any value of b or under any fund behavioural policy. As with the LULD-MWCB policy, the fraction of funds expected to have defaulted prior to the activation of the Level 1 circuit breaker is close to zero for all fund behaviour policies. However, Figure 5.12 (top) shows that at the time of Level 2 circuit breaker activation the LULD-MWCB-BKRK policy induces a substantial reduction in the extent of contagion for all behavioural policies when $b \leq 0.05$. Although the results presented in Figure 5.12 (top) could benefit from further Monte Carlo samples in order to improve statistical significance and reduce anomalous outlier observations (e.g., the $\mathcal{R}$ policy for $b \geq 0.1$), the trend towards reduced systemic risk for lower b is robustly observed for all fund behavioural policies. LULD-MWCB-BKRK has the most pronounced effect on extent of contagion at the time of the Level 2 circuit breaker under the $\mathcal{U}$ and $\mathcal{L}$ behavioural policies. In particular, under the current US regulatory standard LULD-MWCB, the $\mathcal{L}$ fund behavioural policy gives rise to an extent of contagion of $10(\pm 0)\%$, a value sufficient to declare a default cascade event. Under LULD-MWCB-BKRK, however, the extent of contagion under $\mathcal{L}$ is reduced by a factor of almost 80% for $b \leq 0.1$. This finding is consistent with our earlier observation (Figure 5.10) that the $\mathcal{L}$ behavioural policy is especially likely to result in the triggering of a MWCB due to BlackRock's modification, as opposed to triggering due to a large fall in the value of the S&P 500 index under the LULD-MWCB policy. Figure 5.12 (top) also demonstrates that LULD-MWCB-BKRK has a significant positive effect on extent of contagion under $\mathcal{U}$. Under LULD-MWCB the $\mathcal{U}$ policy resulted in an extent of contagion close to the level at which a default cascade is declared, whereas under LULD-MWCB-BKRK it is possible to reduce the $\mathcal{U}$ extent of contagion by a factor of almost 75% for $b \leq 0.05$. In related results (which for the sake of brevity are not presented here), we observed that the probability that a MWCB will be triggered under LULD-MWCB is greatest for the $\mathcal{U}$ and $\mathcal{H}$ policies. These results are consistent with expectations because MWCB activation is conditional on distress to high market-cap assets that are members of the S&P 500 index. The $\mathcal{U}$ and $\mathcal{H}$ policies cause funds to preferentially ($\mathcal{H}$) and unconditionally ($\mathcal{U}$) target high market-cap assets with distressed sell orders, leading to increased probability that the S&P 500 index itself will fall in value, thus triggering the MWCB under LULD-MWCB. However, the preferential targeting of high market-cap assets under $\mathcal{U}$ and $\mathcal{H}$ also implies that high market-cap assets will be more likely to experience LULD halts than under other fund behavioural policies, and so it is to be expected that MWCB activation under $\mathcal{U}$ and $\mathcal{H}$ occurs more readily under the LULD-MWCB-BKRK policy as the activation threshold b is reduced.

These observations are even more striking when we consider the Level 3 circuit breaker. Figure 5.12 (bottom) presents the extent of contagion at the time of the Level 3 circuit breaker, and here we see a high level of statistical significance for the $\mathcal{U}$, $\mathcal{H}$ and $\mathcal{R}$ fund behavioural policies²⁴. At the b_5 threshold, we see reductions in the pre-Level 3 circuit breaker extent of contagion of 50% ($\mathcal{U}$), 52% ($\mathcal{H}$), and 42% ($\mathcal{R}$). Our results thus suggest that, contrary to our observations of the effectiveness of the Level 3 circuit breaker under the current LULD-MWCB policy, under the LULD-MWCB-BKRK policy with threshold $b = 0.05$, the Level 3 circuit breaker can play a highly-relevant and high-impact role in the mitigation of systemic risk and the protection of large fractions of the fund–asset network from catastrophic failure.

The effectiveness of LULD-MWCB-BKRK is further supported through *false positive* testing. In this analysis we identified trials under the baseline LULD-MWCB for which the BlackRock MWCB would have activated (but did not, because $b = \infty$, effectively). We classified the inferred BlackRock activations as *true positive* in cases where a default cascade was subsequently observed (regardless of whether or not the MWCB triggered in the trial), and *false positive* in cases where no default cascade was observed, and the MWCB did not trigger. A false positive BlackRock activation would thus serve to delay the return of the market to a non-crisis quiescent state. We observed no false positive instances for any behavioural scenario, under $\omega_{\text{mcap}}^{\text{low}}$ or $\omega_{\text{mcap}}^{\text{high}}$.

The results we have presented in this section suggest that BlackRock’s proposed policy, that of triggering market-wide circuit breakers in response to a fraction of assets experiencing individual trading halts (BlackRock, 2015), has the potential to make substantial improvements to the resilience of the financial system in the event of a systemic deleveraging crisis over intraday timescales. However, these findings must to be considered in the wider context of the implications of closing the markets under conditions of financial distress that are less severe than would otherwise be necessary under current regulations. We turn to a discussion of these and other issues in the following section.

5.5.5 Discussion of BlackRock’s Policy Proposal

A key strength of our model is the realistically calibrated micro-level algorithmic trading simulation, however, by embedding the microstructure model in a calibrated

²⁴Fund behaviour policy $\mathcal{L}$ did not result in levels of financial distress sufficient to trigger the Level 3 MWCB in our simulations. This observation is as expected and is an important confirmation and sanity check of the model behaviour.

macro-level fund–asset network model we offer a much more comprehensive *in silico* policy testing environment than other microstructure models that operate on one or a small number of assets (e.g., [Kobayashi and Hashimoto \(2011\)](#); [Jacob Leal and Napoletano \(2019\)](#)). By construction, market-wide circuit breaker triggering (real, and simulated) is a function of the intraday price movements of multiple interacting assets. By connecting the detailed price formation activity of multiple assets with systemic properties such as the S&P 500 level, our model offers a unique regulatory window into the connection between macro-level systemic risk and the macro- and micro-level policies designed to mitigate it. Consequently, our results represent the first time an alternative market-wide circuit breaker policy has been subjected to quantitative analysis in a calibrated full real-world-scale micro-macro ABM. Indeed, to the best of our knowledge we have performed the first ever quantitative assessment of BlackRock’s proposal, via any methodology. As such, our results will be of significant interest to regulators as they attempt to assess the utility of BlackRock’s suggestion.

Our results indicate that BlackRock’s proposal has a positive effect on systemic risk, measured by the probability and extent of contagion. Furthermore, this result is consistent across a wide range of fund behavioural policies, and is typically monotonic in the BlackRock parameter b . We find that the policy reduces to the current regulatory LULD-MWCB for $b \geq 20\%$. By considering the extent of contagion at the time of a Level 3 circuit breaker event, our results also offer evidence that BlackRock’s policy can substantially improve the value and relevance of the Level 3 circuit breaker. Without BlackRock’s proposal this circuit breaker was found to be ineffective, and “too little too late”. In contrast, under BlackRock’s policy the Level 3 circuit breaker could potentially reduce the extent of contagion (the fraction of funds entering default and bankruptcy) by a factor of 50%.

However, our results also indicate that systemic risk continues to vary widely under alternative fund behavioural policies with BlackRock’s modification as with the current LULD-MWCB regulatory policy. Regulators would be well advised to account for alternative fund behaviours in any subsequent assessment of BlackRock’s policy, or future competing alternatives. This finding again highlights the need to collect adequate data, both in terms of the regulator’s ability to characterise market participants and calibrate the ABM, but also in terms of performing adequate Monte Carlo replications of the model itself. Computational costs are highly material in a simulation of this scale and level of detail.

However, there is another side to BlackRock's policy that is difficult to quantify. If the market is closed, then it is not performing its function as an orderly price discovery mechanism, and this failure carries a social cost. Sifat and Mohamad (2019, p.1163) state "the closing price of a security is a beacon for the economy. Intervention that distorts or impedes the process to discovering a rational price for a security imposes economic costs on the society". Reducing the threshold at which MWCB events trigger by connecting activation to individual LULD halts may introduce a range of unintended consequences:

1. The *magnet effect* (Sifat and Mohamad, 2019; Hautsch and Horvath, 2019), whereby traders anticipate a trading halt when prices are close to limits, can make it more likely that a limit will indeed be breached. The effect is due to traders wishing to avoid being unable to sell out of positive positions during a halt, and so sell pre-emptively during a price decline, driving prices down further and making a limit breach even more likely. In the case of BlackRock's proposal, if a fraction $b - \epsilon$ of S&P 500 assets are halted, traders could be forgiven for predicting that MWCB activation would likely follow. Assessing this behaviour in an ABM, while fully feasible, would require moving beyond the zero-intelligence paradigm we have adopted in this thesis.
2. A too-low value of b could expose the market to a vulnerability whereby a manipulative attacker could seek to force MWCB activation by targeting specific assets with overwhelming sell orders. As our results demonstrate, high market cap assets could be particularly vulnerable to overwhelming directional liquidity, and such assets are likely those relevant to the BlackRock mechanism.
3. Our results indicate that we do not expect to observe a default cascade by the time of a Level 1 circuit breaker activation. However, such an activation could act as a warning signal to funds of increased future volatility. This could induce funds to take defensive action and reduce risk exposures, which would involve selling into a falling market, thus increasing the probability and extent of systemic collapse.
4. Finally, and perhaps most importantly, closing the market early eliminates the opportunity for markets to rebound in false-positive scenarios (such as during a flash crash, where volatility is due to algorithmic, rather than fundamental, trading (Menkveld and Yueshen, 2019)). Unable to take action in domestic positions, investors would be faced with no choice but to hedge exposures using

international markets. For example, traders wishing to offset positive US market exposure may set up short positions in Asian index futures markets (such as the TOPIX in Japan, or Hang Seng in Hong Kong). In this way, the distress would cross geopolitical boundaries and become a truly global crisis. This possibility was introduced by [Cliff and Northrop \(2012, p.7\)](#) who state “we think that the true nightmare scenario would have been if the [Flash] crash’s 600 point down-spike, the trillion-dollar write-off, had occurred immediately before market close: that is, if the markets had closed just after the steep drop, before the equally fast recovery had a chance to start. Faced with New York showing its biggest ever one-day drop in the final 15 minutes before close of business on May 6th ... traders in Tokyo would have had only one rational reaction: sell. The likelihood is that Tokyo would have seen one of its biggest ever one-day losses. Following this, as the mainland European bourses and the London markets opened on the morning of May 7th, seeing the unprecedented sell-offs that had afflicted first New York and then Tokyo, European markets would have followed into precipitous freefall.” We discuss the necessary extensions to our model to enable quantitative investigation of this scenario in Chapter 6.

5.6 Conclusions

In this chapter we have demonstrated methodology and results that for the first time connect intraday systemic risk and flash crash contagion with the regulatory policies intended to protect major financial markets against panic selling and excess volatility, in the context of a calibrated real-world-scale agent-based model. Circuit breakers and price limits are *only* defined over intraday horizons ([Sifat and Mohamad, 2019](#)), and so in order to connect these regulatory policies to systemic risk it is *necessary* to model systemic risk over intraday timescales. Our model represents the first study to look at real-world-scale circuit breaker and price limit simulation, in the context of systemic risk, or otherwise. We advance the state of the art from two assets ([Torii et al., 2015](#)) to over 6600, and calibrate the model using the empirical portfolios of 4500 asset managers (with 2018 data). We offer the following contributions:

1. We find that systemic risk increases monotonically with leverage, and with the market cap of the asset receiving the initial shock, regardless of fund behaviour. This result confirms our earlier observations in a smaller scale model with synthetic fund-asset network topology ([Paulin et al., 2019](#)).

2. Other things equal, systemic risk is maximised when a shock is delivered to an asset with intermediate degree (i.e., an intermediate number of owners). Shocks to assets with higher degree are mitigated by diversification effects, and shocks to assets with lower degree are mitigated by connectivity effects. This suggests regulators should pay particular attention to large market cap assets with relatively concentrated investor bases.
3. Higher market cap assets are more prone to experiencing flash crashes (a result that is consistent with empirical findings by [Fry and Serbera \(2017\)](#)).
4. Systemic risk (probability and extent of contagion) varies significantly with alternative fund behavioural policies. We find that, contrary to macro-level systemic risk literature (e.g., [Capponi and Larsson, 2015](#)), systemic risk is maximised when funds preferentially sell *high* market cap assets. The importance of considering asset microstructure in systemic risk models is thus established.
5. The current LULD-MWCB regulatory policy is not always sufficient to prevent systemic collapse in all distress scenarios considered. The Level 3 circuit breaker, in particular, triggers too late to be of any practical use in a crisis, as by then a large fraction of funds will already have failed.
6. The optimal fund behavioural policy from a systemic risk perspective, $\mathcal{L}$, is particularly poorly served by the MWCB mechanism, as by the time the S&P 500 moves significantly, under $\mathcal{L}$, it is likely that systemic collapse is unavoidable.
7. We subject MWCB proposals by [BlackRock \(2015\)](#) to quantitative comparison with existing policy, finding that BlackRock's proposals have a beneficial effect on systemic risk, tempered by a set of hard-to-quantify costs.
8. We find evidence that BlackRock's policy can potentially improve the effectiveness of the Level 3 circuit breaker by a factor of approximately 50%, including under the $\mathcal{L}$ fund behavioural policy.
9. We offer regulators a new tool for the quantitative evaluation of future alternative intraday systemic risk policy interventions, and demonstrate comprehensive methodology to facilitate calibrated and validated simulation at full real-world scale.

We have shown how it is possible take a quantitatively calibrated whole-of-market ABM, with heterogeneous assets and funds, and use it to not only test the effectiveness of current regulatory policies, but also to subject novel policy proposals to *in silico* evaluation. Regulators have expressed a strong desire for tools of this nature ([Haldane, 2016](#); [Haldane and Turrell, 2018](#)), and have taken steps to construct their own

approaches (Braun-Munzinger et al., 2018; Collver, 2017; Kårvik et al., 2018). With the increased availability of “Big Data” for calibration, and computational power for Monte Carlo replications and large sets of simulated scenarios, ABM stands poised to deliver a potent and flexible tool into the hands of regulatory authorities. Our results contribute to this rapidly growing enterprise.

As international markets continue to fragment into multitudinous trading venues, it is more important than ever for regulators to have access to techniques that are capable of handling the increasing complexity of the modern financial system. However, it is only possible to unlock the potential of ABM if regulators can be assured of access to sufficiently detailed data on market participants. In this chapter we have been forced to accommodate two key restrictions on available data, and regulators will face similar considerations:

1. *Fund-level leverage arrangements including fund capital, current leverage, and margin account tolerance arrangements:* With the recent introduction of Form PF, which requires detailed disclosure of leverage and capital arrangements by funds, it would seem that regulators are aware of this blind spot, and are taking an active approach to resolve it. It is not clear whether these data will be made available to academia, though the SEC do produce quarterly summary reports which may be of some use, if combined with network reconstruction techniques.
2. *Fund distress behaviour:* Without access to an unprecedented level of detail on the likely crisis behaviour of fund managers (which may be impossible for regulators to collect with any level of reliability), it is necessary to construct alternative scenarios to investigate the effect of fund crisis behaviour. Overcoming this limitation would elevate the practice of regulatory ABM to the third and final level in the taxonomy of Axtell and Epstein (1994), though it is very difficult to envisage how this undertaking could be achieved in the current regulatory climate.

In addition to these considerations, it is also important to consider a possible objection to counterfactual policy experiments, known as the Lucas critique (Lucas, 1976). Briefly, Lucas (1976) argues that agent behaviour will not necessarily continue to exhibit historically observed patterns following a policy intervention that is optimised according to those historical patterns. The Lucas critique claims that the strategic response of agents to the new policy could invalidate the causal basis for the policy, potentially undermining its effectiveness. Agent-based models are not automatically immune to the Lucas critique (Turrell, 2016), however, the zero-intelligence approach

that we take in the present chapter offers a degree of protection. Because circuit breaker activation is such a rare tail-event, we join [Jacob Leal and Napoletano \(2019, p.27\)](#) in arguing that “our model structure does not assume any specific information disclosure about the type of circuit breaker that is implemented in the market, which allows us to discard the strategic reaction of traders and to only focus on the effect of discretionary circuit breakers, when agents hardly trade in anticipation of a market halt”. By moving beyond the zero-intelligence paradigm and directly incorporating strategic agent behaviour, it would be possible to offer a more comprehensive response to the Lucas critique. Our framework could be readily extended with such strategic features, and we discuss this potential future work direction in Chapter 6.

The results we have presented in this chapter support our argument that ABM offers a powerful quantitative tool for regulators, to assist with their policy decisions. As discussed in Chapter 2, the immense flexibility of ABM is arguably both the discipline’s greatest strength, and its greatest weakness — the ability to derive useful inferences from any model rests on the strength of the model’s assumptions. The full-scale ABM we have presented in this chapter could readily be extended to include further relevant phenomena, and we consider key possibilities in the next chapter. However, we would argue that even in its present form, our model, methodology, and results, offer a substantial contribution to knowledge, and go a long way toward addressing deep questions about the interaction between circuit breakers, price limits, systemic risk, and algorithmic trading. It is perhaps the case that the work we have presented here begins to provide resolution to long-standing challenges in the field of regulatory policy evaluation ([Kobayashi and Hashimoto, 2011](#)), and begins to meet the pressing need expressed by [Gould et al. \(2013, p.1733\)](#) “From a regulatory standpoint, it is crucial to understand how electronic trading algorithms affect market stability, yet almost nothing concrete is currently known.”

In the next chapter we offer our conclusions, and suggestions for future work.

Computational costs of this thesis (estimated):

- *Results that we present: 44 CPU-years*
- *All results: 87 CPU-years*
- *Energy use at 5W per core: 4000 kWh*
- *Carbon footprint: 2tCO₂e (offset on 2019-10-09 via www.carbonfootprint.com, #17467)*

6

Conclusions

IN THIS THESIS, we first reviewed and assessed state-of-the-art macro-level systemic risk models based on complex network theory, and micro-level flash crash and price impact models based on market microstructure theory. We identified that these approaches need to be integrated in order to provide a more powerful and realistic model for the simulation of systemic risk over intraday timescales. The central contribution of this thesis is to demonstrate how these approaches can be combined to provide a quantitatively validated platform for understanding the impact of algorithmic trading, and regulatory policy, on systemic risk.

6.1 Summary

In Chapter 3, we proposed and assessed a hybrid micro-macro model consisting of leveraged funds, connected via shared asset holdings. Unlike pure macro-level models, price impact arises endogenously in our model through the emergent interactions of algorithmic trading agents. We showed, for the first time, how extreme price impact events (flash crashes) propagate through the fund-asset network, and offered compelling evidence that leverage management practice by funds facilitates this contagion channel for financial distress. Our results thus bridge the gap between the micro- and macro- literatures, showing (1) how micro-models can make use of complex network theory to extend the range of phenomena they can represent, and (2) how macro-level

models can substantially enhance their realism by replacing non-crisis price impact functions with microstructure components.

The calibration and validation methodology we introduced, deployed on the model, and assessed in Chapter 4, contributes to advancing ABM as-a-discipline by providing a novel example of a model validated at Level 2 in the taxonomy of [Axtell and Epstein \(1994\)](#). Relatively few models have been reported that meet this level of *quantitative* validation, and our work is, to the best of our knowledge, the first demonstration of whole-of-market microstructure calibration and validation. In order to achieve this, we made use of extensive financial data, and defined new methods to manage high-frequency market data at a scope and scale not previously demonstrated in the ABM literature. We successfully calibrated and validated the model across the full market-cap spectrum for the first quarter of each year from 2009–2018 inclusive, finding that genetic algorithms outperformed other methods we considered and assessed for this task, e.g., Nelder-Mead simplex search ([Nelder and Mead, 1965](#)) and Markov Chain Monte Carlo ([Lux, 2018](#)). Our discussion contributes to the drive for consistent terminology in the ABM discipline. We also contribute to the adoption of consistent calibration and validation approaches in the field by advocating for a standard methodological work-flow, incorporating sensitivity, and initial transient, analyses. We hope our results will serve as a roadmap for future academic and regulatory practitioners.

In Chapter 5, the final part of this thesis, we took the parameters established through calibration and demonstrated, for the first time, that it is possible to instantiate a full real-world-scale micro-macro model of the US equity financial system. Our model incorporated regulatory 13-F ownership data in order to represent the empirical fund–asset network, leading us to thus establish a new simulation fidelity benchmark, with a calibrated and validated multi-asset microstructure simulation featuring 6619 concurrent assets (each simulated at 50 millisecond frequency). The validated model facilitated the first ever investigation of the impact of intraday price limits and market-wide circuit breakers on systemic risk. We thus extended the state of the art of circuit breaker simulation into the multi-asset, systemic domain. Our results identified limitations of current policies (Limit-Up-Limit-Down rule, and market-wide circuit breakers) at preventing systemic risk and flash crash contagion. We identified circumstances where existing policy is insufficient (e.g., when funds preferentially unwind the highest market-cap subsets of their portfolios), and found that systemic risk is strongly dependent on fund behavioural policy and the properties of assets receiving exogenous shocks. The final contribution was to quantitatively assess, for

the first time, a novel circuit breaker policy from [BlackRock \(2015\)](#). We found that the policy has the potential to reduce systemic risk, and to add value to the existing “Level 3 circuit breaker” policy, but that it could also give rise to unintended negative consequences.

These results support our claim that, through enhanced realism via more complex models and large data, hybrid micro-macro modelling offers valuable insights to financial regulators unobtainable through the established micro-only or macro-only approaches.

6.2 Key Contributions

This thesis has addressed the following central research question:

What role do algorithmic trading systems have in the contagion of distress through a network of financial institutions, connected via shared-asset investments?

The key contributions that constitute our answer to this question are:

- Leverage management practice by investment funds engenders a contagion channel for intraday systemic risk, by allowing flash crashes to propagate between algorithmically traded assets. Systemic risk increases as funds deploy trading algorithms that enter the market with higher frequency, and with larger volume limits. This distressed sell order flow readily exceeds the ability of counterparty algorithmic traders to provide offsetting liquidity, resulting in flash crashes.
- Using our model to investigate the effect of network topology, we found the surprising result that portfolio crowding can, sometimes, be beneficial to systemic risk. This finding establishes the need for future studies to consider realistic non-uniform asset allocations.
- In a model with price impact that arises endogenously from a calibrated and validated microstructure component, high market cap assets: (1) are more prone to exhibiting flash crashes than low market cap assets, confirming empirical research by [Fry and Serbera \(2017\)](#), and (2) give rise to higher systemic risk if shocked, or if targeted for preferential unwind by distressed funds. This latter result is in *direct contrast* to theoretical macro-level only results from [Capponi and Larsson \(2015\)](#) and [Cont and Schaanning \(2019\)](#), who reported

that systemic risk is instead lower when high market cap assets are preferred for sale.

- Systemic risk monotonically increases with increasing fund leverage, and increasing market cap of distressed assets, however asset connectivity has a mixed effect on systemic risk. In particular, we find that assets with an intermediate level of ownership concentration were the most likely to trigger systemic events when shocked.
- We found that systemic risk is maximised when funds preferentially sell the highest market cap subset of their portfolios, and minimised when funds preferentially sell the lowest market cap subset. However, we also found that market-wide circuit breakers are particularly ineffective at mitigating systemic events when funds preferentially unwind low market cap assets. We find that a policy proposal by [BlackRock \(2015\)](#) is an effective remedy in this situation.
- BlackRock's policy requires the specification of a parameter related to the fraction of assets experiencing distress. We offer the first quantification of the appropriate range of this parameter, finding it becomes ineffective beyond 20%. This result is of potential interest to regulators who are faced with the prospect of assessing BlackRock's policy, and selecting an appropriate parametrisation.
- Our results represent the first study that connects microstructure policies with macro-level systemic risk. We thus offer a new lens through which intraday policy can be examined and assessed.
- Our Big Data micro-macro calibration advances the state-of-the-art of quantitatively validated market microstructure modelling, and shows the feasibility and value of incorporating large-scale network datasets to produce a fully valid fund-asset network model. We thus contribute to the advance of the discipline of ABM from Level 1 to Level 2 of the taxonomy of [Axtell and Epstein \(1994\)](#).

In order to support the micro-macro model integration at the heart of this thesis, it was necessary to overcome a number of conceptual and technical challenges. In particular:

- Which components of the model should be considered agents, and which form part of the environment? What specific behaviour should agents have, and how can their initial conditions be determined from empirical data, from industry

knowledge, or from previously-reported behavioural experiments? How is internal accounting consistency to be achieved in the model? These questions formed part of the input validation of our the model. The key step was to establish leverage management as the key macro-level model component (with the potential for initialisation using regulatory 13-F data), and to identify a realistic LOB-based model to facilitate price discovery in the model in a way that is sufficiently realistic to allow us to address our research questions.

- How are the different timescales over which agents operate best represented in the model? How should the connection be made between funds that need to sell assets, and the asset microstructure model? How should market executions be communicated back to funds? In order to facilitate the deep connection between fund actions and the placement of orders in the LOB component it was necessary to develop a complete order management system for the delegation of fund trading (controlled from the macro, fund level) to agents at the micro level with responsibility for the life-cycle management of individual limit orders.
- How should this all be performed in a manner that is feasible, given CPU, memory, and time constraints? Due to the large memory footprint of the fully calibrated model, it was necessary to utilise concurrent programming in-process in order to achieve parallelisation. It was not feasible to parallelise by simply instantiating multiple Monte Carlo instances. We applied extensive CPU and memory profiling techniques in order to make simulation at full real-world scale feasible with the available computational resources.
- With the model defined, implemented, and optimised, how should it be calibrated with real-world data? Given micro-macro models have rarely been reported in the literature, there were no prior examples of micro-macro calibration and validation to guide this process. There were also no examples of micro-level calibration using a range of assets that is representative of the full market cap spectrum. It was particularly challenging to adapt single-asset microstructure calibration methods to facilitate the all-of-market calibration we reported in Chapter 4 of this thesis.

Having overcome these challenges, the models and approaches that we have developed can be readily extended to consider new research directions, and we consider promising opportunities in the following section.

6.3 Future Work

The need for, and value of, sufficient data has been a recurring theme throughout this thesis. The quantitative calibration and validation of our model was made possible by the scope and fidelity of our Big Data microstructure dataset. Nevertheless, it was not possible to obtain sufficient data to exhaustively calibrate the model — in particular it was not possible to obtain access to data regarding the leverage and capital of individual funds present in the 13-F portfolio ownership dataset. Regulators have started collecting these data, and so in future it may become feasible to complete this final aspect of model calibration. This undertaking would further enhance the realism of the model, and would enable a deeper quantitative analysis of systemic risk exhibited by the model, going beyond trends and patterns, to establishing the absolute level of systemic risk with statistical confidence.

Going hand-in-hand with data considerations, each of the main contributions in this thesis — building the hybrid micro-model; validation and calibration; *in silico* policy analysis and regulatory extensions — has highlighted the extensive computational resources necessary to infer results at an acceptable level of statistical significance. Given the emergent nature of ABM, and the inherent need for repeated Monte Carlo sampling, it seems likely that computational requirements will continue to be material for ABM researchers for the foreseeable future. Approaches that could act to reduce the computational burden, particularly though not only during calibration, include Gaussian process regression and other machine learning techniques (Lamperti et al., 2018; van der Hoog, 2019). It would also be interesting to consider in more depth the trade-offs between competing calibration methods, such as variants of genetic algorithms (Hansen and Ostermeier, 2001; Kimbrough et al., 2008).

A key driver of the increasing traction of ABM is the freedom with which models can be extended to express phenomena of interest. There are many relevant directions in which the models and results presented in this thesis could be extended. Perhaps the most important aspect to consider is the intelligence of the agents themselves. The agents in the model we developed in this thesis are zero-intelligence, however, it would be possible to endow agents with more strategic sophistication in future. This would allow exploration of the effects on systemic risk of intraday decision-making by agents, for example, (1) HFT and market maker agents could condition their liquidity provision on expectations of profit and trading halt activation, (2) fund agents could implement predictive trading strategies and optimise their exposure to overlapping risk factors, and (3) bank agents could exercise discretion about how

and when to unwind the portfolios of defaulted funds. Two key issues arise when agents undertake strategic behaviour. Firstly, it is necessary to consider the critique by Lucas (1976) that strategic agents will change their behaviour in anticipation of a policy amendment, and thus agents need to be endowed with adaptiveness, in addition to predictive capability, in order for model inferences to remain valid. Secondly, if the optimal behaviour of agents is a function of the behaviour of other agents, then the tools of game theory offer the most appropriate methodological framework for reasoning about collective agent behaviours. Efforts to deploy game theoretic analysis to real-world-scale market systems offer a promising future research direction (Wellman and Wah, 2017).

Within the zero-intelligence paradigm, several lines of research are immediately accessible as extensions to our model, given access to appropriate data and computational resources. It would be valuable to add a further agent type to the micro-level model component to represent *arbitrageurs* who trade based on assumptions of asset co-movement. Arbitrage of the S&P 500 index future against the S&P 500 exchange-traded fund (ETF) was implicated in the Flash Crash of 6th May 2010, and it would be interesting to embed this potential additional channel of asset distress contagion within the model. This agent type would also facilitate the study of cross-jurisdictional distress propagation. International arbitrage agents (or indeed, agents seeking to hedge exposures between international markets without the explicit goal of generating profits) have the potential to propagate systemic risk between markets, leading to the possibility of a truly global intraday crisis (Cliff and Northrop, 2012).

Looking to the future, advances in methodology, data access, and computer power mean that agent-based modelling is poised to become a standard approach in the financial regulator's tool-kit. As regulators endeavour to stay one step ahead of the profound transformations impacting modern financial markets, the need for new approaches and deeper understanding is more pressing than ever. It is our hope that this thesis helps the discipline of agent-based modelling move toward this goal.



Trader Agent Details

The behavioural details of the trader agent classes are as follows:

Small (noise) Traders Small traders place minimally-sized limit orders according to a uniform distribution around prevailing bid and ask prices on the LOB and thus are truly zero-intelligence ([Gode and Sunder, 1993](#)). These orders may be aggressive and remove liquidity from the book or passive and add liquidity which may match against subsequent aggressive orders of opposing sign. Small traders have no dependence on price or volume changes, and we do not model inventory limits for this agent type. Their slow interaction speed means that any individual small trader is highly unlikely to build a significant position during a simulation run. We deviate from [Paddrik et al. \(2012\)](#)'s implementation by increasing the spread within which orders can be placed from 10 ticks to 1000 ticks. This change allows our model to better capture the relative price stylised fact, as discussed in Section 3.7, and is our method of modelling *stub quote* order placement. Stub quotes are orders placed far from prevailing prices that are not intended for execution, but that were identified as playing a role in the Flash Crash of 6th May 2010 ([CFTC and SEC, 2010](#)).

Fundamental buyers / sellers Fundamental buyers / sellers are in possession of a private valuation for the asset and will execute when prevailing prices yield a trading surplus ([Wellman, 2011](#)). Fundamental buyers / sellers seek to balance trading impact with immediacy ([Paddrik et al., 2012](#)) and so are represented as gradually building

positions throughout the simulation run. Inventory limits are not modelled for such agents, however the agents do possess slightly above zero intelligence (known as ϵ -intelligence). As in Paddrik et al. (2012), fundamental buyers / sellers are sensitive to *toxic order flow* (Easley et al., 2012). If prices trend either higher or lower by more than 70 ticks over a one-minute horizon, fundamental buyers / sellers will avoid placing any orders until the market stabilises. This behaviour was identified during the real 6th May 2010 Flash Crash by Kirilenko et al. (2017). Fundamental buyers / sellers have an extremely important function in flash crash generation — without the action of fundamental buyers reacting to far from fundamental prices, prices would not rebound following the initial crash.

Opportunistic Herding These agents represent a class of traders that react to external information such as corporate news dissemination. The order placement direction probability of opportunists follows a stochastic process. These agents operate over an intermediate timescale and are modelled with small inventory limits. It is not entirely clear from Paddrik et al. (2012) or later related works (Hayes et al., 2014; Paddrik et al., 2017; Bookstaber and Paddrik, 2015) what the impact of opportunists is on the model dynamics, but since Kirilenko et al. (2017) identify this group of real-world market participants we include them for completeness.

Market Makers Market makers are algorithmic traders that maintain simultaneous limit orders on both sides of the LOB and exclusively provide (rather than remove) liquidity. As such, they have been identified as systemically important for the efficient operation of asset markets (Wah et al., 2017; Bookstaber and Paddrik, 2015). Paddrik et al. (2012) model market makers with small inventory capacity and, furthermore, model the documented tendency for market makers to retreat from trading by widening spreads or indeed suspending order placement entirely when they detect toxic order flow (Easley et al., 2012). Similarly to fundamental buyers / sellers, when the one-minute price trend exceeds 24 ticks, market makers cease trading (calibration based on Kirilenko et al. (2017)). When the market maker inventory is saturated, the agents attempt to liquidate positions passively. They do this by providing liquidity only on the offsetting side of the LOB. We model market makers exiting the liquidation state once their absolute inventory reaches half of its maximum limit.

High Frequency Traders High frequency traders (HFTs) in the model act as very fast market makers first and foremost, though they also possess a simple shared

trading signal (Paddrik et al., 2012) which skews their order placement side probability in the direction of LOB imbalances. An imbalance in this context refers to a difference in the quantity of available shares resting on either side of the LOB. The probability that an HFT will place a buy order in the j th asset is given by $P(buy)_j = Q_j^{bid} / (Q_j^{bid} + Q_j^{ask})$, where Q_j^{bid} (Q_j^{ask}) refers to the total quantity of shares available at the best bid (ask) price (Paddrik et al., 2012). This signal can be considered a crude prediction algorithm. The intuition behind it is that, when $Q_j^{bid} \gg Q_j^{ask}$ it is likely that the price will rise in the near future, and so it is rational for the HFT agent to buy, in order to profit from the expected price move. The public nature of the signal results in a well-documented herding effect (Menkveld, 2016). When their small inventory¹ is saturated, HFTs aggressively reduce positions by taking liquidity. This effect has been implicated in deepening the 6th May 2010 Flash Crash as HFTs passed positions back and forth between themselves in a “hot potato” fashion (Abrol et al., 2016; Aldrich et al., 2016).

Sensitivity of the Model to Trader Behaviour In sensitivity analyses, we found that the model’s qualitative behaviour was robust to the replacement of opportunistic agents with pure zero-intelligence traders of equivalent market activity timescale. We found that replacing market makers with zero-intelligence equivalents caused flash crashes to occur more frequently (which is to be expected as market makers strictly provide liquidity and thus stabilise markets). We further found that replacing fundamental buyer/seller agents with zero-intelligence traders destabilised the market leading to inevitable price collapse to zero value with no reversionary bounce. These results are in agreement with Chiarella et al. (2002) who found (p.350) “these results indicate that all three trader types are necessary in some form to generate realistic looking return dynamics”. The role of HFT agents was also found to be critical to model behaviour. We found that replacing HFT traders with noise traders of equal market entry speed results in extremely unstable behaviour, to which the introduction of distressed sellers at any level of participation (η) has no discernible effect, and flash crashes do not occur (there is no characteristic fall and rebound of prices, just noise). Paddrik et al. (2012), Vuorenmaa and Wang (2014) and Jacob Leal and Napoletano (2019) all find the presence of HFT traders is necessary for the emergence of flash crashes, and, moreover, find that flash crashes become less likely as the number of HFT agents is reduced. We found that removing HFT agents (without replacement

¹The findings of CFTC and SEC (2010) suggests a total inventory across all HFTs of 3000 contracts. However, Paddrik et al. (2012), Hayes et al. (2014), and Paddrik et al. (2017) appear to have utilised 3000 contracts per HFT.

by fast noise traders) results in the suppression of flash crashes except for very high values of distressed seller participation above 40% of rolling one-minute market volume. For comparison, the presence of HFT agents results in a high probability of flash crashes when distressed seller participation is just 9% (the value adopted by the real-world distressed seller in the Flash Crash of 6th May 2010).

B

Bipartite Network Generation

Algorithm 1 presents our method for constructing random bipartite networks with controllable diversification and crowding.

Algorithm 1 makes the following function calls:

***rank*(j, L)** returns the index that the j th element of list L would receive, if the elements of L were sorted in ascending numerical order.

***weightedChoice*($n_a, weights, replacement$)** returns $j' \in (1, \dots, n_a)$ with probability given by the j' th element of the vector *weights*, with or without replacement.

***nextInvestmentSize*(i)** returns values in units of currency representing the sizes of investment fund i makes in each of the k_{fund} assets in its portfolio. Investments are returned in decreasing size order based on repeated samples from a lognormal distribution $\text{Lognormal}(0, 0.66^2)$ constrained such that the fractional investments to the k_{fund} assets sum to unity.

Algorithm 1 Random Bipartite Networks with Preferential Attachment.

Require: $n_f, n_a, \rho, \beta, C^0, \lambda^0$

```

1:  $A_{ij} \leftarrow 0, \forall i \in (1, \dots, n_f), \forall j \in (1, \dots, n_a)$ 
2: for all  $i \in (1, \dots, n_f)$  do
3:   # Calculate total value already invested in each asset
4:   for all  $j \in (1, \dots, n_a)$  do
5:      $total_j \leftarrow \sum_i^{n_f} A_{ij}$ 
6:   end for
7:   # Calculate preferential probability of selecting each asset
8:   for all  $j \in (1, \dots, n_a)$  do
9:      $r_j \leftarrow rank(j, total)$ 
10:    if  $\beta < 0$  then
11:       $r_j \leftarrow n_a - r_j + 1$ 
12:    end if
13:     $p_j \leftarrow r_j^\beta / \sum_j^{n_a} r_j^\beta$ 
14:  end for
15:  # Calculate fund fractional investment to assets
16:   $k_{fund} \leftarrow max(\lfloor \rho n_a \rfloor, 1)$ 
17:  for all  $k \in (1, \dots, k_{fund})$  do
18:     $j' \leftarrow weightedChoice(n_a, p, \text{false})$ 
19:     $A_{ij'} \leftarrow nextInvestmentSize(i)$ 
20:  end for
21:  # Convert normalised fractional investments to currency values
22:   $A_{ij} \leftarrow C^0(1 + \lambda^0)A_{ij} / \sum_j^{n_a} A_{ij}$ 
23: end for
24: return  $A$ 

```

C

Global Sensitivity Analysis Supplement

In Section 4.5 we presented the results of a sample-based global sensitivity analysis of the microstructure ABM. In this supplementary section we investigate the number of samples required to provide sufficient precision in the assessment of the total sensitivity index (see Equation 4.22). We measure precision for parameter k as the ratio of the standard deviation of total order sensitivity $\text{stddev}(S)$ to S_{T_k} . Low values of this ratio indicate higher precision, and are thus to be preferred.

Figure C.1 presents the mean and 95% confidence interval across 17 input parameters for each output moment as the number of sampled parameter vectors θ increases to $N = 97200$. The figure presents separate results for each output moment. The number of Monte Carlo replicates for each parameter vector θ is fixed at $I = 8$. It is clear that precision increases monotonically with an increasing number of samples. This is consistent with expectations. It is also clear that although the sensitivity analysis achieves different levels of precision for different output moments, the precision of all moments improves according to the same monotonic relationship. $N = 97200$ samples is sufficient to achieve a precision of better than 20% for all output moments, and of better than 10% for seven of the ten output moments. Although precision could clearly be further improved with more samples, this analysis suggests $N = 97200$ to be a sufficient number of samples for the sensitivity analysis of our model.

Figure C.2 presents an alternative assessment of the convergence of the SA. The figure shows the Spearman's rank correlation coefficient r_s of S_{T_k} for 17 parameters

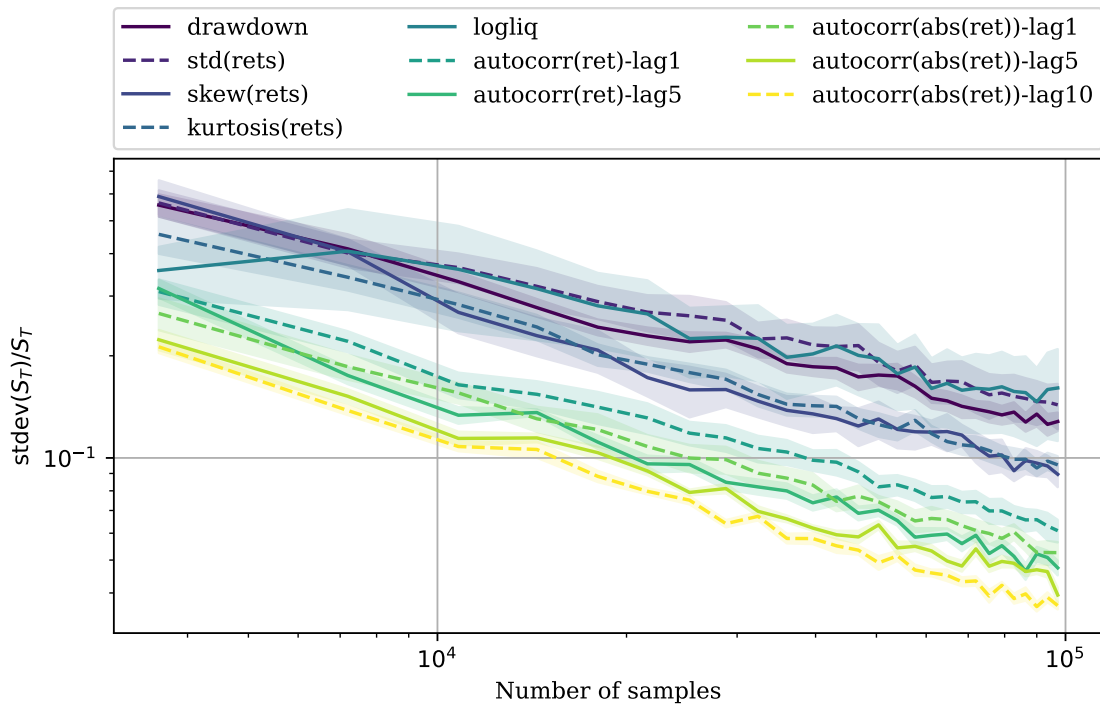


Figure C.1: The ratio of the standard deviation of total order sensitivity $\text{stddev}(S_{T_k})$ to S_{T_k} as we increase the number of parameter vectors θ sampled from a quasi-random Sobol process. We present the mean and 95% confidence interval across 17 parameters (indexed by $k \in (1, \dots, 17)$) for each output moment on a log-log scale. Lower values indicate higher precision.

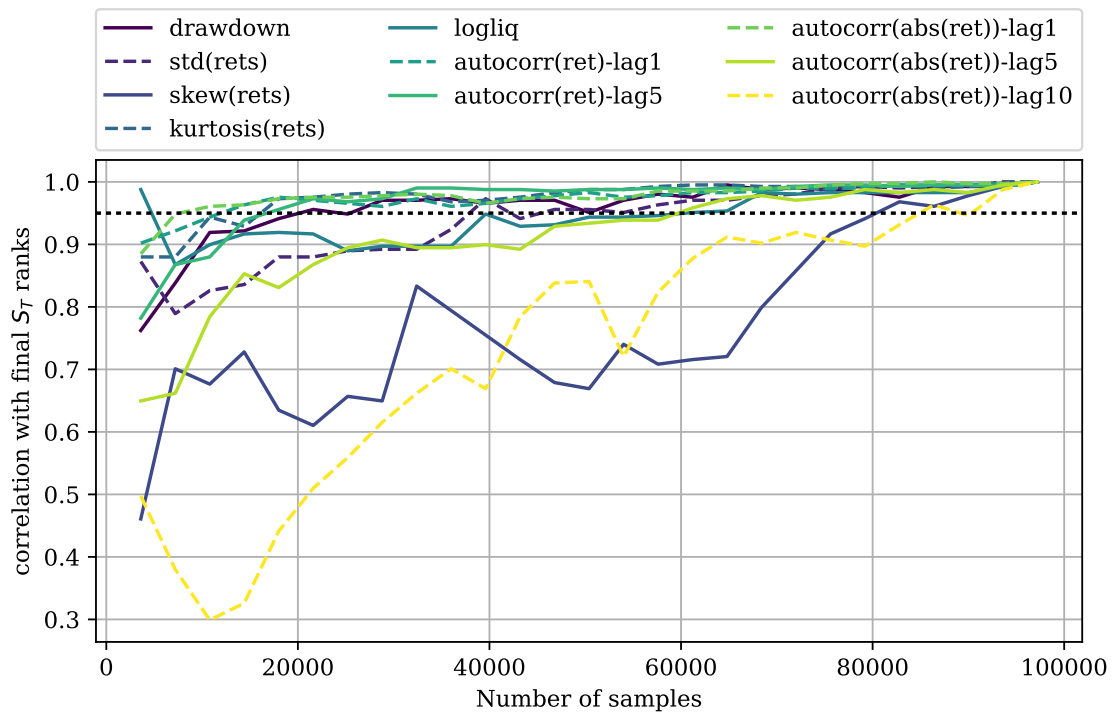


Figure C.2: Spearman's rank correlation coefficient r_s of S_T for 17 parameters to the parameter ranks obtained using the full set of 97200 Sobol samples. Confidence intervals are suppressed for clarity. Correlation at $r_s = 95\%$ is indicated by a horizontal dashed line.

to the parameter ranks obtained using the full set of $N = 97200$ Sobol' samples. We present a series for each output moment, with confidence intervals suppressed for clarity. The figure demonstrates that $N = 60000$ samples is sufficient for sensitivity ranks to exhibit 95% correlation to the sensitivity ranks obtained using the full set of $N = 97200$ samples for eight out of the ten output moments. The remaining two moments (`skew(rets)` and `autocorr(abs(ret))-lag5`) require $N = 80000$ and $N = 85000$ samples respectively to achieve 95% correlation. As with Figure C.1, these results suggest that sensitivity analysis would benefit somewhat from more samples, but that $N = 97200$ appears to be a sufficient sample size for the sensitivity analysis of our model.

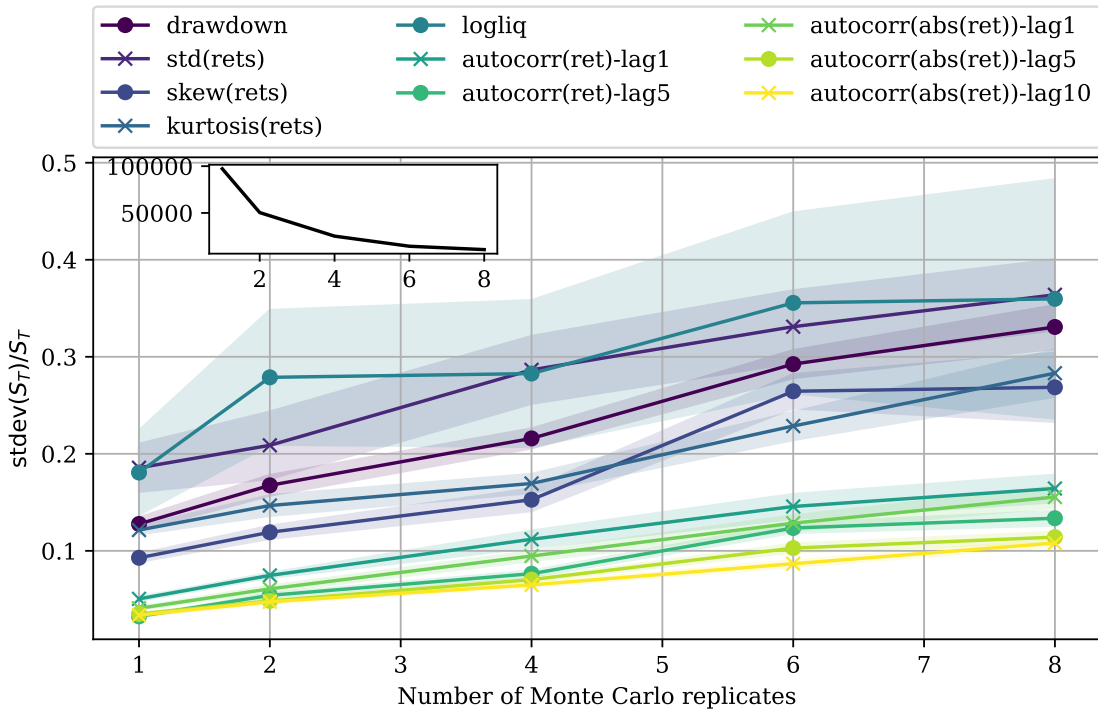


Figure C.3: The effect of increasing the number of Monte Carlo replicates on the ratio of the standard deviation of total order sensitivity $\text{stddev}(S_T)$ to S_T with a fixed computational budget. We present the mean and 95% confidence interval across 17 parameters for each output moment. Lower values indicate higher precision. **Inset:** number of samples used vs number of Monte Carlo replicates.

Figure C.3 presents the SA precision attained as the number of Monte Carlo replicates used to measure $\mu(\theta)$ is varied over $I \in (1, \dots, 8)$, whilst fixing the product $I \times N = 97200$. The inset of Figure C.3 indicates the reduction in N as I is increased. For a biological ABM, and an alternative global sensitivity analysis method known as *partial rank correlation coefficient*, Marino et al. (2008) find that increasing the

number of replicates I yields a greater benefit in SA precision than increasing the number of samples N for 7 of their 12 parameters, under the assumption of a fixed computational budget. However, Figure C.3 clearly shows that precision *decreases* as I increases, which is the opposite finding to Marino et al. (2008). This reduction in SA precision is monotonic for all model output moments. These results strongly suggest that sensitivity analysis of the model using the method of Saltelli et al. (2010) benefits more from increasing the number of samples N , whilst retaining a relatively low number of Monte Carlo evaluations I .

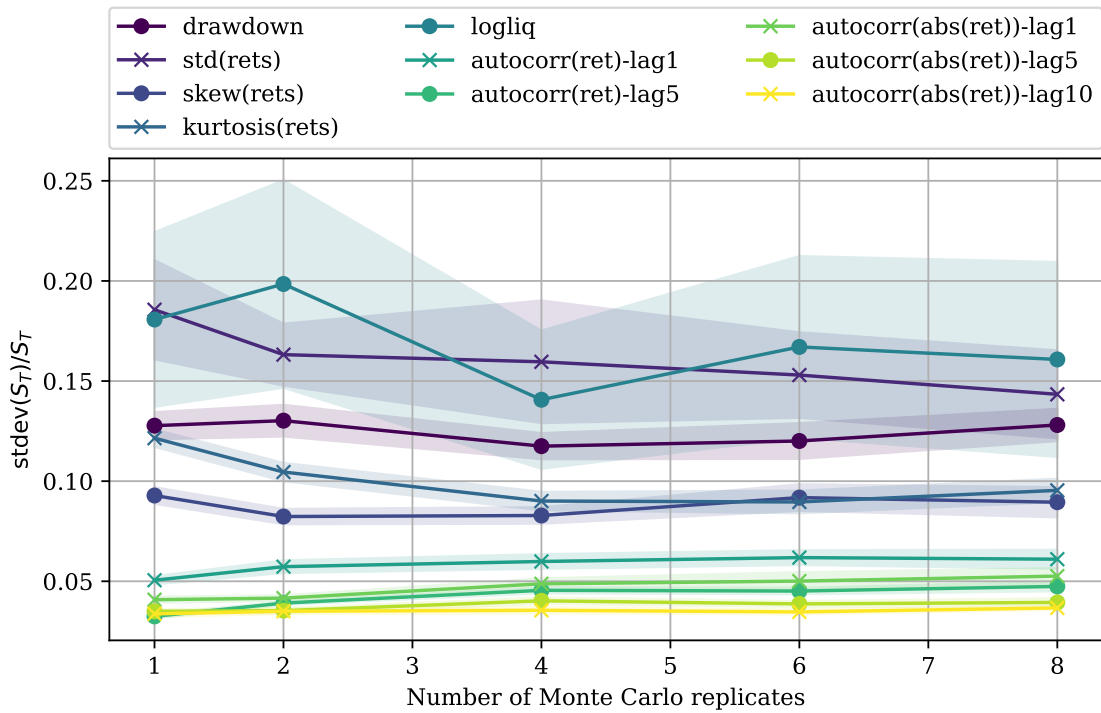


Figure C.4: The effect of increasing the number of Monte Carlo replicates on the ratio of the standard deviation of total order sensitivity $\text{stddev}(S_T)$ to S_T with an increasing computational budget. We present the mean and 95% confidence interval across 17 parameters for each output moment. Lower values indicate higher precision. The number of samples is held fixed at $N = 97200$, and so the total computational budget increases with the number of Monte Carlo replicates.

Figure C.4 addresses the naturally arising question of whether increasing the number of Monte Carlo replicates I is beneficial at all. Our hypothesis is that a reduction in error in the measurement of $\mu(\theta)$ obtained by taking the average of a higher number of Monte Carlo replicates I would lead to a consequent improvement in the precision of subsequent sensitivity analysis. Figure C.4 presents the SA precision as I is increased, while holding fixed $N = 97200$. The results do not exhibit a clear

pattern, and we will not claim that they either support or refute our hypothesis. It could be that $I = 8$ is simply too small for the resulting reduction in measurement error to have a significant influence on sensitivity analysis, but the extremely high computational cost involved in the exercise is prohibitive to deeper investigation.

D

Out of Sample Validation Supplement

Figure D.1 presents an alternative view of the probability of flash crash occurrence to that of Section 4.10. The figure presents the probability of flash crash occurrence $P_{FC}(b, \mathcal{Q}, \eta, \delta)$ in the calibrated ABM as a function of distressed selling agent properties for each market cap band from 2009-Q1 to 2018-Q1. Each panel presents the dependence of P_{FC} on distressed seller volume fraction η , with a line for each value of distressed seller order placement timescale δ (δ colour assignments are given in the figure legend). Market cap band b is shared by panels in a single column, and the date of the calibration dataset is shared by panels along a single row.

As in Figure 4.10, Figure D.1 demonstrates that for values of $\delta < 20$ there exists a critical value of η , above which flash crashes are observed with probability approaching 100%. This figure is included for completeness, and because the manner of presentation facilitates comparison with earlier results in Section 3.8. However, the reader is referred to Section 4.10 for a full discussion of these results, including consideration of dependence on market cap band and date.

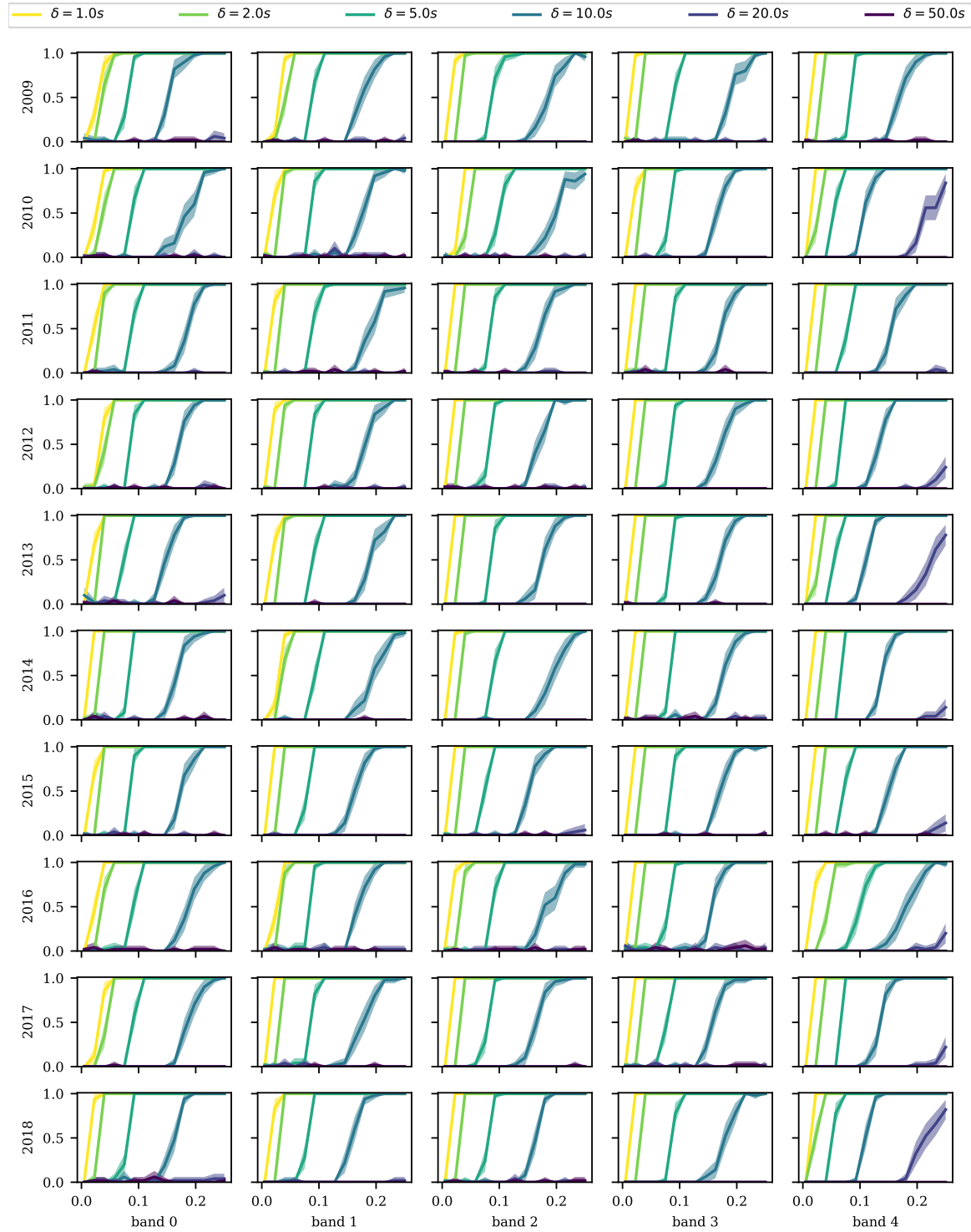


Figure D.1: Probability of flash crash occurrence in the calibrated ABM as a function of distressed selling agent properties for each market cap band from 2009-Q1 to 2018-Q1. The order size limit as a fraction of rolling one-minute market volume, η , is plotted on the horizontal axis. δ refers to the mean interval between sell order placements in seconds. We show the mean and 95% confidence interval sampled over 100 Monte Carlo trials.



Policy and Interventions Supplement

Figure E.1 presents the mean and 95% confidence interval for the extent of contagion ($\Omega_{contagion}$) as a function of leverage and shocked asset market cap, for the same set of experimental scenarios presented in Figure 5.3. The figure demonstrates that the system is *robust-yet-fragile*, and systemic collapse is effectively total whenever a default cascade occurs (Section 5.2.2).

Figure E.2 presents the mean and 95% confidence interval for the probability that at least one fund enters the margin call state, as a function of leverage and shocked asset market cap, for the same set of experimental scenarios presented in Figure 5.3. We observe as-expected trends that increasing fund leverage, and increasing size of shocked asset, both give rise to increased probability that one or more funds will experience margin calls. These results are discussed in Section 5.4.4.

Figure E.3 presents the mean and 95% confidence interval for the fraction of assets experiencing flash crashes, as a function of leverage and shocked asset market cap, for the same set of experimental scenarios presented in Figure 5.3. It is apparent from comparison with Figure 5.3 that the expected extent of flash crash contagion is extremely highly correlated with the probability of default contagion.

Figure E.4 presents the mean and 95% confidence interval for the fraction of Monte Carlo replicates where at least one fund enters the distressed margin call state, for the set of experimental scenarios discussed in detail in Section 5.4.2. The figure demonstrates that the probability that at least one fund enters a distressed state is a function of ω_{mcap} (as we would expect), and is not a function of fund

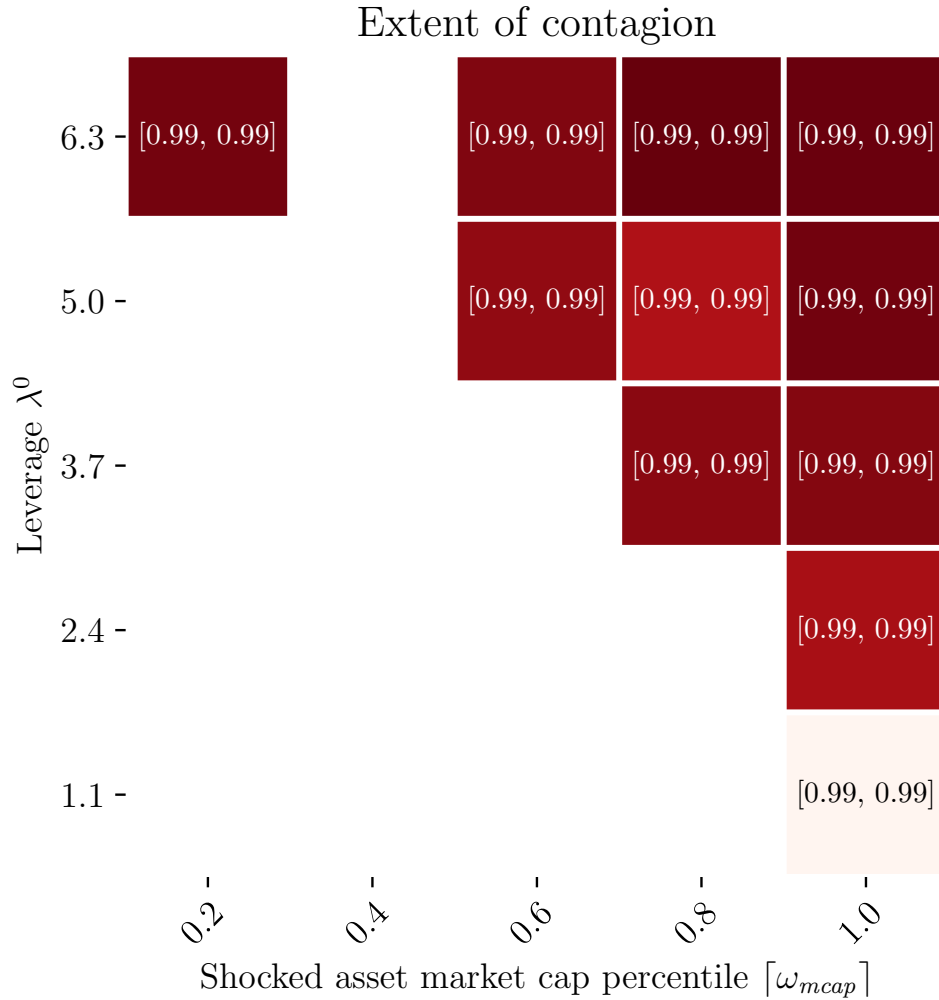


Figure E.1: Comparing the extent of contagion for leverage $\lambda^0 \in [1.1, 6.3]$ and shocked asset centrality $\omega_{mcap} \in \{[0, 20), [20, 40), [40, 60), [60, 80), [80, 100]\}\%$. Cells shading represents the mean across 31 Monte Carlo replicates. Cells are labelled with 95% confidence intervals.

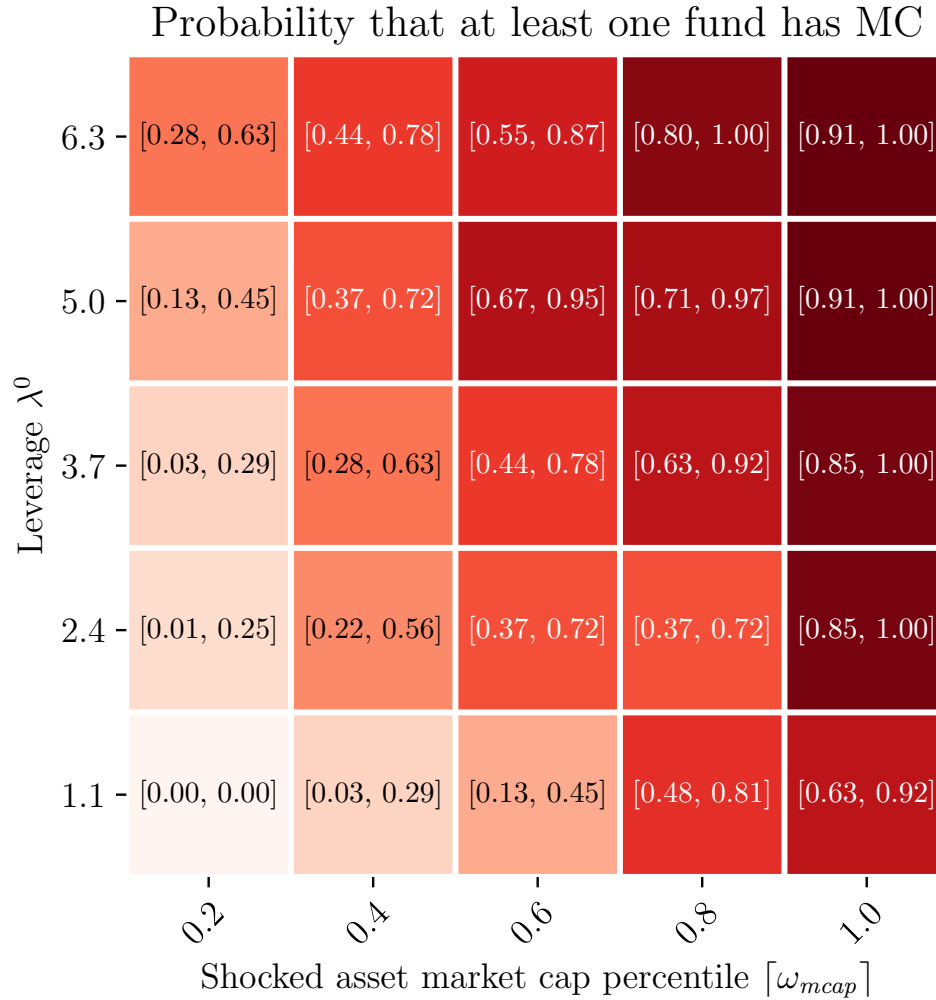


Figure E.2: Comparing the probability of distress for leverage $\lambda^0 \in [1.1, 6.3]$ and shocked asset centrality $\omega_{mcap} \in \{[0, 20), [20, 40), [40, 60), [60, 80), [80, 100]\}\%$. Cells shading represents the mean across 31 Monte Carlo replicates. Cells are labelled with 95% confidence intervals.

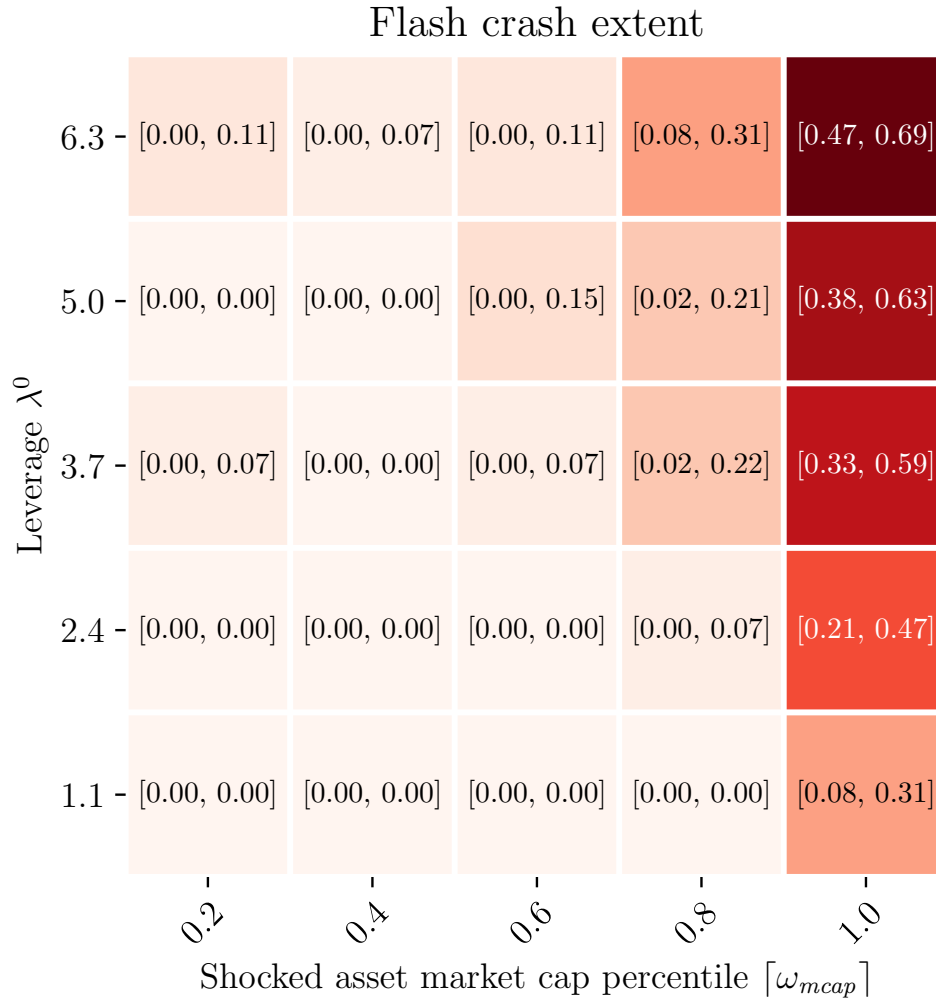


Figure E.3: Comparing the extent of flash crash occurrence for leverage $\lambda^0 \in [1.1, 6.3]$ and shocked asset centrality $\omega_{mcap} \in \{[0, 20), [20, 40), [40, 60), [60, 80), [80, 100]\}\%$. Cells shading represents the mean across 31 Monte Carlo replicates. Cells are labelled with 95% confidence intervals.

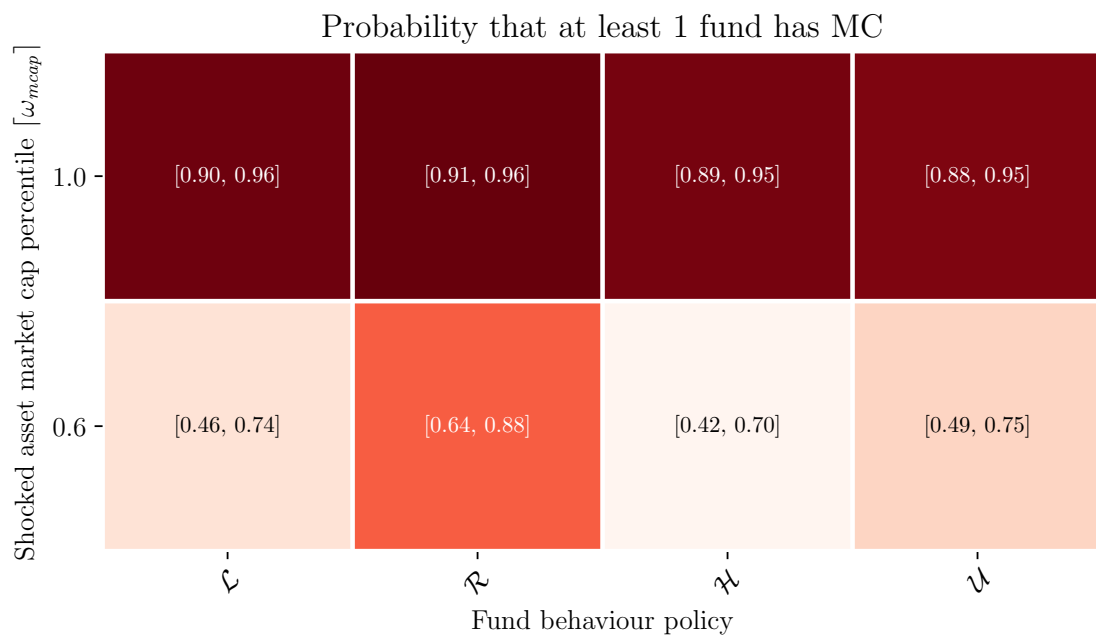


Figure E.4: Comparing the probability of fund distress for shocked asset centrality quintiles of ω_{mcap} and fund behavioural policies $\mathcal{F}$ at leverage $\lambda^0 = 5$, under the current US regulatory standard LULD-MWCB trading halt policy. Results are based on 318 Monte Carlo replicates ($\omega_{mcap}^{\text{high}}$) and 50 Monte Carlo replicates ($\omega_{mcap}^{\text{low}}$) with cell shading indicating the mean and labelling indicating the 95% confidence interval.

deleverage behaviour $\mathcal{F}$. This is also as expected, because the consequent selling behaviour of the first fund to enter distress cannot affect the probability that that fund originally entered the distressed state. We observe for $\omega_{\text{mcap}}^{\text{high}}$ that the expectation that at least one fund enters distress is approximately 93%, with insignificant variation across fund behaviours. For $\omega_{\text{mcap}}^{\text{low}}$, at least one fund is expected to enter distress with approximately 64% probability. The variation between fund behaviours is an artefact of the low number of Monte Carlo samples for $\omega_{\text{mcap}}^{\text{low}}$ (50). The composite 64% estimate is within the 95% confidence interval of the results associated with each fund behaviour.

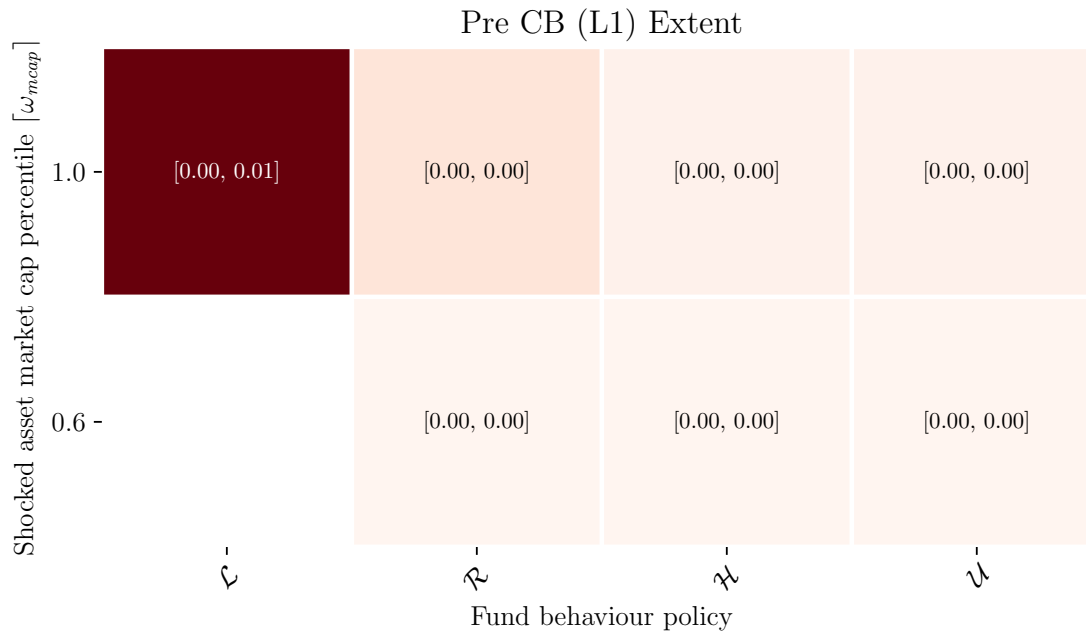


Figure E.5: Comparing pre market-wide circuit breaker contagion extent for L1 circuit breaker activation, for shocked asset centrality quintiles of ω_{mcap} and fund behavioural policies $\mathcal{F}$ at leverage $\lambda^0 = 5$, under the current US regulatory standard LULD-MWCB trading halt policy. Results are based on 318 Monte Carlo replicates ($\omega_{\text{mcap}}^{\text{high}}$) and 50 Monte Carlo replicates ($\omega_{\text{mcap}}^{\text{low}}$) with cell shading indicating the mean and labelling indicating the 95% confidence interval.

Figure E.5 presents the extent of contagion immediately prior to the triggering of Level 1 circuit breakers. The result shows that no funds are expected to have defaulted at the time of a Level 1 circuit breaker, except in the case of $\omega_{\text{mcap}}^{\text{high}}$ under

the $\mathcal{L}$ policy where $0.5(\pm 0.5)\%$ of funds are expected to enter default prior to the Level 1 MWCB. These findings are discussed in Section 5.4.5.

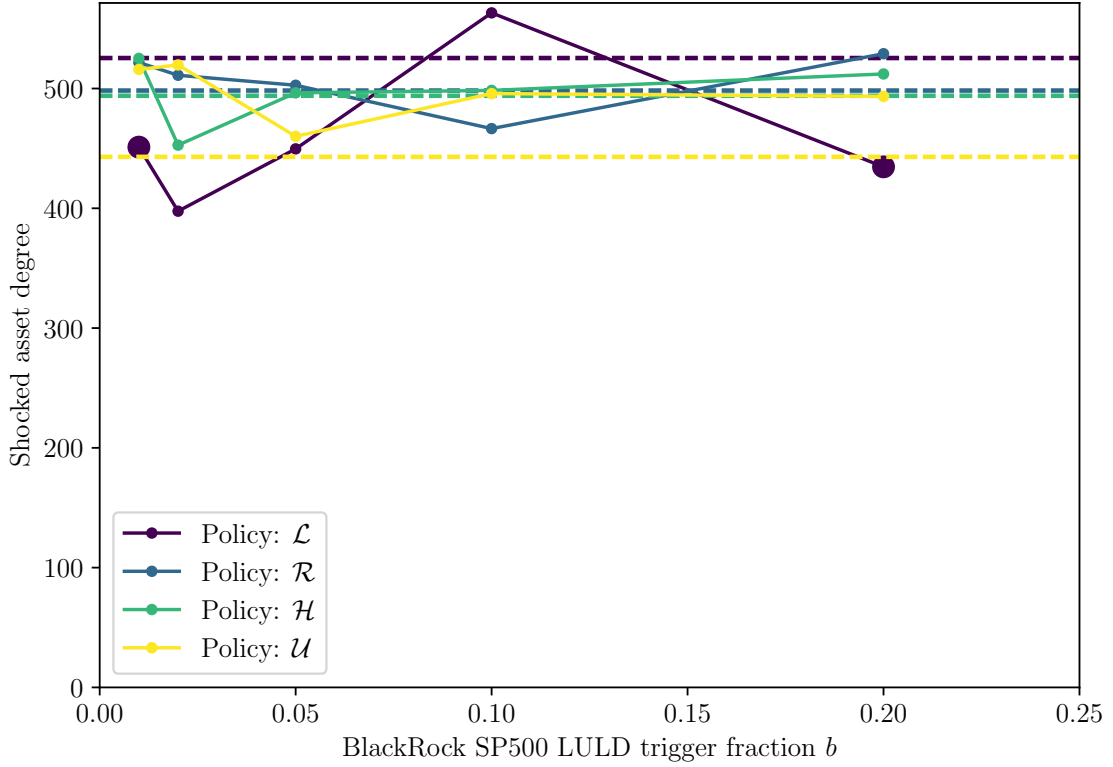


Figure E.6: Mood's χ^2 comparison between shocked asset degree distributions for real-world LULD-MWCB policy (dashed lines) and under different choices for the BlackRock threshold parameter b (solid lines), for fund behavioural policies $\mathcal{F}$. Statistically significant differences between the LULD-MWCB-BKRK and LULD-MWCB policies from 150 Monte Carlo replicates at $p < 0.05$ are indicated with large hollow circles, and at $p < 0.1$ with large filled circles.

Figure E.6 presents a comparison of the mean shocked asset degree between LULD-MWCB and LULD-MWCB-BKRK experimental scenarios. The figure is augmented with robust χ^2 test results (Mood's median test), showing that no scenario is different from the LULD-MWCB baseline at $p < 0.05$, and experiments $(\mathcal{L}, b_1)$ and $(\mathcal{L}, b_{20})$ exhibit borderline significance at $p < 0.1$. This could be resolved, if necessary, with further Monte Carlo replicates for those specific scenarios.

Bibliography

- Abrol, S., Chesir, B., and Mehta, N. (2016). High frequency trading and us stock market microstructure: A study of interactions between complexities, risks and strategies residing in us equity market microstructure. *Financial Markets, Institutions & Instruments*, 25(2):107–165.
- Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015a). Networks, shocks, and systemic risk. Technical report, National Bureau of Economic Research.
- Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015b). Systemic risk and stability in financial networks. *The American Economic Review*, 105(2):564–608.
- Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *The Review of Financial Studies*, 30(1):2–47.
- Ackert, L. F. (2012). The impact of circuit breakers on market outcomes. *UK Government Office for Science, Foresight Driver Review-The Future of Computer Trading in Financial Markets*.
- Admati, A. R. and Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies*, 1(1):3–40.
- Adrian, T. and Brunnermeier, M. K. (2016). CoVaR. *The American Economic Review*, 106(7):1705–1741.
- Adrian, T. and Shin, H. S. (2010). Liquidity and leverage. *Journal of financial intermediation*, 19(3):418–437.
- Adrian, T. and Shin, H. S. (2013). Procyclical leverage and value-at-risk. *The Review of Financial Studies*, 27(2):373–403.
- Aldrich, E. M., Grundfest, J., and Laughlin, G. (2016). The flash crash: a new deconstruction. *Available at SSRN 2721922*.

- Allen, F. and Babus, A. (2008). Networks in finance. *Wharton Financial Institutions Center Working Paper*, (07-08).
- Allen, F. and Gale, D. (2000). Financial contagion. *Journal of political economy*, 108(1):1–33.
- Aloud, M., Fasli, M., Tsang, E., Dupuis, A., and Olsen, R. (2017). Modeling the high-frequency FX market: An agent-based approach. *Computational Intelligence*, 33(4):771–825. 10.1111/coin.12114.
- Amihud, Y. and Mendelson, H. (1991). Liquidity, asset prices and financial policy. *Financial Analysts Journal*, 47(6):56–66.
- Anand, A., Irvine, P., Puckett, A., and Venkataraman, K. (2013). Institutional trading and stock resiliency: Evidence from the 2007–2009 financial crisis. *Journal of financial Economics*, 108(3):773–797.
- Anand, K., van Lelyveld, I., Banai, Á., Christiano Silva, T., Friedrich, S., Garratt, R., Halaj, G., Hansen, I., Howell, B., Lee, H., Martínez Jaramillo, S., Molina-Borboa, J. L., Nobili, S., Rajan, S., Rubens Stancato de Souza, S., Salakhova, D., and Silvestri, L. (2018). The missing links: A global study on uncovering financial network structure from partial data. *Journal of Financial Stability*, 35:107–119.
- Ang, A., Gorovyy, S., and Van Inwegen, G. B. (2011). Hedge fund leverage. *Journal of Financial Economics*, 102(1):102–126.
- Arinaminpathy, N., Kapadia, S., and May, R. M. (2012). Size and complexity in model financial systems. *Proceedings of the National Academy of Sciences*, 109(45):18338–18343.
- Axtell, R. and Epstein, J. (1994). Agent based modeling: Understanding our creations. *The Bulletin of the Santa Fe Institute*, pages 28–32.
- Aymanns, C., Farmer, J. D., Kleinnijenhuis, A. M., and Wetzter, T. (2018). Models of financial stability and their application in stress tests. In Hommes, C. and LeBaron, B., editors, *Handbook of Computational Economics*, volume 4 of *Handbook of Computational Economics*, pages 329 – 391. Elsevier.
- Banwo, O., Caccioli, F., Harrald, P., and Medda, F. (2017). The effect of heterogeneity on financial contagion due to overlapping portfolios. *Advances in Complex Systems*, 19(08):1650016.

- Baptista, R., Farmer, J. D., Hinterschweiger, M., Low, K., Tang, D., and Uluc, A. (2016). Macroprudential policy in an agent-based model of the UK housing market. *Bank of England Working Paper*.
- Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Barde, S. and van der Hoog, S. (2017). An empirical validation protocol for large-scale agent-based models. *Bielefeld Working Papers in Economics and Management*. Available at SSRN 2992473.
- Bardoscia, M., Battiston, S., Caccioli, F., and Caldarelli, G. (2017). Pathways towards instability in financial networks. *Nature Communications*, 8:14416.
- Bargigli, L., Riccetti, L., Russo, A., and Gallegati, M. (2018). Network calibration and metamodeling of a financial accelerator agent based model. *Journal of Economic Interaction and Coordination*, pages 1–28.
- Battiston, S., Caldarelli, G., D’Errico, M., and Gurciullo, S. (2016a). Leveraging the network: a stress-test framework based on debtrank. *Statistics & Risk Modeling*, 33(3-4):117–138.
- Battiston, S., Farmer, J. D., Flache, A., Garlaschelli, D., Haldane, A. G., Heesterbeek, H., Hommes, C., Jaeger, C., May, R., and Scheffer, M. (2016b). Complexity theory and financial regulation. *Science*, 351(6275):818–819.
- Battiston, S., Gatti, D. D., Gallegati, M., Greenwald, B., and Stiglitz, J. E. (2012a). Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of Economic Dynamics and Control*, 36(8):1121–1141.
- Battiston, S., Puliga, M., Kaushik, R., Tasca, P., and Caldarelli, G. (2012b). Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Scientific Reports*, 2(541):1–6.
- Beale, N., Rand, D. G., Battey, H., Croxson, K., May, R. M., and Nowak, M. A. (2011). Individual versus systemic risk and the regulator’s dilemma. *Proceedings of the National Academy of Sciences*, 108(31):12647–12652.
- Ben-David, I., Franzoni, F., and Moussawi, R. (2012). ETFs, arbitrage, and shock propagation. Available at <https://bit.ly/324anae> Accessed 2018-09-19.

- Benoit, S., Colliard, J.-E., Hurlin, C., and Pérignon, C. (2017). Where the risks lie: A survey on systemic risk. *Review of Finance*, 21(1):109–152.
- Bertschinger, N., Mozzhorin, I., and Sinha, S. (2018). Reality-check for econophysics: Likelihood-based fitting of physics-inspired market models to empirical data. *arXiv preprint arXiv:1803.03861*.
- Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*.
- Bian, J., Da, Z., Lou, D., and Zhou, H. (2019). Leverage network and market contagion. *PBCSF-NIFR Research Paper*.
- Billio, M., Casarin, R., Costola, M., and Pasqualini, A. (2016). An entropy-based early warning indicator for systemic risk. *Journal of International Financial Markets, Institutions and Money*, 45:42–59.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, volume 1. Springer-Verlag New York.
- Bisias, D., Flood, M. D., Lo, A. W., and Valavanis, S. (2012). A survey of systemic risk analytics. *Annu. Rev. Financ. Econ.*, 4(1):255–296.
- BlackRock (2015). US equity market structure: lessons from August 24. *BlackRock ViewPoint*, pages 1–16.
- Blum, C. and Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM computing surveys (CSUR)*, 35(3):268–308.
- Boas, M. L. (1996). *Mathematical methods in the physical sciences, second edition*. John Wiley & Sons.
- Boguth, O. and Simutin, M. (2018). Leverage constraints and asset prices: Insights from mutual fund risk taking. *Journal of Financial Economics*, 127(2):325–341.
- Bookstaber, R., Foley, M. D., and Tivnan, B. F. (2016). Toward an understanding of market resilience: market liquidity and heterogeneity in the investor decision cycle. *Journal of Economic Interaction and Coordination*, 11:205–227.

- Bookstaber, R., Paddrik, M., and Tivnan, B. (2018). An agent-based model for financial vulnerability. *Journal of Economic Interaction and Coordination*, 13(2):433–466.
- Bookstaber, R. and Paddrik, M. E. (2015). An agent-based model for crisis liquidity dynamics. *OFR Working Paper*, 15(18).
- Bookstaber, R. M. (2012). Using agent-based models for analyzing threats to financial stability. *OFR. Available at SSRN 2642420*.
- Booth, A. (2016). *Automated algorithmic trading: machine learning and agent-based modelling in complex adaptive financial markets*. PhD thesis, University of Southampton.
- Bouchaud, J.-P. (2018). Agent-based models for market impact and volatility. In *Handbook of computational economics*, volume 4, pages 393–436. Elsevier.
- Bouchaud, J.-P. (2019). Econophysics: still fringe after 30 years? *Europhysics News*, 50(1):24–27.
- Bouchaud, J.-P. and Challet, D. (2017). Why have asset price properties changed so little in 200 years. In *Econophysics and Sociophysics: Recent Progress and Future Directions*, pages 3–17. Springer.
- Bouchaud, J.-P., Mézard, M., Potters, M., et al. (2002). Statistical properties of stock order books: empirical results and models. *Quantitative finance*, 2(4):251–256.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis : Forecasting and control*. John Wiley & Sons, New York.
- Braouezec, Y. and Wagalath, L. (2018). Risk-based capital requirements and optimal liquidation in a stress scenario. *Review of Finance*, 22(2):747–782.
- Braun-Munzinger, K., Liu, Z., and Turrell, A. E. (2018). An agent-based model of corporate bond trading. *Quantitative Finance*, 18(4):591–608.
- Braverman, A. and Minca, A. (2018). Networks of common asset holdings: aggregation and measures of vulnerability. *The Journal of Network Theory in Finance*, 4(3).
- Brede, M. and Sinha, S. (2005). Assortative mixing by degree makes a network more unstable. *arXiv preprint cond-mat/0507710*.

- Breedon, F., Chen, L., Ranaldo, A., and Vause, N. (2018). Judgement day: algorithmic trading around the Swiss franc cap removal. *Bank of England Staff Working Paper No. 711*.
- Brewer, P., Cvitanic, J., and Plott, C. R. (2013). Market microstructure design and flash crashes: A simulation approach. *Journal of Applied Economics*, 16(2):223–250.
- Broeke, G. t., Voorn, G. v., and Ligtenberg, A. (2016). Which sensitivity analysis method should I use for my agent-based model? *Journal of Artificial Societies and Social Simulation*, 19(1):5.
- Brogaard, J., Carrion, A., Moyaert, T., Riordan, R., Shkilko, A., and Sokolov, K. (2018). High frequency trading and extreme price movements. *Journal of Financial Economics*, 128(2):253–265.
- Brown, L. D., Cai, T. T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 16(2):101–133.
- Brown, S. J., Goetzmann, W., Ibbotson, R. G., and Ross, S. A. (1992). Survivorship bias in performance studies. *The Review of Financial Studies*, 5(4):553–580.
- Brunnermeier, M. K. and Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial studies*, 22(6):2201–2238.
- Bruno, S., Chincarini, L. B., and Davis, J. (2015). Crowded spaces and copycat risk management. *Available at SSRN 2616579*.
- Bryan, B., Nichol, R. C., Genovese, C. R., Schneider, J., Miller, C. J., and Wasserman, L. (2006). Active learning for identifying function threshold boundaries. In *Advances in neural information processing systems*, pages 163–170.
- Bucci, F., Mastromatteo, I., Eisler, Z., Lillo, F., Bouchaud, J.-P., and Lehalle, C.-A. (2019). Co-impact: Crowding effects in institutional trading activity. *Quantitative Finance*, Latest Articles:1–13.
- Buchanan, M. (2015). Trading at the speed of light. *Nature*, 518(7538):161–163.
- Budish, E. B., Cramton, P., and Shim, J. J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621.

- Caccioli, F., Barucca, P., and Kobayashi, T. (2018). Network models of financial systemic risk: a review. *Journal of Computational Social Science*, 1(1):81–114.
- Caccioli, F., Farmer, J. D., Foti, N., and Rockmore, D. (2015). Overlapping portfolios, contagion, and financial stability. *Journal of Economic Dynamics and Control*, 51:50–63.
- Caccioli, F., Shrestha, M., Moore, C., and Farmer, J. D. (2014). Stability analysis of financial contagion due to overlapping portfolios. *Journal of Banking & Finance*, 46:233–245.
- Cahan, R. and Luo, Y. (2013). Standing out from the crowd: Measuring crowding in quantitative strategies. *The Journal of Portfolio Management*, 39(4):14–23.
- Calcagnile, L. M., Bormetti, G., Treccani, M., Marmi, S., and Lillo, F. (2018). Collective synchronization and high frequency systemic instabilities in financial markets. *Quantitative Finance*, 18(2):237–247.
- Calvez, B. and Hutzler, G. (2006). Automatic tuning of agent-based models using genetic algorithms. In Sichman, J. S. and Antunes, L., editors, *Multi-Agent-Based Simulation VI*, pages 41–57, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Capponi, A. and Larsson, M. (2015). Price contagion through balance sheet linkages. *Review of Asset Pricing Studies*, 5(2):227–253.
- Cespa, G. and Foucault, T. (2014). Illiquidity contagion and liquidity crashes. *Review of Financial Studies*, 27(6):1615–1660.
- CFTC and SEC (2010). Findings regarding the market events of May 6, 2010. *Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*.
- Chen, C., Iyengar, G., and Moallemi, C. C. (2014). Asset-based contagion models for systemic risk. *Operations Research (forthcoming)*. Available at <https://www8.gsb.columbia.edu/researcharchive/articles/25467>, accessed 2017-10-30.
- Chen, S.-H., Chang, C.-L., and Du, Y.-R. (2012). Agent-based economic models and econometrics. *The Knowledge Engineering Review*, 27(02):187–219.
- Cheng, F. and Wellman, M. P. (2017). Accounting for strategic response in an agent-based model of financial regulation. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 187–203. ACM.

- Cheng, S., Hameed, A., Subrahmanyam, A., and Titman, S. (2017). Short-term reversals: The effects of past returns and institutional exits. *Journal of Financial and Quantitative Analysis*, 52(1):143–173.
- Chiarella, C., Iori, G., et al. (2002). A simulation analysis of the microstructure of double auction markets. *Quantitative finance*, 2(5):346–353.
- Chiarella, C., Iori, G., and Perelló, J. (2009). The impact of heterogeneous trading rules on the limit order book and order flows. *Journal of Economic Dynamics and Control*, 33(3):525–537.
- Čihák, M. (2007). Systemic loss: A measure of financial stability. *Czech Journal of Economics and Finance*, 57(1-2):5–26.
- Cimini, G., Squartini, T., Garlaschelli, D., and Gabrielli, A. (2015). Systemic risk analysis on reconstructed economic and financial networks. *Scientific reports*, 5(15758):1–12.
- Cliff, D., Bruten, J., et al. (1997). Zero not enough: On the lower limit of agent intelligence for continuous double auction markets. *HP Laboratories Technical Report HPL*.
- Cliff, D. and Northrop, L. (2012). The global financial markets: an ultra-large-scale systems perspective. In *Proceedings of the 17th Monterey conference on Large-Scale Complex IT Systems: development, operation and management*. Springer-Verlag.
- Colander, D. (2005). The future of economics: the appropriately educated in pursuit of the knowable. *Cambridge Journal of Economics*, 29(6):927–941.
- Collver, C. (2017). An application of agent-based modeling to market structure policy: the case of the U.S. tick size pilot program and market maker profitability. *SEC white paper*.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236.
- Cont, R. (2017). Central clearing and risk transformation. *Financial Stability Review (Banque de France)*.
- Cont, R. and Schaanning, E. (2019). Monitoring indirect contagion. *Journal of Banking & Finance*, 104:85–102.

- Cont, R. and Schaanning, E. F. (2017). Fire sales, indirect contagion and systemic stress testing. *Available at SSRN 2541114*.
- Cont, R. and Wagalath, L. (2013). Running for the exit: distressed selling and endogenous correlation in financial markets. *Mathematical Finance*, 23(4):718–741.
- Costa, L. d. F., Rodrigues, F. A., Travieso, G., and Villas Boas, P. R. (2007). Characterization of complex networks: A survey of measurements. *Advances in physics*, 56(1):167–242.
- Cristelli, M., Alfi, V., Pietronero, L., and Zaccaria, A. (2010). Liquidity crisis, granularity of the order book and price fluctuations. *The European Physical Journal B-Condensed Matter and Complex Systems*, 73(1):41–49.
- Crotty, J. (2009). Structural causes of the global financial crisis: a critical assessment of the new financial architecture. *Cambridge journal of economics*, 33(4):563–580.
- Da, Z., Gao, P., and Jagannathan, R. (2011). Impatient trading, liquidity provision, and stock selection by mutual funds. *The Review of Financial Studies*, 24(3):675–720.
- Dalko, V. (2016). Limit up–limit down: an effective response to the flash crash? *Journal of Financial Regulation and Compliance*, 24(4):420–429.
- Danielsson, J. and Zer, I. (2012). Systemic risk arising from computer based trading and connections to the empirical literature on systemic risk. *UK Government Office for Science, Foresight Driver Review-The Future of Computer Trading in Financial Markets*.
- Darley, V. and Outkin, A. V. (2007). *A NASDAQ market simulation: insights on a major market from the science of complex adaptive systems*, volume 1. World Scientific.
- Dechow, P. M., Hutton, A. P., Meulbroek, L., and Sloan, R. G. (2001). Short-sellers, fundamental analysis, and stock returns. *Journal of Financial Economics*, 61(1):77–106.
- Delpini, D., Battiston, S., Caldarelli, G., and Riccaboni, M. (2019). Systemic risk from investment similarities. *PloS one*, 14(5):e0217141.

- Di Gangi, D., Lillo, F., and Pirino, D. (2018). Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. *Journal of Economic Dynamics and Control*, 94:117–141.
- Duarte, F. and Eisenbach, T. M. (2018). Fire-sale spillovers and systemic risk. *FRB of New York Staff Report*, (645).
- Easley, D., de Prado, M. M. L., and O’Hara, M. (2012). Flow toxicity and liquidity in a high-frequency world. *Review of Financial Studies*, 25(5):1457–1493.
- Ebner, M., Shackleton, M., and Shipman, R. (2001). How neutral networks influence evolvability. *Complexity*, 7(2):19–33.
- Eisenberg, L. and Noe, T. H. (2001). Systemic risk in financial systems. *Management Science*, 47(2):236–249.
- Enriques, L., Romano, A., and Wetzter, T. (2019). Network-sensitive financial regulation. *The Journal of Corporation Law (forthcoming)*.
- Fabretti, A. (2013). On the problem of calibrating an agent based model for financial markets. *Journal of Economic Interaction and Coordination*, 8(2):277–293.
- Fagiolo, G., Guerini, M., Lamperti, F., Moneta, A., and Roventini, A. (2019). *Validation of Agent-Based Models in Economics and Finance*, pages 763–787. Springer International Publishing, Cham.
- Fagiolo, G., Moneta, A., and Windrum, P. (2007). A critical guide to empirical validation of agent-based models in economics: methodologies, procedures, and open problems. *Computational Economics*, 30(3):195–226.
- Fagiolo, G. and Roventini, A. (2017). Macroeconomic policy in DSGE and agent-based models redux: New developments and challenges ahead. *Journal of Artificial Societies and Social Simulation*, 20(1):1.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56.
- Farmer, J. D. and Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256):685–686.
- Farmer, J. D., Gillemot, L., Lillo, F., Mike, S., and Sen, A. (2004). What really causes large price changes? *Quantitative finance*, 4(4):383–397.

- Farmer, J. D., Patelli, P., and Zovko, I. I. (2005). The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6):2254–2259.
- Filimonov, V. and Sornette, D. (2012). Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108.
- Foucault, T., Kadan, O., and Kandel, E. (2005). Limit order book as a market for liquidity. *The Review of Financial Studies*, 18(4):1171–1217.
- Franke, R. and Westerhoff, F. (2012). Structural stochastic volatility in asset pricing dynamics: Estimation and model contest. *Journal of Economic Dynamics and Control*, 36(8):1193–1211.
- Fry, J. and Serbera, J.-P. (2017). Modelling and mitigation of flash crashes. *MPRA preprint 82457*.
- Gai, P., Haldane, A., and Kapadia, S. (2011). Complexity, concentration and contagion. *Journal of Monetary Economics*, 58(5):453–470.
- Gai, P. and Kapadia, S. (2010). Contagion in financial networks. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 466, pages 2401–2423. The Royal Society.
- García, D. and Norli, Ø. (2012). Crawling EDGAR. *The Spanish Review of Financial Economics*, 10(1):1–10.
- Gatti, D. D., Desiderio, S., Gaffeo, E., Cirillo, P., and Gallegati, M. (2011). *Macroeconomics from the Bottom-up*, volume 1. Springer Science & Business Media.
- Gatti, D. D., Fagiolo, G., Gallegati, M., Richiardi, M., and Russo, A. (2018). *Agent-based Models in Economics: A Toolkit*. Cambridge University Press.
- Geanakoplos, J., Axtell, R., Farmer, D. J., Howitt, P., Conlee, B., Goldstein, J., Hendrey, M., Palmer, N. M., and Yang, C.-Y. (2012). Getting at systemic risk via an agent-based model of the housing market. *The American Economic Review*, 102(3):53–58.
- Gerig, A. (2015). High-frequency trading synchronizes prices in financial markets. *Available at SSRN 2173247*.

- Gilli, M. and Winker, P. (2003). A global optimization heuristic for estimating agent based models. *Computational Statistics & Data Analysis*, 42(3):299–312.
- Glasserman, P. and Young, H. P. (2016). Contagion in financial networks. *Journal of Economic Literature*, 54(3):779–831.
- Gode, D. K. and Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of political economy*, pages 119–137.
- Goldberg, D. E. (2006). *Genetic algorithms*. Pearson Education India.
- Golub, A., Keane, J., and Poon, S.-H. (2012). High frequency trading and mini flash crashes. *arXiv preprint arXiv:1211.6667*.
- Gould, M. D., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., and Howison, S. D. (2013). Limit order books. *Quantitative Finance*, 13(11):1709–1742.
- Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of applied econometrics*, 8(S1):S85–S118.
- Gratiet, L. L., Marelli, S., and Sudret, B. (2016). Metamodel-based sensitivity analysis: polynomial chaos expansions and gaussian processes. *Handbook of Uncertainty Quantification*, pages 1–37.
- Grazzini, J. and Richiardi, M. (2015). Estimation of ergodic agent-based models by simulated minimum distance. *Journal of Economic Dynamics and Control*, 51:148 – 165.
- Grazzini, J., Richiardi, M. G., and Tsionas, M. (2017). Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control*, 77:26–47.
- Greenwood, R., Landier, A., and Thesmar, D. (2015). Vulnerable banks. *Journal of Financial Economics*, 115(3):471–485.
- Gualdi, S., Cimini, G., Primicerio, K., Di Clemente, R., Challet, D., et al. (2016). Statistically validated network of portfolio overlaps and systemic risk. *Scientific reports*, 6(39467):1–14.
- Gualdi, S., Tarzia, M., Zamponi, F., and Bouchaud, J.-P. (2015). Tipping points in macroeconomic agent-based models. *Journal of Economic Dynamics and Control*, 50:29–61.

- Gultekin, M. N. and Gultekin, N. B. (1983). Stock market seasonality: International evidence. *Journal of financial economics*, 12(4):469–481.
- Guo, W., Minca, A., and Wang, L. (2016). The topology of overlapping portfolio networks. *Statistics & Risk Modeling*, 33(3-4):139–155.
- Haldane, A. G. (2016). The dappled world. Speech given at GLS Shackle Biennial Memorial Lecture.
- Haldane, A. G. and May, R. M. (2011). Systemic risk in banking ecosystems. *Nature*, 469(7330):351–355.
- Haldane, A. G. and Turrell, A. E. (2018). An interdisciplinary model for macroeconomics. *Oxford Review of Economic Policy*, 34(1-2):219–251.
- Haldane, A. G. and Turrell, A. E. (2019). Drawing on different disciplines: macroeconomic agent-based models. *Journal of Evolutionary Economics*, 29(1):39–66.
- Hansen, N. and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2):159–195.
- Hanson, T. A. (2011). The effects of high frequency traders in a simulated market. In *Midwest Finance Association 2012 Annual Meetings Paper*.
- Hautsch, N. and Horvath, A. (2019). How effective are trading pauses? *Journal of Financial Economics*, 131(2):378–403.
- Hayek, F. A. (1945). The use of knowledge in society. *The American economic review*, 35(4):519–530.
- Hayes, R., Todd, A., Chaidarun, N., Tepsuporn, S., Beling, P., and Scherer, W. (2014). An agent-based financial simulation for use by researchers. In *Proceedings of the 2014 Winter Simulation Conference*, pages 300–309. IEEE Press.
- Helton, J. C. and Davis, F. J. (2002). Illustration of sampling-based methods for uncertainty and sensitivity analysis. *Risk Analysis*, 22(3):591–622.
- Herman, J. and Usher, W. (2017). SALib: An open-source python library for sensitivity analysis. *The Journal of Open Source Software*, 2(9):97.
- Holland, J. H. (1992). Genetic algorithms. *Scientific American*, 267(1):66–73.

- Hu, D., Schwabe, G., and Li, X. (2015). Systemic risk management and investment analysis with financial network analytics: research opportunities and challenges. *Financial Innovation*, 1(1):2.
- Huang, X., Vodenska, I., Havlin, S., and Stanley, H. E. (2013). Cascading failures in bi-partite graphs: model for systemic risk propagation. *Scientific Reports*, 3:12–19.
- Hué, S., Lucotte, Y., and Tokpavi, S. (2019). Measuring network systemic risk contributions: A leave-one-out approach. *Journal of Economic Dynamics and Control*, 100:86–114.
- Iskrev, N. (2019). What to expect when you’re calibrating: Measuring the effect of calibration on the estimation of macroeconomic models. *Journal of Economic Dynamics and Control*, 99:54–81.
- Jackson, R., Crowley-Reidy, L., and Schrimpf, A. (2017). The sterling ‘flash event’ of 7 October 2016. Bank for International Settlements <https://www.bis.org/publ/mktc09.pdf>. Accessed 2018-02-27.
- Jacob Leal, S. and Napoletano, M. (2019). Market stability vs. market resilience: Regulatory policies experiments in an agent-based model with low- and high-frequency trading. *Journal of Economic Behavior & Organization*, 157:15–41.
- Jacob Leal, S., Napoletano, M., Roventini, A., and Fagiolo, G. (2016). Rock around the clock: An agent-based model of low- and high-frequency trading. *Journal of Evolutionary Economics*, 26(1):49–76.
- Jacquier, E., Kane, A., and Marcus, A. J. (2003). Geometric or arithmetic mean: A reconsideration. *Financial Analysts Journal*, 59(6):46–53.
- Jain, P. K., Jain, P., and McInish, T. H. (2016). Does high-frequency trading increase systemic risk? *Journal of Financial Markets*, 31:1–24.
- Jiang, W. (2018). Leveraged speculators and asset prices. *Available at SSRN 2525986*.
- Johnson, N., Zhao, G., Hunsader, E., Meng, J., Ravindar, A., Carran, S., and Tivnan, B. (2012). Financial black swans driven by ultrafast machine ecology. *Available at SSRN 2003874*.
- Johnson, N., Zhao, G., Hunsader, E., Qi, H., Johnson, N., Meng, J., and Tivnan, B. (2013). Abrupt rise of new machine ecology beyond human response time. *Scientific Reports*, 3(2627).

- Johnson, N. F., Jefferies, P., Hui, P. M., et al. (2003). Financial market complexity. *OUN Catalogue*.
- Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, 47(2):263–291.
- Kalnina, I. (2011). Subsampling high frequency data. *Journal of Econometrics*, 161(2):262–283.
- Kårvik, G.-A., Noss, J., Worlidge, J., and Beale, D. (2018). The deeds of speed: an agent-based model of market liquidity and flash episodes. *Bank of England working paper*, (743). Accessed 2018-10-17.
- Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, volume 115, pages 700–721. The Royal Society.
- Khandani, A. and Lo, A. (2007). What happened to the quants in August 2007? *Journal of investment management*, 5(4):29–78.
- Kim, T.-H. and White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters*, 1(1):56–73.
- Kimbrough, S. O., Koehler, G. J., Lu, M., and Wood, D. H. (2008). On a feasible-infeasible two-population (FI-2Pop) genetic algorithm for constrained optimization: Distance tracing and no free lunch. *European Journal of Operational Research*, 190(2):310–327.
- Kirilenko, A., Kyle, A. S., Samadi, M., and Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3):967–998.
- Kirilenko, A. A., Kyle, A. S., Samadi, M., and Tuzun, T. (2011). The flash crash: The impact of high frequency trading on an electronic market. *Available at SSRN 1686004*.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.
- Kirman, A. and Teyssiere, G. (2002). Microeconomic models for long memory in the volatility of financial time series. *Studies in Nonlinear Dynamics & Econometrics*, 5(4).

- Kobayashi, S. and Hashimoto, T. (2011). Benefits and limits of circuit breaker: Institutional design using artificial futures market. *Evolutionary and Institutional Economics Review*, 7(2):355–372.
- Koijen, R. S. and Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, 127(4):1475–1515.
- Kononovicius, A. and Ruseckas, J. (2019). Order book model with herd behavior exhibiting long-range memory. *Physica A: Statistical Mechanics and its Applications*, 525:171–191.
- Korattikara, A., Chen, Y., and Welling, M. (2014). Austerity in MCMC land: Cutting the metropolis-hastings budget. In *International Conference on Machine Learning*, pages 181–189.
- Krapivsky, P. L., Redner, S., and Leyvraz, F. (2000). Connectivity of growing random networks. *Phys. Rev. Lett.*, 85:4629–4632.
- Krugman, P. (1991). Increasing returns and economic geography. *The Journal of Political Economy*, 99(3):483–499.
- Kukacka, J. and Barunik, J. (2017). Estimation of financial agent-based models with simulated maximum likelihood. *Journal of Economic Dynamics and Control*, 85:21 – 45.
- Lamperti, F. (2018). Empirical validation of simulated models through the GSL-div: an illustrative application. *Journal of Economic Interaction and Coordination*, 13(1):143–171.
- Lamperti, F., Roventini, A., and Sani, A. (2018). Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control*, 90:366–389.
- Langton, C. G. (1986). Studying artificial life with cellular automata. *Physica D: Nonlinear Phenomena*, 22(1-3):120–149.
- Leal, R. P. C. and de Melo Mendes, B. V. (2005). Maximum drawdown: Models and applications. *The Journal of Alternative Investments*, 7(4):83–91.
- LeBaron, B. (2000). Agent-based computational finance: Suggested readings and early research. *Journal of Economic Dynamics and Control*, 24(5):679–702.

- LeBaron, B. (2016). Financial price dynamics and agent-based models as inspired by Benoit Mandelbrot. *The European Physical Journal Special Topics*, 225(17–18):3243–3254.
- LeBaron, B. and Tesfatsion, L. (2008). Modeling macroeconomies as open-ended dynamic systems of interacting agents. *The American Economic Review*, 98(2):246–250.
- Lee, W. B., Cheng, S.-F., and Koh, A. (2011). Would price limits have made any difference to the ‘flash crash’ on May 6, 2010. *The Review of Futures Markets*, 19(special IFM issue):55–93.
- Levin, S. A. and Lo, A. W. (2015). Opinion: A new approach to financial regulation. *Proceedings of the National Academy of Sciences*, 112(41):12543–12544.
- Levine, Z. S., Hale, S. A., and Floridi, L. (2017). The October 2014 United States treasury bond flash crash and the contributory effect of mini flash crashes. *PloS one*, 12(11):e0186688.
- Lillo, F. and Pirino, D. (2015). The impact of systemic and illiquidity risk on financing with risky collateral. *Journal of Economic Dynamics and Control*, 50:180–202.
- Llacay, B. and Pepper, G. (2018). Using realistic trading strategies in an agent-based stock market model. *Computational and Mathematical Organization Theory*, 24(3):308–350.
- Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30(5):15–29.
- Lo, A. W. (2017). Moore’s law vs. Murphy’s law in the financial system: who’s winning? *Journal Of Investment Management*, 15(1):17–38.
- Lucas, R. (1976). Econometric policy evaluation: A critique. In *The Phillips Curve and Labor Markets*, pages 19–46. Amsterdam: North-Holland.
- Lux, T. (2018). Estimation of agent-based models using sequential Monte Carlo methods. *Journal of Economic Dynamics and Control*, 91:391–408. Special Issue in Honour of Prof. Carl Chiarella.
- Lux, T. and Zwinkels, R. C. (2018). Empirical validation of agent-based models. In *Handbook of Computational Economics*, volume 4, pages 437–488. Elsevier.

- Ma, J., He, J., Liu, X., and Wang, C. (2019). Diversification and systemic risk in the banking system. *Chaos, Solitons & Fractals*, 123:413–421.
- Madhavan, A. (2012). Exchange-traded funds, market structure, and the flash crash. *Financial Analysts Journal*, 68(4):20–35.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *The journal of business*, 36(4):394–419.
- Marino, S., Hogue, I. B., Ray, C. J., and Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology*, 254(1):178–196.
- Matsumoto, M. and Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30.
- McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, 57(5):995–1026.
- Menkveld, A. J. (2015). Crowded trades: an overlooked systemic risk for central clearing counterparties. In *AFA 2015 Boston Meetings*.
- Menkveld, A. J. (2016). The economics of high-frequency trading: Taking stock. *Annual Review of Financial Economics*, 8:1–24.
- Menkveld, A. J. and Yueshen, B. (2019). The flash crash: A cautionary tale about highly fragmented markets. *Management Science*, 65(10):4470–4488.
- Mo, S. Y. K., Paddrik, M., and Yang, S. Y. (2013). A study of dark pool trading using an agent-based model. In *2013 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, pages 19–26. IEEE.
- Nagel, S. (2012). Evaporating liquidity. *The Review of Financial Studies*, 25(7):2005–2039.
- Namatame, A. and Chen, S.-H. (2016). *Agent Based Modelling and Network Dynamics*. Oxford University Press.
- Nanex (2010). Nanex flash crash summary report. <https://bit.ly/2C3gvoo>, Accessed: 2017-04-18.

- Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313.
- Newman, M. (2018). Network structure from rich but noisy data. *Nature Physics*, 14(6):542.
- Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2):167–256.
- Nier, E., Yang, J., Yorulmazer, T., and Alentorn, A. (2007). Network models and financial stability. *Journal of Economic Dynamics and Control*, 31(6):2033–2060.
- Oriol, N. and Veryzhenko, I. (2019). Market structure or traders’ behavior? a multi agent model to assess flash crash phenomena and their regulation. *Quantitative Finance*, 19(7):1075–1092.
- Paddrik, M., Hayes, R., Todd, A., Yang, S., Beling, P., and Scherer, W. (2012). An agent based model of the e-mini S&P 500 applied to flash crash analysis. In *2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, pages 1–8. IEEE.
- Paddrik, M. E., Hayes, R., Beling, P., and Scherer, W. (2017). Effects of limit order book information level on market stability metrics. *Journal of Economic Interaction and Coordination*, 12(2):221–247.
- Panayi, E., Harman, M., and Wetherilt, A. (2012). Agent-based modelling of stock markets using existing order book data. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*, pages 101–114. Springer.
- Paulin, J., Calinescu, A., and Wooldridge, M. (2018). Agent-based modeling for complex financial systems. *IEEE Intelligent Systems*, 33(2):74–82.
- Paulin, J., Calinescu, A., and Wooldridge, M. (2019). Understanding flash crash contagion and systemic risk: A micro-macro agent-based approach. *Journal of Economic Dynamics and Control*, 100:200–229.
- Pele, D. T. and Mazurencu-Marinescu-Pele, M. (2019). Using high-frequency entropy to forecast bitcoin’s daily value at risk. *Entropy*, 21(2):102.
- Platt, D. (2019). A comparison of economic agent-based model calibration methods. *arXiv preprint arXiv:1902.05938*.

- Platt, D. and Gebbie, T. (2018). Can agent-based models probe market microstructure? *Physica A: Statistical Mechanics and its Applications*, 503:1092–1106.
- Poledna, S. and Thurner, S. (2016). Elimination of systemic risk in financial networks by means of a systemic risk transaction tax. *Quantitative Finance*, 16(10):1599–1613.
- Poledna, S., Thurner, S., Farmer, J. D., and Geanakoplos, J. (2014). Leverage-induced systemic risk under Basel II and other credit risk policies. *Journal of Banking & Finance*, 42:199–212.
- Raman, V., Robe, M., and Yadav, P. K. (2014). Man vs. machine: liquidity provision and market fragility. *Law Review*, 68:101–163.
- Recchioni, M. C., Tedeschi, G., and Gallegati, M. (2015). A calibration procedure for analyzing stock price dynamics in an agent-based framework. *Journal of Economic Dynamics and Control*, 60:1–25.
- Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535.
- Robert, C. P. and Casella, G. (1999). The Metropolis-Hastings algorithm. In *Monte Carlo Statistical Methods*, pages 231–283. Springer.
- Roseman, B. S., Ness, B. F. V., and Ness, R. A. V. (2018). Odd-lot trading in U.S. equities. *The Quarterly Review of Economics and Finance*, 69:125–133.
- Rosenbluth, G. (1955). Measures of concentration. In *Business concentration and price policy*, pages 57–99. Princeton University Press.
- Rozeff, M. S. and Kinney, W. R. (1976). Capital market seasonality: The case of stock returns. *Journal of Financial Economics*, 3(4):379–402.
- Salle, I. and Yildizoglu, M. (2014). Efficient sampling and meta-modeling for computational economic models. *Computational Economics*, 44(4):507–536.
- Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., and Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling and Software*, 114:29–39.

- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270.
- Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.
- Schmitt, N. and Westerhoff, F. (2014). Speculative behavior and the dynamics of interacting stock markets. *Journal of Economic Dynamics and Control*, 45:262–288.
- Schwarcz, S. L. (2008). Systemic risk. *Geo. LJ*, 97:193.
- Secchi, D. and Seri, R. (2017). Controlling for false negatives in agent-based models: a review of power analysis in organizational research. *Computational and Mathematical Organization Theory*, 23(1):94–121.
- Sergeyev, Y. D., Kvasov, D., and Mukhametzhanov, M. (2018). On the efficiency of nature-inspired metaheuristics in expensive global optimization with limited budget. *Scientific Reports*, 8(1):453.
- Serri, M., Caldarelli, G., and Cimini, G. (2017). How the interbank market becomes systemically dangerous: an agent-based network model of financial distress propagation. *Journal of Network Theory in Finance*, 3(1):1–15.
- Shleifer, A. and Vishny, R. (2011). Fire sales in finance and macroeconomics. *The Journal of Economic Perspectives*, 25(1):29–48.
- Sias, R., Turtle, H. J., and Zykaj, B. (2016). Hedge fund crowds and mispricing. *Management Science*, 62(3):764–784.
- Sifat, I. M. and Mohamad, A. (2019). Circuit breakers as market stability levers: A survey of research, praxis, and challenges. *International Journal of Finance & Economics*, 24(3):1130–1169.
- Sirignano, J. and Cont, R. (2019). Universal features of price formation in financial markets: perspectives from deep learning. *Quantitative Finance*, 19(9):1449–1459.
- Smith, V. L. (1962). An experimental study of competitive market behavior. *The Journal of Political Economy*, pages 111–137.

- Sobol', I. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1):271 – 280. The Second IMACS Seminar on Monte Carlo Methods.
- Solis, F. J. and Wets, R. J.-B. (1981). Minimization by random search techniques. *Mathematics of operations research*, 6(1):19–30.
- Sörensen, K. (2013). Metaheuristics — the metaphor exposed. *International Transactions in Operational Research*, 22(1):3–18.
- Squartini, T., Caldarelli, G., and Cimini, G. (2016). Stock markets reconstruction via entropy maximization driven by fitness and density. *arXiv preprint arXiv:1606.07684*.
- Stein, J. C. (2009). Presidential address: Sophisticated investors and market efficiency. *The Journal of Finance*, 64(4):1517–1548.
- Subrahmanyam, A. (2013). Algorithmic trading, the flash crash, and coordinated circuit breakers. *Borsa Istanbul Review*, 13(3):4–9.
- Tesfatsion, L. (2002). Agent-based computational economics: Growing economies from the bottom up. *Artificial life*, 8(1):55–82.
- Thurner, S., Farmer, J. D., and Geanakoplos, J. (2012). Leverage causes fat tails and clustered volatility. *Quantitative Finance*, 12(5):695–707.
- Torii, T., Izumi, K., and Yamada, K. (2015). Shock transfer by arbitrage trading: analysis using multi-asset artificial market. *Evolutionary and Institutional Economics Review*, 12(2):395–412.
- Tóth, B., Lemperiere, Y., Deremble, C., De Lataillade, J., Kockelkoren, J., and Bouchaud, J.-P. (2011). Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1(2):021006.
- Tsang, E. and Martinez-Jaramillo, S. (2004). Computational finance. *IEEE computational intelligence society newsletter*, 3(8).
- Turrell, A. (2016). Agent-based models: understanding the economy from the bottom up. *Bank of England Quarterly Bulletin*, page Q4.
- UK Government Office for Science (2012). The future of computer trading in financial markets. *Foresight Driver Review*.

- van der Hoog, S. (2019). Surrogate modelling in (and of) agent-based models: A prospectus. *Computational Economics*, 53(3):1245–1263.
- Van Kervel, V. and Menkveld, A. J. (2019). High-frequency trading around large institutional orders. *The Journal of Finance*, 74(3):1091–1137.
- Virgilio, G. P. M. (2020). When spread bites fast — volatility and wide bid-ask spread in a mixed high-frequency and low-frequency environment. *Research in International Business and Finance*, 51:101066.
- Vuorenmaa, T. A. and Wang, L. (2014). An agent-based model of the flash crash of May 6, 2010, with policy implications. *Available at SSRN 2336772*.
- Wah, E. and Wellman, M. P. (2016). Latency arbitrage in fragmented markets: A strategic agent-based analysis. *Algorithmic Finance*, 5(3-4):69–93.
- Wah, E., Wright, M., and Wellman, M. P. (2017). Welfare effects of market making in continuous double auctions. *Journal of Artificial Intelligence Research*, 59:613–650.
- Wellman, M. P. (2011). Trading agents. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 5(3):1–107.
- Wellman, M. P. (2016). Putting the agent in agent-based modeling. *Autonomous Agents and Multi-Agent Systems*, 30(6):1175–1189.
- Wellman, M. P. and Wah, E. (2017). Strategic agent-based modeling of financial markets. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 3(1):104–119.
- Wigglesworth, R. and McLannahan, B. (2018). Liquidity ousts leverage as the big market worry. *Financial Times*. <https://on.ft.com/33gwD25>. Accessed: 2018-09-19.
- Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212.
- Winker, P., Gilli, M., and Jeleskovic, V. (2007). An objective function for simulation based inference on exchange rate data. *Journal of Economic Interaction and Coordination*, 2(2):125–145.
- Wooldridge, M. and Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The knowledge engineering review*, 10(02):115–152.

- Wooldridge, M. J. (2001). *Multi-agent systems: an introduction*. Wiley Chichester.
- Xiong, X., Nan, D., Yang, Y., and Yongjie, Z. (2015). Study on market stability and price limit of chinese stock index futures market: An agent-based modeling perspective. *PloS one*, 10(11):e0141605.
- Yeh, C.-H. and Yang, C.-Y. (2010). Examining the effectiveness of price limits in an artificial stock market. *Journal of Economic Dynamics and Control*, 34(10):2089 – 2108.
- Zarinelli, E., Treccani, M., Farmer, J. D., and Lillo, F. (2015). Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate. *Market Microstructure and Liquidity*, 1(2):1550004.
- Zhou, Y. and Li, H. (2019). Asset diversification and systemic risk in the financial system. *Journal of Economic Interaction and Coordination*, 14(2):247–272.
- Zovko, I., Farmer, J. D., et al. (2002). The power of patience: a behavioural regularity in limit-order placement. *Quantitative Finance*, 2(5):387–392.