Mathematical Modelling, Scenario Simulation and Policy Analysis

Artur Kotlicki

St Hugh's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2021

To Alice, Aneta and Arek

Acknowledgements

Supported by the Mathematical Institute Scholarship, University of Oxford.

This thesis would not be complete without due thanks to the people who have made it possible.

First and foremost my deepest gratitude goes to Professor Rama Cont – you have been an absolute inspiration to work with. You have shown me a vast depth of expertise and a great passion for research with impact. I will be forever grateful for your invaluable guidance and for our many stimulating and fruitful discussions over the years, starting from my undergraduate project up until now – time flies!

A warm thank you goes to everyone I have had the pleasure to work with at the Bank of England. Special thanks go to Andrea Austin, David Humphry, and Alan Sheppard – thank you for the opportunity to intern at the Bank and for our travels through the world of Solvency II. I will fondly remember our discussions and of course, the Christmas parties! It has been a great pleasure to be part of the Insurance Policy Division, a truly welcoming team. A big thank you also to the fellow researchers at the Bank for their inspiring comments and invaluable feedback.

I extend my thanks to Laura Valderrama – it was a great pleasure to work with you on the *Liquidity at Risk* paper. Thank you for sharing your cutting-edge perspective on the world of financial supervision and for showing me the strong partnership between academia and supervision in action.

A heartfelt thanks to all those who reviewed and provided valuable comments on my work, both during seminars and in the run up to journal publication.

Thank you to my friends at the Mathematical and Computational Finance group at Oxford – the seminars were great and the discussions were too! A particular mention to my office buddies, Purba and Song: our coffee breaks, extensive mathematical discussions and chess games made every moment at the office special.

Warm thoughts go to Rama's squad, consisting of his PhD students both old and new – thanks for forming the great community that became my family away from home.

Thank you to my friends around the world, who have helped put the 'balance' into work-life. Thank you for the evenings of stimulating discussions spanning economics and world politics, and of course for the many board game nights!

Last but not least, I thank my family and partner whose love and care have been instrumental in my development, ultimately leading to the fulfilling path on which I have embarked. Thank you, for always being there with me on this journey, and for the many journeys yet to come.

I submit this thesis in loving memory of my grandfather, Ignacy Belta. *Dziękuję Ci Dziadku.*

Statement of Originality

- Chapter 1 is entirely my own work.
- Chapter 2 is based on the following published work: Rama Cont, Artur Kotlicki, and Laura Valderrama. Liquidity at risk: Joint stress testing of solvency and liquidity. *Journal of Banking & Finance*, 118: 105871, 2020. All co-authors contributed equally.
- Chapter 3 is based on: Artur Kotlicki, Andrea Austin, David Humphry, Hannah Burnett, Philip Ridgill, Sam Smith. Network Analysis of the UK Reinsurance Market. Staff Working Paper, Bank of England, (forthcoming), for which I am the first author.
- Chapter 4 is based on the following published work: Rama Cont, Artur Kotlicki, and Renyuan Xu. Modelling COVID-19 contagion: risk assessment and targeted mitigation policies. *Royal Society Open Science* 8: 201535, 2021. All co-authors contributed equally.

Abstract

Scenario simulation and stress testing have become indispensable tools for policy makers for the monitoring and management of risk in complex socio-economic systems with heterogeneous, interrelated components. This thesis contains several methodological contributions to this increasingly important field at the intersection of mathematical modelling and policy, with applications focusing on three specific areas: the stress testing of large financial institutions, the financial stability of the UK reinsurance sector and the analysis of control policies for the COVID-19 pandemic in England.

Chapter 2 proposes a new framework for the joint stress testing of liquidity and solvency risk for financial institutions. Rather than being specified independently from solvency shocks and applied in parallel to them, liquidity shocks are instead endogenously generated through mechanisms that model the liquidity-solvency nexus. The framework is applied to balance sheets of large financial institutions and provides interesting insights into the links between solvency and liquidity risk.

Chapter 3 proposes a network model for counterparty credit risk in the UK reinsurance market. A multi-layered network approach is used to incorporate information on reinsurance contract risk types and ownership structure for life and non-life insurers. The UK reinsurance sector is found to exhibit a 'small-world' property with a scale-free, core-periphery structure and topological characteristics common to other financial networks, making it 'robust-yet-fragile' to financial shocks. A stress simulation exercise shows the network to be robust to a shock to the value of total investments, and to idiosyncratic shocks to large, highly interconnected reinsurers.

Chapter 4 proposes a spatial epidemic model with demographic and geographic heterogeneity to study the regional dynamics of COVID-19 in England. The model provides a framework for assessing the impact of policies targeted towards sub-populations or regions. We define a concept of efficiency for comparative analysis of epidemic control policies and show targeted mitigation policies based on local monitoring to be more efficient than country-level or non-targeted measures.

Contents

List of Figures

Chapter 1

Overview

This chapter provides an overview of the motivation and themes underlying the thesis, as well as its key contributions. Detailed literature reviews and discussion can be found in each of the following chapters.

1.1 Scenario Simulation: a Tool for Risk Management

The last two decades have witnessed spectacular examples of economic and social crises that have posed major challenges to policymakers, most notably the 2008 financial crisis, the failure of large financial institutions across the United States and Europe and, most recently, the COVID-19 pandemic. A common feature of these crises has been the importance of *contagion* and risk amplification mechanisms, as well as the associated *systemic risk*, and large scale instabilities, which have focused the attention of policy makers due to the associated large scale mortality risk, and large scale financial and economic losses.

Faced with such challenges, policymakers have had to devise strategies for monitoring and controlling the systemic risk that may arise from such crisis situations. A common feature across these episodes has been the widespread adoption of *mathematical models* by policymakers, both in the area of financial regulation and in public health. The use of mathematical models is indispensable when examining policy for the management of risks that may not have occurred in the past, and thus for which the sole examination of historical data may not be sufficient. This

is especially clear in light of the 2008 financial crisis, the Greece credit crisis, and the COVID-19 pandemic.

On the other hand, given the complexity and heterogeneity of the underlying systems, it has also been progressively recognised that simple 'homogeneous' models – which may be useful to make a point about risk management or epidemic dynamics in a textbook setting – may fail to capture adequately the impact of various policies in a realistic manner. The alternative, which has been increasingly embraced by practitioners of risk management in various areas, has been to develop more granular – and more realistic – heterogeneous models incorporating in particular the contagion mechanisms at play. Network models, which provide a natural theoretical framework for modelling the interconnectedness of components in heterogeneous systems, have provided useful insights for studying such systems (Acemoglu et al., 2015; Cont et al., 2013; Keeling and Eames, 2005).

In consequence, large scale heterogeneous stochastic models have been developed and adopted, among other fields, in financial risk management and epidemiology, and have been used by policy makers for designing new regulations and intervention policies. For example, during the COVID-19 pandemic it was quickly recognised among experts that simple homogeneous epidemic models such as the SIR or SEIR models, deployed at the level of a country, while illustrating certain qualititative points, do not yield realistic results, and that geographic and demographic heterogeneities need to be incorporated (Thomas et al., 2020; Britton et al., 2020; Hebert-Dufresne et al., 2020).

In the field of financial risk management, the 2008 crisis led to a renewed interest in more realistic risk models, in particular with the incorporation of institutional details such as funding costs, collateral mechanisms, counterparty risk, liquidity risk and the resulting contagion mechanisms, leading to a more granular view of the balance sheets of financial institutions and their risk exposures.

However, the price to pay for more realistic models is increasing model complexity and the corresponding loss of analytical tractability. Unlike simple low-dimensional toy models that may be studied in detail by analytical means, heterogeneous stochastic models lead to complex dynamics and make the impact of various regulations and policies less transparent, requiring a systematic simulation-based approach for the quantitative analysis of the impact of various policies.

This has led to the systematic use of *scenario simulation* techniques for analysing the behaviour of such systems. Simulation-based approaches for financial risk management have been in use since the early 2000s with the advent of model-based approaches for bank capital adequacy in the Basel II framework, and the concurrent availability of software tools for stochastic scenario simulation. Model-based scenario simulation has the benefit of exploring scenarios which may not have occurred in the past and which may constitute robustness tests for policies designed to manage potential risks. This approach is now adopted by most large financial institutions, which make extensive use of scenario simulation for the assessment and monitoring of their risk exposures. The underlying question is then the design of such models and whether they capture the underlying risks correctly.

Given the model risk underlying such model-based approaches, their use was complemented early on by *stress testing*, an approach in which historical or model-based scenarios are complemented with 'adverse scenarios' chosen by experts and which may correspond to the anticipations of policymakers. This is the approach adopted for instance by the International Monetary Fund (IMF) in its periodic Financial Stability Assessment Programmes (International Monetary Fund, 2019) and by most central banks and financial regulators since 2008 for monitoring the stability of financial institutions. Regulators select one of two 'adverse scenarios' and examine the adequacy of capital and liquidity reserves of financial institutions in such scenarios. The validity of this approach crucially depends on the choice of a realistic 'extreme but plausible' stress scenario. The design of such stress scenarios is a challenging task especially if the underlying system is exposed to many risk factors and has many interrelated components: this is the case for example for the balance sheet of large financial institutions, and for the financial system as a whole.

1.2 Scenario Simulation and Stress Testing for Complex Heterogeneous Systems

One of the striking features of the 2008 financial crisis was the sudden drying up of liquidity in the market for short-term debt rollover. However, it was not

until August 9, 2007 that the full effects of the system-wide financial crisis of the 2006 sub-prime market meltdown have been recognised fully (Gale, 2015). In particular, it was the announcement from BNP-Paribas on August 9, 2007 of a temporary suspension on redemptions by three of its hedge funds that led to a panic in the Euro interbank market. As a response to the sudden spike in liquidity demand and in an attempt to stabilise the financial system, the European Central Bank (ECB) injected 95 billion Euro in overnight lending on August 9, 2007, and a further 150 billion Euro injection followed a week later (European Central Bank, 2009a).

In the aftermath of seven consecutive credit events happening to financial entities within a month of each other in 2008,[1] a discussion on the underlying cause of the crisis emerged in literature. The interplay between bank fundamentals and default events have been extensively studied, where liquidity shortages have been shown to be a major contributor to the crisis (Morris and Shin, 2008; Pierret, 2015).

In particular, following an adverse shock to the system a complex series of adjustments in prices and quantities are triggered, which are required for the market to clear in the new context (Gale, 2015). Such an interaction of insolvency and illiquidity risk can lead to a systemic event, where an amplification of the initial perturbation in the system leads to incurred losses that may differ from the initial impact significantly.

The potential for systemic events to create widespread disorder and instability leading to impairments in both economic growth and social welfare makes them prime candidates for modelling and study. It is important to understand the scenarios and conditions under which they can occur, and the network characteristics which make a system particularly fragile to systemic risk. Sources of vulnerability can then be mitigated, and guided through the implementation of policy. Indeed, systemic events have been a focal topic of study both before and after the 2008 crisis (for example, De Brandt and Hartmann (2000); Hellwig (2009); Haldane and May (2011); Cont et al. (2013); Brownlees and Engle (2017)) and can be

[1]From September 7, 2008 to October 8, 2008 there were seven credit events involving major financial institutions: Fannie Mae, Freddie Mac, Lehman Brothers, Washington Mutual, Glitnir, Kaupthing, and Landsbanki (Brigo et al., 2010).

characterised in a multitude of ways, whether through idiosyncratic or systematic factors, exogenous or endogenous triggers, and with sequential or simultaneous impacts (European Central Bank, 2009b). The multi-faceted nature of these events calls for complex models that can capture adequately the underlying interactions to not only inform supervisors and government bodies in policy making, but also to gauge the effectiveness of policies aimed at curbing its spread. The European Central Bank (2009b) categorised three (not mutually exclusive) shock scenarios that can lead to the materialisation of systemic events which we will reference to frame the onward discussion.

Shock from the Unfolding of Systemic Imbalances
Endogenous trigger as a result of the unravelling of imbalances in the financial system which have built up over time. This unravelling may affect multiple intermediaries or markets at the same time.

A prominent example from the financial crisis of an imbalance to have built up over time is the over reliance of financial institutions on short-term funding, which when unravelled during a credit panic subjects them to a higher risk of failure (Rodrik and Velasco, 1999; Shin, 2009; Fahlenbrach et al., 2012). In particular, Gorton and Metrick (2012) argue that in the case of the recent financial crisis it was the fear of market freezes extending to other non-subprime related asset-backed securities that resulted in a public shock causing a significant increase in expected future spread volatility, and subsequently leading to increases in the repurchase agreement haircuts – the result of which is considered tantamount to a run on the financial system, and illustrates the amplifying impact of liquidity constraints on solvency shocks.

The fundamental maturity transformation role of financial institutions subjects them to liquidity risk related to the loss of short-term funding. Beginning with Diamond and Dybvig (1983), literature on bank runs made various attempts to explain the circumstances under which financial institutions can experience sudden liquidity withdrawals that lead to their failure. The classical setting of Diamond and Dybvig (1983) considers a rollover decision of depositors endowed with uncertainty about the timing of their consumption needs. As these consumption needs

are typically uncorrelated, the bank expects – and thus provisions for – a relatively small number of withdrawals in the short-term, and uses the remaining funds to finance long-term investments. Such an inherent maturity mismatch creates the possibility of a run in the model. This is because each depositor's incentive to withdraw depends on the anticipated action of others. In particular, when each agent expects other agents to withdraw early, they too will try to withdraw early as a result of panic stemming from an anticipated bank failure. This leads to an interesting feature of the model, in which multiple Nash equilibria exist, with at least one desirable and one undesirable equilibrium. A desirable equilibrium arises when the confidence is maintained among depositors, and withdrawals are only made by depositors with a current consumption need. On the other hand, an undesirable equilibrium relates to a bank run, which arises as a result of a self-fulfilling creditor panic. In such a case, a solvent financial institution is forced to unexpectedly liquidate its assets early leading a consequent failure.

However, it can also be argued that creditor panics are not irrational events but rather a response to an adverse signal about the economy (Gorton, 2012). In light of this, the bank run model of Gorton (1985) presents a contrasting view to Diamond and Dybvig (1983), which relates the creditor panic to the observed information about the bank's fundamentals. This information-based explanation of panic is founded on rational fears of capital losses. Without an underlying signal about deterioration of the bank's investments, optimal portfolio allocations are argued to provide sufficient incentive for depositors not to withdraw early. In a similar vein, Allen and Gale (1998) relate bank runs to business cycles, which change perception of depositors' risk. An economic downturn increases the possibility that a bank will be unable to meet their commitments, and as such banking panics are argued to be an inevitable consequence of the standard unsecured deposit contract. As detailed in Gorton (1988), empirical evidence from banking panics in the eighteenth and nineteenth century corroborates this view. Using the change in pig iron production as an indicator, the five worst recessions are shown to be always accompanied by panics.

In contrast to the early models of bank runs that put emphasis on maturity transformation and insurance of small depositors, Rochet and Vives (2004) consider a more modern view of bank runs, posed as a debt rollover problem of large

and well-informed investors in the interbank market. The decision about the renewal of credit is then related to an event that casts doubt about the repayment capacity of the financial institution. By letting creditors observe a private noisy signal about the bank's fundamentals, Rochet and Vives (2004) introduce a strategic uncertainty about the actions of other creditors. This approach, inspired by the methodology of global games proposed by Morris and Shin (1998),[2] addresses the main caveat of the model of Diamond and Dybvig (1983), where in the presence of multiple equilibria it is impossible to clearly identify the factors that affect the probability of a run. Goldstein and Pauzner (2005) use an analogous approach in which the fundamentals of the economy uniquely determine whether a bank run occurs. Consequently, the authors find that the probability of a run increases with the short-term payment on the deposits offered by the bank.

Although fundamental weaknesses may indeed exacerbate illiquidity problems, the converse is also true. The theoretical framework of Diamond and Rajan (2005) examines how liquidity shortages and solvency problems can interact in the context of a simple financial system. In particular, the anticipated insolvency of a financial institution can trigger a bank run, in which depositors withdraw liquid funds immediately. As such, this can force banks to foreclose on loans that otherwise would generate liquidity in the future. In consequence, this shrinks the common liquidity pool, and thus exacerbates the aggregate liquidity shortage in the economy. These system-wide spillover effects can then lead to a contagion of failures, which can be detrimental to the entire financial system. Despite highlighting the important interplay of solvency and liquidity, the model of Diamond and Rajan (2005) fails to capture adequately the various mechanisms through which solvency and liquidity can interact in a modern financial system. For example, the model does not consider liquidation costs caused by fire sales, the existence of securitised funding markets, nor the context of a modern regulation that introduces variation margin requirements.

Empirical research sheds more light on some of these key mechanisms through which liquidity and solvency interact in practice. In particular, by examining the

[2]Global game methodology in a general setting was firstly introduced by Carlsson and van Damme (1993). See also subsequent developments in Morris and Shin (2003).

short-term balance sheets of fifty US financial institutions over the period of 2000–2013, Pierret (2015) shows that the amount of short-term funding available to a bank becomes limited with an increase in its solvency risk. Conversely, financial institutions with a higher proportion of short-term debt are more vulnerable to insolvency. The former observation is corroborated by the study of Du et al. (2019) on counterparty risk management practices in the credit default swap (CDS) market during the post-crisis period. The authors show that although counterparty risk is found to have a minor impact on the pricing of contracts, it has a significant effect on the client's choice of dealer counterparties. In light of this, solvency problems can significantly limit the volume of new short-term funding available to the financial institution, and thus exacerbate its liquidity distress. Indeed, this view is further supported by Blickle et al. (2019) who study the evidence from the the German Crisis of 1931. In particular, the authors show that the system-wide bank run is centered on the withdrawal of interbank funding by large institutional investors. Despite the absence of deposit insurance, regular retail depositors are seen to withdraw only after the information on deterioration of bank balance sheet characteristics, such as its equity and liquidity positions, have became public.

The evidence from past financial crises demonstrates the strong interactions between the solvency and liquidity risks that need to be taken into account by policymakers. Although theoretical considerations provide a useful framework from which policy decisions for financial stability can be made, most existing models tend to focus on liquidity or solvency and not the interaction between the two. A notable exception to this is the theoretical framework of Morris and Shin (2016) that emphasises the importance of modelling the liquidity-solvency nexus. In particular, the authors draw on ideas from the classical literature on bank run models to underline the importance of the illiquidity component of credit risk. As exemplified by the demise of Bear Stearns in March 2008, liquidity shortfall – created or exacerbated by a bank run – is a typical route to failure for financial institutions.[3] However, as liquidity and solvency are tightly linked, it is often impossible to distinguish in practice whether a run is the cause of failure for an otherwise solvent bank, or if it merely precipitated the default of an already troubled institution.

[3]We refer to Duffie (2010); Gorton (2012); Bernanke (2013); Geithner (2014) for a general discussion in this context.

Morris and Shin (2016) thus attempt to address this issue by means of a tractable theoretical framework that incorporates the probability of a run when pricing total credit risk. As emphasised by the authors, the inclusion of run probability is paramount to the pricing of total credit risk, since its occurrence not only undermines the debt value but also affects the recovery values through disorderly liquidation in a fire sale. In particular, Morris and Shin (2016) consider a leveraged financial institution that funds its investment in risky assets through a mix of short-term debt, which subjects the bank to rollover risk, as well as long-term debt and equity. Notably, the authors use global games methodology to establish a unique run equilibrium in their framework, and as such the rollover risk depends on the expected return of the creditors, which is influenced by the bank's fundamentals and its liquidity holdings. Another important factor affecting the run probability is the 'outside option' that defines the opportunity cost faced by short-term creditors. Consequently, Morris and Shin (2016) show that the total credit risk can be decomposed into two components. The first component is the insolvency risk that stems from the uncertainty in the eventual asset value realisation, and the second being the illiquidity risk. An important aspect of the model is that the illiquidity component relates to not only the run risk, defined as the probability of default due to a loss of short-term funding in the case of an otherwise solvent bank, but also the fire sale risk that relates to a failure caused by losses suffered from the excessive cost of disorderly asset liquidation.

The model of Morris and Shin (2016) provides a useful analytical framework that can inform policy analysis. In particular, the authors show that the illiquidity risk of a financial institution depends on the 'liquidity ratio', defined as the amount of available liquidity divided by the total value of short-term debt. We remark that the notion of the amount of available liquidity encompasses both the current liquid holdings of the bank as well as the discounted value of the risky asset, representing the amount of liquidity that can be realised in a fire sale. In consequence, the framework of Morris and Shin (2016) is useful in quantifying the impact on the total credit risk stemming from the shift in balance sheet composition from safe but low-yielding cash holdings to high-yielding risky assets. During the period of greater fundamental uncertainty (that is, when returns of the risky assets are

highly volatile), reducing the proportion of cash held increases the total credit risk significantly.

Morris and Shin (2016) highlight vulnerabilities introduced with excess reliance on short-term funding, which can be quantified using their liquidity ratio measure. This concept of the liquidity ratio is analogous to the Liquidity Coverage Ratio (LCR), introduced as part of the Basel III regulatory response to the 2008 financial crisis. Together with the Net Stable Funding Ratio (NSFR), LCR aims to create a substantial liquidity buffer to reduce the over-reliance of banks on unstable short-term funding (Schmieder et al., 2012). However, a notable drawback of these measures is that both LCR and NSFR are estimated based on historical data, and as such they are backward looking measures whose usefulness depend on the statistical assumptions about future stress scenarios. In Chapter 2 we introduce instead a new[4] forward looking measure called *Liquidity at Risk* that quantifies the expected net outflow taking into account a given scenario of the evolution of liquid balances and maturing liabilities. The scenario is model agnostic and for example may be generated from a stochastic model or be based on historical scenarios.

Furthermore, in Chapter 2 we capture the inextricable link between solvency and liquidity in our stress test framework through the combination of endogenous liquidity shocks as a result of solvency shocks, and the amplification of solvency shocks through funding costs arising from liquidity constraints. Our method allows us to distinguish between cases that are insolvent but liquid, and illiquid but solvent – areas which are not captured in typical stress tests that focus mainly on solvency.

Our work has significant policy implications. We demonstrate that the risk of failure can be significantly underestimated if considering solvency and liquidity in independent channels, or with liquidity as an add-on to solvency stress tests (as is the case with current practices). The framework also enables supervisors to identify sources of systemic spillover and make better quantification of emergency

[4]Incidentally, Liquidity at Risk was previously defined by Conzen (2009) as the maximum net liquidity drain relative to the expected liquidity buffer that should not be exceeded at a given confidence level. This definition is analogous to the statistical Value at Risk measure, and as such is affected by the same drawbacks: that is, the inability to capture tail risks, and its reliance on bank specific views on risk that may not be aligned at the systemic level (Basel Committee on Banking Supervision, 2012).

liquidity assistance requirements. Adoption effort for our framework is low, as it can be readily implemented through leveraging data already available to supervisors. Furthermore, the framework is comprehensively implemented and made available online to aid custom analysis at:

`http://liquidityatrisk.kotlicki.pl/`

Idiosyncratic Shocks

Exogenous idiosyncratic trigger becoming more widespread (typically sequentially).

Idiosyncratic shocks give rise to contagion risk, which can be categorised into direct and indirect channels (De Brandt and Hartmann, 2000). The former is a result of contractual agreements that are not fulfilled following an adverse shock, whereby one firm's failure to meet its obligations triggers distress on its counterparties, in turn leading to their consequent failure (the domino effect). On the other hand, indirect contagion covers cases where no such contractual agreements exist. For example, indirect contagion via the market-price channel (fire sale externality) stems from unorderly asset liquidation that goes on to impact the solvency of other parties with correlated exposures, which can potentially trigger a spiral of further fire sales. In addition, the resultant stress on the financial system can be amplified in the presence of margin calls, as demonstrated in Chapter 2. Another channel of indirect contagion is the information channel, triggered when bad news about a firm leads to hedging behaviour towards direct counterparties of the firm in question, or towards other firms with similar business or operating models (Clerc et al., 2016).

Although fire sales are extensively studied in literature (Shleifer and Vishny, 1992, 2010; Cont and Wagalath, 2016; Cont and Schaanning, 2017), most tend to focus on a homogeneous financial market with little attention given to the wide variety of market participants. A study by Timmer (2018) shows that insurance companies and pension funds are counter-cyclical investors – they buy when prices are low (such as during a fire sale) and sell when prices are high. They are able to invest in this way due to the relative stability of the liabilities on their balance

sheet which makes them better at absorbing short-term losses. In contrast, banks and investment funds are shown to be pro-cyclical investors. Insurers thus play an important role of liquidity provision in times of crisis, and in general contribute to financial stability through diversification of idiosyncratic risk for both individuals and institutions. Insurers can however be subject to contagion – not only through indirect channels (for example, stemming from common exposures with other financial institutions), but also through direct ones (for example, when reinsurers fail and reinsurance contracts can no longer be met). This makes the insurance sector an interesting area for study given the vital role insurers play during periods of stress. Network analysis tools have proven to be extremely useful for studying the dynamics of direct contagion in financial systems (Boss et al., 2004; Amini et al., 2012; Cont et al., 2013).

Early research by Allen and Gale (2000) provides microeconomic foundations for the analysis of contagion through direct linkages in financial systems. In their setting, consumers have the liquidity preferences as introduced by Diamond and Dybvig (1983), and banks perfectly insure against liquidity shocks by exchanging interbank deposits, which in turn exposes them to contagion. By studying different network structures involving four banks, Allen and Gale (2000) show that the spread of contagion depends strongly on the characteristics of interconnectedness among banks. Complete network structures, with all banks having evenly spread exposures to each other, are found to be the most robust. On the other hand, incomplete systems are found to be more fragile: concentration of losses in neighbouring regions prompts premature liquidation of illiquid assets, triggering loss spillovers to other, initially unaffected banks. However, insights based on simple and rigid network structures may fail to capture real-world contagion dynamics and accurate magnitude of losses.

In presence of heterogeneity among financial institutions – such as varying exposures or a complex network topology – the spread of contagion may be non-linear and highly sensitive to small changes in model parameters. Therefore, past adverse shocks to the system may not be indicative of the future resilience of the system. This point is emphasised in the study of Gai and Kapadia (2010), where the authors develop a network model of contagion in complex financial networks with arbitrary

structure. In particular, only the degree distribution is fixed, while other topological characteristics of the network are assumed to form randomly. Under these conditions, the authors argue that financial networks exhibit the *robust-yet-fragile* property, whereby the probability of contagion is low, but the potential impact once contagion occurs is high. In other words, below a tipping point, connections serve as a shock-absorber – disturbances are dispersed and attenuated within the system, and risk-sharing and diversification practices strengthen the robustness of the system. Beyond this tipping point, however, interconnectedness aids shock amplification and helps the contagion to spread across the entire financial system.

Caccioli et al. (2012) extend the above analysis of Gai and Kapadia (2010) (that makes arbitrary network structure assumptions) to account for empirically observed properties of financial networks, such as the effect of power law distributions in degree[5] and balance sheet size. Heterogeneous networks are shown to be more robust with respect to a failure of random banks but more fragile with respect to a failure of the highly connected hubs. These findings are corroborated by Acemoglu et al. (2015) who investigate how network topology affects the phase transition in financial contagion: densely connected networks with a higher level of diversification of exposures are shown to enhance stability under sufficiently small shocks. However, beyond a certain threshold in shock size, higher interconnectedness facilitates contagion and renders the financial system more fragile.

That said, the relation between network topology and its robustness is intricate. In particular, Roukny et al. (2013) compare scale-free networks with more homogeneous random networks to analyse default cascades under various circumstances. By varying several key characteristics, such as the initial adverse shock, degree distribution and market liquidity, they show that the network topology alone does not determine the stability of the system – scale-free networks can be both more robust and more fragile depending on initial endowments and in- and out-degree correlations. The impact of various variables – such as the connectivity, concentrations, and capital levels – on the global level of systemic risk has been examined also in earlier studies by Nier et al. (2007) and Battiston et al. (2012a) on simulated network structures. However, their main results related to

[5]A network whose degree distribution $P(k)$ follows a power law, that is $P(k) \sim k^{-\gamma}$, is known as a *scale-free* network.

the contagion and the roles of interconnectedness are strongly dependent on the exact network structure and details of the model. It is therefore unclear if these conclusions would hold in actual financial networks.

In order to capture the nuances of real-world systems, a significant emphasis has recently been placed on simulation-based approaches to contagion in financial networks using central bank data. Elsinger et al. (2006) for example asses contagion risk at the banking system level by applying a range of macroeconomic risk factors to the network as a whole. They find that correlated portfolios are the main source of direct contagion. Low bankruptcy costs and effective crisis resolution play a critical role in reducing the losses from contagious defaults.

Financial systems are far from random, and are observed to exhibit common characteristics across different countries and time periods. In the early study of Boss et al. (2004) on the network structure of the Austrian interbank market, the authors find that node degree and contract size are highly skewed and best described by power law distributions. Cluster analysis reveals further heterogeneity of banking sectors within the Austrian banking system. These findings are corroborated by Caccioli et al. (2015), who consider multiple contagion channels – counterparty risk and overlapping portfolios – based on information on the balance sheet of Austrian banks over the 2006–2008 time period. Although neither channel of contagion results in large effects on their own, bankruptcies are much more common and have large systemic effects in the presence of both channels. A strong heterogeneity of degrees and interbank exposures is also highlighted in the study of Cont et al. (2013) on the Brazil banking system during the 2008–2009 time period.

A scale-free network topology is shown by Barabási and Albert (1999) to emerge as a result of preferential attachment, where new nodes in the system are more likely to connect with already highly connected counterparties. Such non-random formation of links often leads to a core-periphery structure in financial networks (Craig and von Peter, 2014; Caccioli et al., 2018). Core nodes form a sub-graph of densely connected nodes, while the remaining nodes in the periphery exhibit a preference in choosing their counterparties: they typically connect with the core but not with other peripheral nodes. Consequently, core nodes are often found to be systemically important, as their failure facilitates the spread of contagion to

the other parts of the financial system. Craig and von Peter (2014) find support for a core-periphery structure in the German interbank network. They utilise Bundesbank statistics of more than 2000 banks and find evidence of intermediation between banks in the network. Banks typically deal with each other bilaterally in a decentralised manner, with intermediary banks extending credit between banks that are not linked. The authors model the resulting tiered structure via a core-periphery model. Fricke and Lux (2015) also find a core-periphery structure in the Italian interbank network by using the overnight interbank transactions over the period of 1999–2000. The authors find the role of banks in both the core and periphery are persistent over time, as well as signs of preferential lending which supports the view of the non-random network structure.

Despite sharing common topological network characteristics with interbank markets, little evidence is found on systemic vulnerabilities of the insurance markets in the existing, albeit limited, literature. Notably, the scenario-based analysis of Lelyveld et al. (2011) observes no default cascade within the Dutch reinsurance market over the period of 2003–2005. Similarly, Kanno (2016) finds limited contagion within the global non-life insurance market over the period of 2006–2013. Both Park and Xie (2014) and Chen et al. (2020) consider the data on US property-causality reinsurance market over the period of 2003–2009, and find the system to be robust even to extreme stress scenarios. These findings are in stark contrast to banking systems, which are often found to be vulnerable to system-wide contagion in more severe scenarios.

Modelling of a direct contagion on a financial network often relies on the notion of clearing to compute the allocation of debtor's assets among creditors. In particular, following an adverse scenario, a defaulted financial institution is no longer able to meet its contractual obligations in full, consequently generating losses to its counterparties. The incurred losses by the counterparties may in turn result in their default, leading to a wide-spread cascade of losses. Furfine (2003) presents an iterative procedure for modelling loss contagion, in which the losses are allocated proportionally among the creditors of a defaulted firm and then new defaults in the network are computed. However, this approach does not recognise that the higher order defaults increase losses at the firms that have failed in the previous iteration step. In particular, computing the number of banks that will default due

to contagion is non-trivial, since each consecutive default may affect the value of assets of already defaulted firms, and hence the respective loss given default (Upper, 2011). Eisenberg and Noe (2001) solve this simultaneity problem by providing a framework to compute a market clearing equilibrium through defining a clearing payment vector – that is, a vector that associates to each bank the total amount that it is able to repay to its debtors. However, this approach is criticised by Cont et al. (2013) as unsuitable in the context of modelling bankruptcies on the network: the framework corresponds to a hypothetical and unrealistic situation where all bank portfolios are simultaneously liquidated resulting in an endogenous recovery rate. It is also unrealistic to assume absence of additional costs related to a default. In light of this, Rogers and Veraart (2013) provide an extension to the Eisenberg and Noe (2001) framework by introducing default costs in the system. Consequently, their clearing procedure does not simply redistribute the initial losses within the system, but instead introduces additional costs to the economy that are related to the financial contagion. This procedure is utilised in the stress test we apply to the UK insurance market, taking into account the priority of debt claims and bankruptcy costs, which may arise from inefficiencies in realising asset values (see Chapter 3). In their recent work, Klages-Mundt and Minca (2020) develop a model for contagion in reinsurance markets by which primary insurers' losses are spread through the network. The framework accounts for more general forms of exposures, such as insurance contracts, where liabilities are not known in advance and exhibit complex interrelations and non-linearities. The model, however, relies on more granular information on reinsurance contracts, which often is unknown or at least highly uncertain in practice.

The heterogeneity demonstrated in the various empirical studies (Chen et al., 2020; Cont et al., 2013; Caccioli et al., 2015; Boss et al., 2004) supports the need for simulation-based methods to verify the susceptibility to contagion of specific financial networks. In light of this, in Chapter 3 we investigate the resilience of the UK insurance market, focusing on direct contagion through exposures via reinsurance contracts. We consider the effect of varying bankruptcy costs, and both the contagion impact of a network wide shock and the impact of default of the most interconnected insurers. To do this we make use of Solvency II regulatory returns

from UK insurers. Solvency II forms the regulatory context for insurance supervision, codifying capital requirements and risk management practices. Solvency II replaces Solvency I and improves upon the risk sensitivity in areas such as market, credit, and operational risk, whilst also improving the area of transparency through greater data disclosure to both supervisors and the public. In collaboration with the Bank of England, we make use of this wider data disclosure under Solvency II to characterise empirically the network properties of the UK reinsurance market and to study its systemic resilience under more stringent regulatory requirements.

Our results corroborate the view that the UK insurance network shares similar characteristics with other financial networks (Castiglionesi and Navarro, 2007; Allen and Babus, 2008; Elliot et al., 2014; Fricke and Lux, 2015). In particular, we observe the core-periphery structure across both life and non-life reinsurers – densely connected hubs with sparsely connected reinsurers in the periphery – and the small world property – few connections are required to reach each reinsurer in the network (Watts, 1999) – which typically gives rise to the robust-yet-fragile property studied in the literature. Despite this property being a potential driver for instability, we find defaults from contagion to be low when considering both system-wide shock (scenario inspired by the EIOPA[6] stress test) and idiosyncratic shock in the form of default of the most interconnected insurers. This is in line with the traditional view that insurers are not considered systemically risky due to the combination of long-term liabilities, diversified assets and the extent of its interconnection to the financial system (International Monetary Fund, 2016).

However, following the near collapse of AIG during the 2008 financial crisis driven by the liquidity issues associated with its credit default swap contracts, attention is drawn to the growing trend of insurers taking on 'Non-traditional Non-insurance' (NTNI)[7] activities (Billio et al., 2011; Acharya et al., 2016; Cummins and Weiss, 2014; Swain and Swallow, 2015). Insurers created offshoots that

[6]The European Insurance and Occupational Pensions Authority (EIOPA) is a supervisory authority with remit to ensure stability in the insurance and pensions sectors in Europe. It typically runs a bi-annual stress test of the European (re)insurance sector, the exercise in 2018 had a market coverage of around 75% (EIOPA, 2018).

[7]We use this initial NTNI classification published in 2013 by the International Association of Insurance Supervisors (IAIS) for succinctness. Following stakeholder consultation, the classification was later superseded in 2016 by a more granular and less binary framework of systemic importance concerning substantial increases in macroeconomic and liquidity risk (IAIS, 2016).

competed with banks and engaged in activities such as 'insuring financial products, credit default swaps, derivatives trading and investment management' (Billio et al., 2011).

For (re)insurers involved in such NTNI activities, primarily solvency based stress tests, such as the one conducted by EIOPA, are not adequate in capturing all channels of contagion. Indeed, a consultation initiated by EIOPA (2020) itself on the methodological principles of the insurance stress test identified a gap and need for a conceptual approach to liquidity stress testing in the insurance industry. The sources of liquidity risk identified include those stemming from NTNI activities. The joint framework of stress testing for solvency and liquidity that we propose in Chapter 2 may be applied to these institutions, since our methodology is not limited to a particular type of financial institution. As we show in that chapter, failure to incorporate the interactions of liquidity and solvency in a joint framework can significantly underestimate the risks of a financial institution. Furthermore, the proposed framework is more comprehensive than current stress tests through capturing indirect channels of contagion, such as the effect of fire sales on asset and collateral valuation, and increasing funding costs. These channels of contagion have been found to be relevant for both banks and non-traditional insurers (Jobst et al., 2014).

Systemic Macro Shocks

Exogenous systematic trigger that is widespread and simultaneously affects a range of intermediaries or markets.

The ongoing COVID-19 pandemic is a staple example of a severe and widespread macroeconomic shock. The combination of wider international integration and the possibility of asymptomatic carriers led to a more rapid viral transmission compared to past epidemics. The relatively sudden onset and rapid escalation of the disease to pandemic status created fear amongst many. Worrying projections of overwhelmed health systems set off the trigger for containment policies applied almost in tandem globally, in essence prioritising health outcome above all else.

This in turn created a sudden stop in economic activity.[8] Despite the severity of the pandemic, the lifetime economic cost of COVID-19 remains highly uncertain.[9]

In addition to the impact on health and the economic cost, there is also a social dimension associated with every policy decision. In an independent review commissioned by the British government on the long-term impacts on society of the COVID-19 pandemic, the British Academy (2021) emphasises nine key areas, including low levels of trust[10] and widening inequalities. Therefore, an efficacious disease control measure should holistically consider health, social and economic impacts. These complex interactions warrant the need for bespoke models to guide the design of effective intervention policies.

Indeed, the merits of using mathematical models to capture epidemic dynamics have long been recognised. The modelling of infectious diseases began as early as in the eighteenth century, where Bernoulli (1766) used mathematical modelling to demonstrate the benefits of smallpox inoculation (Brauer, 2008). Since then there has been a myriad of epidemiological models falling into three broad classes of methods: mechanistic models, statistical models and machine learning based models.[11] Aside from different model use cases (for example, whether a model is used to monitor or simulate disease progression), one driver for the variety of epidemiological models is the heterogeneity of infectious diseases (for example, whether infection results in immunity, their mode of transmission, and their varying impacts across geographies and demographics). Another key driver is the challenging trade-off between realism and model complexity.

As the name suggests, mechanistic models make fundamental assumptions about the underlying system mechanics. Compartmental models are one such example, where a population is segmented by the state of health. One of the simplest examples is the SIR model which accommodates three states of health

[8] See, for example, Boissay and Rungcharoenkitkul (2020).

[9] For example, Cutler and Summers (2020) estimate the loss in GDP exceeds $7.5 trillion (£5.5 trillion) in the US alone.

[10] Low levels of trust has long lasting spillover effects on the society. For example, the rate of vaccine uptake and the compliance level of public health interventions may be lowered as a result of a lack of trust in the government's actions.

[11] Literature presenting general overviews of epidemic models in these areas include Brauer and Castillo-Chavez (2012); Britton et al. (2019); Wiemken and Kelley (2020); Clayton and Hills (2013); Siettos and Russo (2013).

– susceptible, infectious, recovered. This model stems from the work of Kermack and McKendrick (1927), in which the authors define homogeneous parameters of infectivity, recovery and death rate across a population, capturing the mechanism of travel between the compartments in a deterministic manner. Their work points to the existence of a critical population density threshold, a function of the three parameters, above which an epidemic occurs should an infected person be introduced to the population, and below which no epidemic occurs. This contradicts an early view that attributes the termination of an epidemic purely to the removal of all susceptible individuals. Whilst being parsimonious, their model adequately captures the typical epidemic dynamics – an initial gradual build up of cases, transitioning to an exponential rise in cases and deaths, followed by a decline which leaves a portion of the population untouched.[12] Kermack and McKendrick build on this initial paper in two further works introducing vital dynamics, capturing population changes not directly related to the modelled disease such as via births, immigration and deaths (Kermack and McKendrick, 1932, 1933). They find that out of most scenarios considered, a single endemic steady state is possible and is again a function of the population density. We note here that whilst vital dynamics add realism to a model, these factors may behave unpredictably during extraordinary times such as under a pandemic. For example, the dramatic variations of migration may occur prior to lockdown as people rush to return home, and flat-line as lockdown measures are put in place. In cases where the model time horizon is short enough and the population is large enough, vital dynamics may add more model complexity than predictive value.

There are additional ways in which the SIR compartmental model can be extended. The compartments could be adapted for different diseases, for example in the case where recovery from infection does not confer immunity, the SIS (Susceptible-Infectious-Susceptible) model may be useful. One may also introduce new types of compartments. For instance, in cases where there is an incubation period between the susceptible and infected stage, an 'exposed' stage may be added, resulting in the SEIR model. The homogeneous population assumption considered thus far is easily challenged, as it assumes the entire population has the same

[12] Allen (2008) provides examples of this dynamic visualised for the Great Plague of London and the New York measels epidemic.

social contact rates irrespective of factors such as age and geography. A direct extension would be to segment the population into distinct 'patches' and to fit a compartmental model to each patch. The governing set of dynamics of these so-called 'multi-patch' or 'metapopulation' models consider both the within patch behaviour but also importantly the interactions between patches. Multi-patch models allow a greater granularity in capturing system dynamics, where the asynchrony across patches sheds light on the role of heterogeneity in the persistence of diseases, something which homogeneous models are not able to capture (Lloyd and May, 1996; Lloyd and Jansen, 2004). Network models offer one further level of flexibility in capturing heterogeneous behaviour in a population, since they consider the specific structure of the underlying framework that intermediates disease transmission. As a result, they are well placed to model human interaction and sources of contagion within the population (El-Sayed et al., 2012). Bansal et al. (2007) conduct analysis on the deviation of real world populations from the classical homogeneous assumptions using a network perspective. In general, the authors observe real-world populations to behave heterogeneously, and thus find network models to offer more accuracy in modelling disease progression through a population. At its simplest, a homogeneous compartmental model can be approximated by a regular random network in which all individuals have an identical number of contacts. On the other hand, a complex network structure may characterise the specific contact behaviour of each person. However, such network based models rely on specific contact pattern data, which often is not available in practice at the required level of granularity. That said, many countries have introduced contact tracing programmes that collect near real-time data from personal devices with in-built GPS functionality in response to the recent COVID-19 pandemic.[13]

Whilst population segmentation provides a more accurate representation of specific contact patterns between individuals, the inherent randomness of these contacts cannot be captured through a deterministic characterisation of the system dynamics. Importantly, at any point in time, deterministic models offer a view of the average behaviour of a system, and hence fail to account for the randomness

[13]We recognise the ethical debate related to such data collection and usage. However, detailed discussion on this subject is out of scope for this thesis.

of real-world disease progression. Identical conditions may lead to different outcomes, as evidenced, for example, in the recent outbreak of COVID-19 pandemic by localised random flare-ups. Notably, the impact of randomness is especially evident at the early stages of a pandemic, where a small number of initial cases can result in drastically varying results across different geographical regions. Similarly, the timing and nature of a potential resurgence in cases after containment policies are lifted is highly uncertain. Therefore, it is paramount to consider the full spectrum of possible scenarios, including considerations of the worst-case outcomes, and take these into consideration as part of effective policy planning. In the case of the compartmental framework, randomness can be incorporated by modelling transitions between states in a probabilistic manner, through a stochastic (point) process that accounts for varying latency and infectious periods across the population.[14]

Given the need of policymakers to capture the complexity and heterogeneity of the underlying system adequately, the use of agent-based models in epidemiological modelling has gained recent attention. The underlying idea of these models is to represent complex systems via individual agents and then to model the specific interactions between them. Depending on the scale of the population modelled, the notion of an agent can range from a single person to a collection of individuals (such as a household or a city). A key benefit of this class of models is its ability to capture the 'emergent' collective systemic behaviour arising from the adaptive actions of individuals (Bonabeau, 2002). However, the increased model complexity comes at a cost. The main challenge with the practical application of agent-based models is the intensive data demands required to underpin realistic agent interactions. Data is rarely available at the granularity and frequency required, especially at the onset of a pandemic, prompting the need for proxies and assumptions based on expert judgement or comparable data sets (such as data from other countries and past pandemics).

A key contribution made towards the modelling of COVID-19 progression in the UK at the beginning of the pandemic is that of Ferguson et al. (2020), in which the authors predict extremely severe health-related consequences should no

[14]For a detailed overview on the topic of stochastic epidemic models see, for example, Britton et al. (2019); Allen (2017).

action be taken to limit the spread of the virus. This influential work prompted a rapid response from the British government that led to the introduction of a nation-wide lockdown. In particular, their counterfactual scenario predicts in total 510,000 fatalities in the UK. The modelling approach of Ferguson et al. (2020) uses a modified agent-based framework, which was previously developed in Ferguson et al. (2005) to investigate the conditions required to halt a potential H5N1 influenza pandemic in Southeast Asia. The disease transmission is modelled and simulated at the individual level using population density data with 1 km resolution. The authors consider contacts in the household, school, workplace as well as 'random' contacts representing travel that can cross national borders. However, the results of Ferguson et al. (2020) should be taken with caution due to the difficulty in parameter estimation at the early stages of a pandemic. Disease dynamics are highly sensitive to parameter uncertainty, especially in the case of complex heterogeneous models.

In a follow-up work by the same working group, Walker et al. (2020) use updated and improved data from a wide variety of sources (for example, mortality and healthcare data from China and high-income countries) to calibrate an SEIR model to inform policy making. Interestingly, their results are quantitatively and qualitatively similar to the previous estimates of Ferguson et al. (2020) for the baseline 'no intervention' scenario, and provide similar policy recommendations to mitigate and suppress the disease spread, including a blend of social distancing, self-isolation and quarantine. Nonetheless, the policies considered are mostly limited to the nation-wide level (the exception being shielding policies for the elderly), and fail to utilise targeted levers such as geography specific interventions and adaptive policies. Moreover, neither Walker et al. (2020) nor Ferguson et al. (2020) capture the socio-economic impact of these policies, and give no guidance on how they should be considered in the context of policy making.

In Chapter 4, we use a stochastic multi-patch compartmental model with demographic and geographical heterogeneity to show that the impact of the epidemic in England is heterogeneous in nature and can vary dramatically depending on the action taken. Importantly, we also acknowledge the challenge of every policy decision as it involves an inherent trade-off between health and socio-economic outcomes. To directly address this, we use our concept of *efficient policy frontiers*

as a tool to help policy makers identify the optimal course of action under their constraints. The outcomes of policy decisions can be explored using our interactive app available at:

`http://covid19.kotlicki.pl`

The COVID-19 pandemic is a reminder of why supervisors and financial institutions use stress tests to gauge the resilience of a firm or system under stress, ensuring the system does not collapse following severe shock. The stress test scenarios used aim to be 'severe but plausible' – for example, in the European Banking Authority (EBA) EU-wide stress test (European Banking Authority, 2018), and the Bank of England stress test (Dent et al., 2016). However, in a recent audit of the 2018 EBA EU-wide stress test, the European Court of Auditors (ECA) found that certain systemic risks were not sufficiently captured, where the adverse scenario represented stress following an economic downturn rather than a severe macroeconomic shock (European Court of Auditors, 2019). Given the variability of outcomes observed in our stochastic model of the COVID-19 epidemic, we argue that the main issue may not be the scenario plausibility, but rather that a wide range of scenarios is not considered.

This is where the approach of reverse stress testing excels, where a range of adverse scenarios can be considered to obtain a pre-specified outcome of a financial institution (for example, failure due to insolvency). In the COVID-19 setting, knowing the boundaries of the system and where the stress becomes 'critical' can enable policy makers to make better informed decisions in the health and socio-economic trade-off. In our work on the joint stress testing of solvency and liquidity in Chapter 2, we show that our proposed framework can be easily adopted for the purpose of reverse stress testing, where 'critical' shock amplitudes are identified. Our proposed visualisation of solvency-liquidity regions offers an informative view on the combinations of shifts in risk factors, such as equity market and interest rate shocks, which will lead to events such as insolvency and illiquidity.

1.3 Main Contributions of the Thesis

In this section we summarise the core contributions of this thesis.

1.3.1 A Framework for Joint Stress Testing of Liquidity and Solvency Risk

In the following we present the contributions of a joint stress testing framework for solvency and liquidity as covered in Chapter 2. This framework has been implemented by the IMF as part of its Financial Sector Assessment Program for Switzerland – an assessment of financial stability and regulatory adequacy of the IMF's member countries (International Monetary Fund, 2019).

New Stress Test Framework We present a novel stress testing framework considering the interactions of liquidity and solvency risk in one coherent framework. We capture endogenous liquidity shocks arising from solvency shocks stemming from margin requirements, contractual obligations triggered due to a credit downgrade, as well as credit sensitive funding. The solvency shocks are then amplified through funding costs arising from liquidity constraints. Our approach avoids the need to introduce exogenous liquidity stress scenarios (which may not be aligned with the solvency scenario) as is common in current stress test practices.

New Measure of Liquidity Risk We introduce *Liquidity at Risk*, a *conditional* measure of net outflow given a stressed scenario. The measure can be compared with the potentially available liquidity resources to gauge the liquidity shortfall. This is in contrast to current regulatory liquidity measures, such as the liquidity coverage ratio (LCR) which is derived using historical data on margin calls and run-off rates (a backward looking approach). As Liquidity at Risk is agnostic to the model used in generating risk scenarios, the measure has the flexibility of being applied to both historical and hypothetical stress scenarios.

Reverse Stress Testing Our stress testing framework includes an extension to reverse stress testing – that is to quantify the impact on solvency and liquidity across a range of adverse scenarios, including regions of failure that arise through

the interactions of solvency and liquidity. We find a non-linear amplification effect of solvency and liquidity interactions. In particular, an increasing amplification effect is observed beyond the fire sale threshold and around the credit downgrade threshold important for credit-sensitive funding. The shock amplification is especially large around the illiquidity default region.

Policy Implications We show that the joint modelling of solvency and liquidity results in a more accurate estimation of the aggregate impact of a stress scenario on a financial institution's equity, as well as enables regulators to adequately capture the illiquidity component of credit risk. We extend the literature by finding that neglecting the solvency-liquidity nexus leads to not just the underestimation of insolvency risk, but also illiquidity risk, rendering the current solvency focused stress tests insufficient for sound risk management. Our framework can be used to identify sources of systemic spillovers that can ultimately threaten financial stability, and can benefit supervisors in planning emergency liquidity assistance for illiquid but solvent financial institutions. Our framework would also benefit industry as it is more in line with bank internal stress tests (which typically consider the liquidity component), allowing supervisors to reward prudence in liquidity provisions and reduce supervisory burden through a convergence in the scope of stress testing.

Accessibility and Interactive App Our framework is designed to be easily implemented by supervisors as it relies on the already existing reporting data. We illustrate the concepts using both the data of a global systemically important bank (G-SIB) and synthetic balance sheet data. The results are accessible from an interpretability perspective through our proposed *solvency-liquidity diagrams*, which can be used to effectively visualise balance sheet dynamics during a joint solvency and liquidity stress test. The stress test methodology developed is implemented in an online tool with the functionality to ingest user provided data to enable custom analysis; the app is available at:

`http://liquidityatrisk.kotlicki.pl`

1.3.2 Network Analysis of the UK Reinsurance Market

In the following we present the contributions of an analysis of the UK reinsurance market as covered in Chapter 3, conducted as part of a collaboration with the Bank of England.

Novel Data This is the first empirical analysis of the UK reinsurance network of its kind using Solvency II data, where we leverage the wider data disclosure requirements to look at reinsurance contracts by line of business (categories of risk). Our contributions include the standardisation of the underlying data and data enrichment using commercial data sources, where the data can be readily used by supervisors and the methodology can be applied to more recent reporting for further study.

Network Characteristics Our work studies the key topological properties of the UK reinsurance network and finds them in line with existing literature of financial networks. The network is similar to theoretical scale-free and small-world networks that exhibit the 'robust-yet-fragile' property to financial shocks. We provide comprehensive observations of the core-periphery structure, including the tendency for most insurers to belong to the same risk sharing network, whilst also identifying the existence of local cliques and when they typically occur. We apply community (clustering) analysis and find that clusters tend not to be organised around lines of business, suggesting a diversification of insurance risk.

Heterogeneity of the Insurance Sector We highlight the differences in risk sharing between life and non-life insurance networks, where typically the literature has focused on non-life only. We find that the life reinsurance network is more hierarchical, with higher exposures and is more sparse than the non-life network. This indicates life insurers share risk via central reinsurers to a larger degree than non-life insurers.

Retrocession Cycles We identify network cycles in the UK insurance industry whereby risks are ceded back to the original reinsurer. That said, the line of business data available show that the type of risks being ceded back to the original

reinsurer typically differ from the risks initially ceded by that reinsurer and hence should not lead to the underestimation of the risk undertaken by the reinsurer. Whilst this does not point to the existence of retrocession cycles, we do identify limitations in the Solvency II line of business data: the limited granularity on the risk type inhibits the full identification of risk amplification and retrocession spirals and thus we suggest standardisation of certain free text fields for supervisor consideration.

Stress Simulation We apply stress simulations to the UK insurance sector and find default rates from contagion to be low and in line with existing literature (for example, Lelyveld et al. (2011); Chen et al. (2020)). Even in the case of a severe shock which resulted in a loss of 93% capital in the system only 0.4% of entities defaulted due to the effect of contagion. Furthermore, we simulated targeted idiosyncratic shocks resulting in the default of highly connected entities, and found the network to be also robust to this type of targeted stress. We acknowledge limitations in the data available (for example, we do not have sight of risks ceded outside of the UK) and propose potential workarounds for future research.

Network Analysis Tool We developed an interactive tool for advanced network analysis, including the functionality for cluster analysis and detection of network cycles.[15] The tool can be used to recreate the analysis presented in the thesis or perform similar studies using other data sets, including those outside of the insurance domain. It doubles as a useful network visualisation tool, utilised by the Bank of England insurance supervision team.

1.3.3 Modelling COVID-19 Contagion: Risk Assessment and Targeted Mitigation Policies

In the following we present the contributions of a model of the COVID-19 epidemic in England that can be used to aid policy making, as outlined in Chapter 4.

[15] See the Banks of England's GitHub repository, `https://github.com/bank-of-england/NetworkApp`.

COVID-19 Model]]Our stochastic compartmental SEIAR model[16] stratifies the population by age and region, focusing on the heterogeneity of epidemic dynamics. Randomness is incorporated into the model to reflect the possibility of different outcomes given identical conditions. These elements cannot be captured by homogeneous compartmental models, which are often cited in policy discussions based around the reproduction number and herd immunity. We show that homogeneous models should not be used to guide policy decisions at the regional level owing to their potential for misleading insights. We demonstrate that our model is capable of accurately reproducing early regional dynamics of COVID-19 in England, both pre-lockdown and one month into lockdown, and is robust to changes in model granularity.

Reporting of COVID-19 Cases Our model takes as input the cumulative number of deaths and the daily *reported cases*. Our model shows that the actual number of cases is much higher than suggested by media reporting, and we infer that less than 5% of cases in England had been detected prior to June 2020.

Efficient Policy Frontier Our framework supports a wide variety of policies with varying levels of severity and targeting, which lends itself to counterfactual simulations and policy planning. We introduce the concept of an *efficient policy frontier* that allows us to identify decision parameters that lead to the most efficient outcome for each policy type (specified by bounds in fatalities and socio-economic cost). Our approach differs from other recent works by not specifying assumptions about the economic value of human life, which is riddled in both ethical and subjectivity issues.

Policy Implications We conduct a detailed comparative analysis of policy levers. We find the most effective policies for reducing fatalities involves decentralised and regional confinement based on the adaptive monitoring of new cases (as opposed to country level and static pre-planned policies). Of the policy types considered, we find shielding the seniors to be by far the most effective in reducing

[16]Susceptible, exposed, infectious, asymptotic, recovered, deceased.

fatalities. In fact, without shielding we find other restrictions in venues such as schools, pubs and workplaces to be inefficient.

Accessibility and Interactive App Our work is implemented in an accessible online tool hosting both pre-defined settings and functionality to accept custom inputs, which can be used to explore the wider suite of scenarios and policy outcomes. The tool can be found at:

`http://covid19.kotlicki.pl`

1.4 Structure of the Thesis

The remainder of this thesis is organised as follows:

- In Chapter 2 we develop a coherent framework for the joint stress testing of solvency and liquidity, including an extension for reverse stress testing. We introduce the concept of *Liquidity at Risk*, a forward-looking measure of liquidity risk conditional on a stress scenario. We demonstrate our framework using data from a G-SIB as well as synthetic balance sheet data. We compare our framework with current supervisory stress tests and detail the policy implications of our work.

- In Chapter 3 we present the network analysis of the UK (re)insurance sector. We standardise a new data set based on Solvency II data and use it to study the topological properties of the UK reinsurance network, including the possibility for retrocession cycles. We apply a stress simulation and verify the robustness of the UK insurance network to direct contagion due to reinsurance contracts.

- In Chapter 4 we develop a stochastic heterogeneous model for COVID-19 contagion and build a framework to aid policy makers in counterfactual simulations and policy planning. We define the concept of the *efficient policy frontier* which highlights the best policies given a trade-off between health outcomes and socio-economic cost. We outline the most efficient policy levers and highlight specific policies found to be effective.

Chapter 2

Joint Stress Testing of Liquidity and Solvency Risk

Chapter based on:
Rama Cont, Artur Kotlicki, and Laura Valderrama. Liquidity at risk: Joint stress testing of solvency and liquidity. *Journal of Banking & Finance* 118: 105871, 2020.

The traditional approach to the stress testing of financial institutions focuses on capital adequacy and solvency. Liquidity stress tests have been applied in parallel to and independently from solvency stress tests, based on scenarios which may not be consistent with those used in solvency stress tests.

We propose a structural framework for the joint stress testing of solvency and liquidity: our approach exploits the mechanisms underlying the solvency-liquidity nexus to derive relations between solvency shocks and liquidity shocks. These relations are then used to model liquidity and solvency risk in a coherent framework, involving external shocks to solvency and endogenous liquidity shocks arising from these solvency shocks.

We define the concept of "Liquidity at Risk", which quantifies the liquidity resources required for a financial institution facing a stress

scenario. Finally, we show that the interaction of liquidity and solvency may lead to the amplification of equity losses due to funding costs that arise from liquidity needs.

2.1 Introduction

Stress testing of banks has become a pillar of bank supervision. Bank stress testing has mainly focused on solvency: a commonly used approach is to evaluate the exposure of bank portfolios to a macro-stress scenario and compare this exposure with the bank's capital in order to assess capital adequacy (Schuermann, 2014). This approach is in line with structural credit risk models which, following Merton (1974), have mainly emphasised solvency.

However it has become clear, especially in the wake of the 2008 global financial crisis, that a typical route to failure for financial institutions may be a lack of *liquidity* triggered by a loss of short-term funding (Duffie, 2010; Gorton, 2012). A famous letter by the SEC Chairman to the Basel Committee emphasises that the failure of Bear Stearns was triggered by the lack of liquidity resources, not capital.[1] The failure of insurance giant AIG, which had a trillion dollar balance sheet, may be traced to a lack of liquidity resources to face margin calls resulting from its credit downgrade (McDonald and Paulson, 2015). This phenomenon is not new. As shown by Postel-Vinay (2016), Chicago state bank failures during the Great Depression were linked to lack of liquid assets to face deposit withdrawals. Blickle et al. (2019) show that the German banking crisis of 1931 was centered around the collapse of interbank and wholesale funding. More recently, the Spanish bank Banco Popular, which displayed a capital ratio of 6.6 percent in the 2016 European stress test, failed because of a lack of liquidity in 2017.[2] These examples illustrate the importance of accurately modelling various channels of liquidity stress for stress testing of banks.

Regulators have taken initiatives for the monitoring and regulation of bank liquidity, such as the Liquidity Coverage Ratio (LCR) and the Net Stable Funding

[1] Letter to the Chairman of the Basel Committee on Banking Supervision on March 20, 2008 (`https://www.sec.gov/news/press/2008/2008-48.htm`).

[2] See `https://srb.europa.eu/en/content/banco-popular`.

Ratio (NSFR), as well as liquidity stress testing to assess the adequacy of liquidity resources of banks. Liquidity stress tests focus on a bank's ability to withstand hypothetical liquidity shocks. The usefulness of such stress tests hinges on the choice of the stress scenarios used for the liquidity shocks. While the Basel III framework emphasises the need for a unified stress testing approach, the assessment of solvency and liquidity risk has remained largely fragmented. Calibration of liquidity shocks is based on supervisory experience rather than a forward-looking assessment of market risk, notwithstanding the increased significance of margin requirements for derivatives under the new European (EMIR) and US (CFTC) rules (Cont, 2017). Current practice is to calibrate such scenarios based on stressed cash in-/outflows and depositor run-offs in recent crisis episodes, using a backward looking approach (European Central Bank, 2019). Although the implementation of the LCR ratio has imposed more stringent liquidity requirements and strengthened banks' liquidity risk practices, the calibration of these requirements is insensitive to the solvency position of the reporting bank and restricted to a prescribed scenario that may differ from the scenario that would deplete the bank capital buffers.

Many theoretical and empirical studies have pointed to the importance of interactions between solvency and liquidity risk (Bernanke, 2013; Cecchetti and Kashyap, 2018; Farag et al., 2013; Imbierowicz and Rauch, 2014; Morris and Shin, 2016; Pierret, 2015; Rochet and Vives, 2004; Schmitz et al., 2019; Basel Committee on Banking Supervision, 2015). Interactions between solvency and liquidity are present in models of bank runs and debt roll-over coordination failures (Allen and Gale, 1998; Diamond and Rajan, 2005; Rochet and Vives, 2004). In a two-period model with short- and long-term liabilities, Morris and Shin (2016) identify two components of credit risk: the "insolvency risk" associated with asset value realisation being below debt value, and the "illiquidity risk" associated with a run by short-term creditors irrespective of the actual solvency state of the institution. Liang et al. (2013) present an extension of the approach by Morris and Shin (2016) to a multi-period dynamic bank run setting where a financial institution is financed through a mix of short-term and long-term debt. A noteworthy implication of this model is that total default risk increases in both rollover frequency and the short-term debt ratio. Cont (2017) describes the role of margin requirements in the transformation of solvency risk into liquidity risk, thereby linking solvency

and liquidity. The importance of the interplay between solvency and liquidity in the context of financial stability also has been evidenced in empirical studies (Cornett et al., 2011; Du et al., 2019; Imbierowicz and Rauch, 2014; Pierret, 2015). Pierret (2015) shows that firms with increased solvency risk are more susceptible to liquidity problems and that the availability of short-term funding decreases with solvency risk. Du et al. (2019) present empirical evidence that indicators of credit quality affect counterparty choice, with the consequence that creditworthiness affects the volume rather than the price of short-term funding. Schmitz et al. (2019) present evidence on the relationship between bank solvency and funding costs and show that neglecting the solvency-liquidity nexus leads to a significant underestimation of the impact of shocks on bank capital ratios.

Despite the evidence on the close link between liquidity and solvency, liquidity and capital requirements are calibrated more or less independently (Cecchetti and Kashyap, 2018) and liquidity stress tests are conducted separately from solvency stress tests (European Central Bank, 2019; Schuermann, 2014). The methodology used in the calibration of liquidity requirements and stress tests either fails to model the interaction of solvency and liquidity risk or includes only a limited channel for such interactions (Basel Committee on Banking Supervision, 2015). For example, in the Bank of Canada's MacroFinancial Risk Assessment Framework (MFRAF) solvency risk affects roll-over risk (Fique, 2017), while in the Austrian Central Bank's stress test, solvency risk limits the access of a financial institution to funding (Basel Committee on Banking Supervision, 2015).[3]

Our goal is to tackle this issue in a systematic manner and build a joint stress testing framework for solvency and liquidity that addresses the interrelations between them. Building on ideas introduced in Cont (2017), we introduce a model in which shocks to asset values generate endogenous liquidity shocks arising from multiple solvency-liquidity interactions channels, thus affecting both the solvency and liquidity of a financial institution.

Contributions We propose a joint stress testing framework for solvency and liquidity: rather than modelling solvency and liquidity stress separately, we integrate

[3]See Table 18 in the online appendix to Baudino et al. (2018) for a summary of solvency-liquidity interactions in current bank stress-testing procedures.

the mechanisms through which they interact and analyse the implications of these interactions for the dynamics of a balance sheet under stress. These mechanisms, summarised in Figure 2.1, lead to

- Endogenous liquidity shocks arising from solvency shocks; and
- The amplification of solvency shocks through funding costs arising from liquidity constraints.

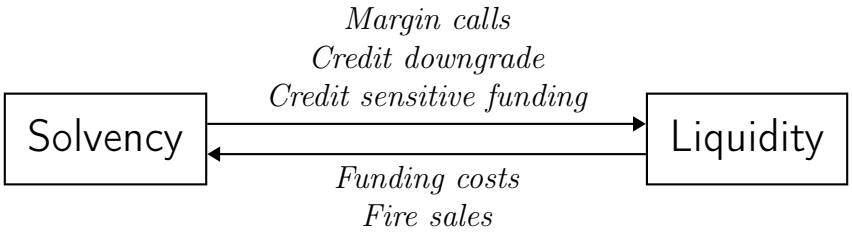


Figure 2.1: Mechanisms governing the solvency-liquidity nexus. Deterioration in solvency may result in a net outflow of variation margin, credit-contingent cash outflows, as well as a loss in short-term funding. Liquidity stress may deteriorate solvency through the inherit cost of liquidity management activities, thus resulting in a feedback loop and a spiral of losses.

We start from a representation of a balance sheet, distinguishing various components in terms of their interaction with the firm's liquidity. We then express the various mechanisms through which these balance sheet components may be affected in a stress scenario, described as a shock to asset values ("solvency shock").[4] Solvency shocks affect liquidity through margin requirements, via the firm's ability to raise short-term funding and through the cost of this funding, leading to *endogenous liquidity shocks*.

In addition to ensuring the coherence between liquidity and solvency stress scenarios, our approach has the benefit of avoiding the introduction of exogenous liquidity stress scenarios. In particular, we observe that identical shocks to risk factors may lead to different endogenous liquidity stress across banks, depending on their balance sheet composition. This feature is easily captured in our approach

[4]By contrast with MFRAF, we model contingent outflows from margin requirements, and allow the firm to raise funding in secured markets before resorting to fire sales.

but would be difficult to implement in an approach where liquidity stress scenarios are specified exogenously.

Our approach also allows us to quantify the amplification of equity losses due to funding costs that arise from liquidity shortfalls. This illustrates how solvency risk may be underestimated by stress tests that do not account for the solvency-liquidity nexus.

The resilience of a balance sheet to the resulting combination of solvency shocks and endogenous liquidity shocks may be visualised through *solvency-liquidity diagrams*, introduced in Section 2.2.3. We define the concept of *Liquidity at Risk*, which quantifies the liquidity resources required for a financial institution facing a stress scenario. In contrast to the current methodology underlying the Liquidity Coverage Ratio (LCR), Liquidity at Risk is a forward-looking measure of liquidity stress *conditional* on a scenario defined in terms of co-movements in risk factors.

The stress testing methodology presented in this chapter has been implemented as an online application available at `http://liquidityatrisk.kotlicki.pl/`.

Chapter outline Section 2.2 introduces the model and explains the various mechanisms through which solvency and liquidity interact. Section 2.3 discusses the extraction of model inputs from balance sheet and regulatory data. Section 2.4 introduces the concept of Liquidity at Risk and illustrates it with two examples: a synthetic balance sheet and the balance sheet of a global systemically important bank (G-SIB).

2.2 A Framework for Joint Stress Testing of Solvency and Liquidity

Figure 2.1 represents various mechanisms through which liquidity and solvency interact with each other. In this section, we introduce a stress testing methodology that aims to capture these mechanisms.

2.2.1 Balance Sheet Representation

A coarse-grained representation of the balance sheet in terms of total assets and total liabilities turns out to be insufficiently detailed to model the mechanisms indicated in Figure 2.1. For instance, in order to quantify potential funding through repurchase agreements, one needs to distinguish unencumbered from encumbered assets and general collateral (GC) from other assets. In order to identify potential sources of margin calls, one needs to distinguish assets subject to margin requirements from other assets. In particular, our focus on distinguishing assets subject to variation margin is motivated by the balance sheets of global systemically important banks (G-SIBs) using publicly reported data as of October 2019. As shown in Figure 2.2, assets subject to variation margin typically form a large amount of total assets for G-SIBs.[5] In turn, these requirements may lead to large endogenous liquidity shocks during a stress and hence should be addressed adequately in liquidity stress tests.

This warrants a more granular decomposition of the balance sheet into components based on their interactions with the firm's solvency and liquidity, shown in Table 2.1.

<table>
<tr><th>Assets</th><th>Liabilities and equity</th></tr>
<tr><td rowspan="2">Illiquid/encumbered assets:
• Subject to margin requirements, I
• Not subject to margin requirements, J</td><td>Maturing liabilities, S</td></tr>
<tr><td rowspan="2">Other liabilities, L</td></tr>
<tr><td>Marketable unencumbered assets:
• Subject to margin requirements, M
• Not subject to margin requirements, N</td></tr>
<tr><td>Liquid assets, C</td><td>Equity, E</td></tr>
</table>

Table 2.1: Stylised balance sheet of a financial institution. We distinguish between assets that generate endogenous liquidity shocks through margin calls as well as those that can be used to raise short-term funding in a stress scenario.

On the asset side, we distinguish:

- *Liquid assets*, which include cash holdings; High Quality Liquid Assets (HQLA), easily convertible into cash; and balances with central banks.

[5] We note a weighed average (by the size of total assets) of 23 percent for the 13 most systemic European banks as of October 2019.

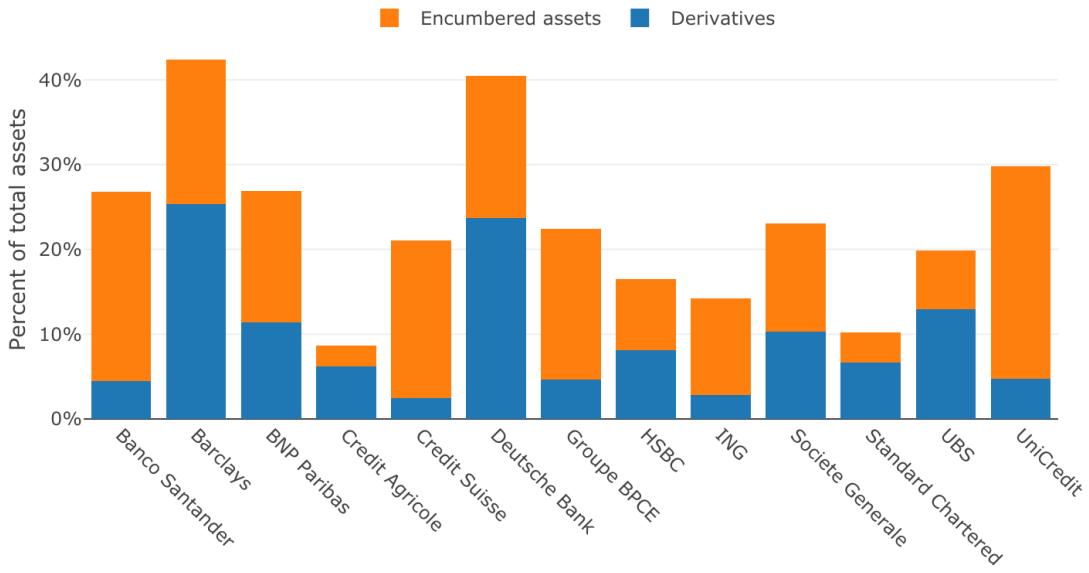


Figure 2.2: Assets subject to variation margin expressed as a fraction of the total value of assets for the 13 European G-SIBs (October 2019). Encumbered assets have been pledged by the bank, for example, to secure a loan, and thus are subject to a legal claim by another institution. Consequently, only unencumbered assets remain available to be used as a collateral in a repurchase agreement.

- *Marketable assets*, defined as assets not in the above category but available for repo or sale. In particular such assets need to be unencumbered by existing repurchase agreements and derivatives transactions. In the context of stress testing, it is conservative to assume that only (unencumbered) assets, mainly in the general collateral (GC) category (subject to a low haircut under stress) would be available for repo in a stress scenario, which is what we shall assume in the examples below. Among these marketable assets we further distinguish:
  - Marketable assets subject to margin requirements, which include for example exchange-traded derivatives and standardised OTC derivatives; and
  - Marketable assets not subject to margin requirements, which include for example financial investments and equity.

- *Illiquid assets* defined as assets that are not "marketable" as defined above. In particular, encumbered assets shall be considered under this category. Among these assets we further distinguish:
  - Illiquid assets subject to margin requirements, which include for example non-standard OTC derivatives; and
  - Illiquid assets not subject to margin requirements (typically loans).

On the liability side, we consider a (short) time horizon and we distinguish:

- *Maturing liabilities*, due within this short-term time horizon.
- *Other liabilities* maturing beyond this time horizon.

As explained below, to model the liquidity risk associated with margin calls and rollover of short-term (for example, overnight) funding, we consider a time horizon of the order of a few days.

The difference between total assets and total liabilities is represented by the firm's equity, denoted by E.

We discuss in Section 2.3 the mapping of balance sheet data and regulatory data to the format presented in Table 2.1.

2.2.2 Dynamics of Balance Sheet Components under Stress

We now describe the dynamics of balance sheet components in a stress scenario over a short-time horizon. It is helpful to break up the sequence of balance sheet transformations occurring over this time horizon into two steps, as shown in Figure 2.3.

This description of the evolution of balance sheet components over a short time horizon may be considered as a building block for a multi-period stress test.

Consider a leveraged financial institution with a balance sheet as in Table 2.1. We denote the initial value of balance sheet components by I_0, J_0, M_0, and N_0, the subscript 0 indicating their initial values at $t = 0$. The initial value of maturing liabilities S_0 represents the amount of liabilities maturing at $t = 2$, while L_0 represents the amount of liabilities maturing after $t = 2$. C_0 denotes the current level of cash reserves and balances with central banks.

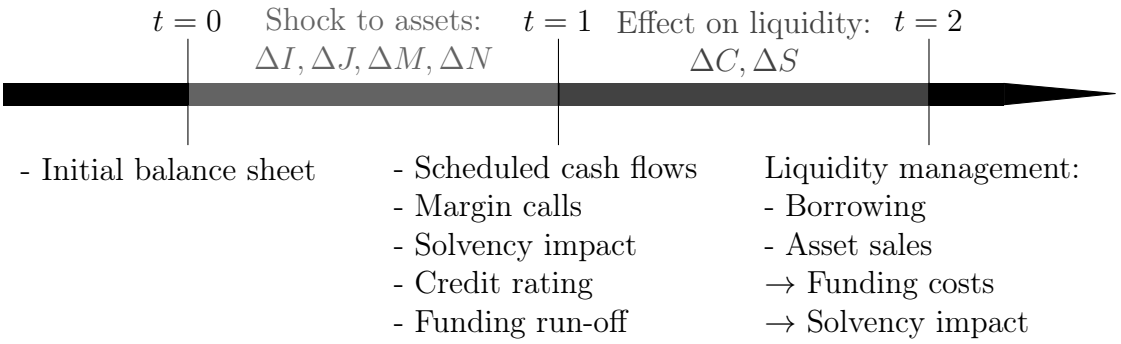


Figure 2.3: Evolution of balance sheet components in the model. The initial balance sheet is firstly subjected to a shock on the assets defined in a stress scenario, which may then result in a credit rating downgrade at time $t = 1$. This initial shock gives rise to endogenous liquidity stress that includes, for example, outflows due to margin calls and loss of short-term funding in the form of deposit run-off. To cover the resultant payables at time $t = 2$, the bank may need to engage in liquidity management activities and thus face the inherit costs of raising new liquidity.

We now consider the impact of an adverse market scenario on this balance sheet.

Stress scenarios Stress scenarios are typically defined in terms of shifts to risk factors such as real GDP, interest rates, credit spreads, equity prices, exchange rates, and other economic variables to which portfolio components are sensitive. Denoting these risk factors by $X = (X_1, \ldots, X_d)$, each stress scenario may be described in terms of shocks $\Delta X = (\Delta X_1, \ldots, \Delta X_d)$ to risk factors.

Direct impact on solvency The reaction of portfolio components to such a stress scenario is evaluated using models calibrated to the risk structure of the portfolio. The models used to derive stress impacts differ across default shocks and market shocks. While the effect of default shocks on credit exposures may take time to materialise, market shocks immediately affect the fair valuation of market exposures. To produce an integrated risk modelling framework, we assume

that firms assess the impact of default shocks on equity using a forward-looking approach (rather than an incurred loss method), and thus the horizon over which shocks hit P&L is the same across risk types. This view is consistent with the Basel III regulatory framework for internal-ratings based models, and the newly implemented accounting IFRS 9 provisions.[6]

For credit shocks, defaults are considered in lending positions (in general valued according to accrual accounting), traded credit positions ("issuer default"), positions measured at fair value and counterparty exposures like OTC derivatives and Securities Financing Transactions. Impairment losses reduce the carrying amount of credit risk positions affecting the value of equity. Impairment charges can be computed as the impact of stressed credit risk parameters – that is, probability of default (PD), loss given default (LGD) and exposure at default (EaD) – on the initial value of the position. Shifts to PDs, LGDs, and EaDs can be expressed in terms of sensitivities to underlying risk factors.

For market shocks, the impact of the shocks on bank portfolios at partial or full fair valuation measurement, can be calculated either by revaluation of the positions in the portfolio under the stress scenario (full valuation method) as computed in firm internal stress tests and regulatory bottom-up stress tests or, as done frequently in top-down regulatory stress tests, by using a linear approximation of the dependence of portfolio components with respect to risk factors, in terms of sensitivities to risk factors. Market shocks are exogenous shifts to risk factors prescribed in the stress test scenario that generate initial market losses. These shocks are different from the endogenous fire sales shocks derived endogenously from banks' liquidity risk mitigation actions.

Denoting $\partial_k A$ the sensitivity of balance sheet component A to risk factor X_k, the change in the value of this balance sheet component in the risk scenario is then given by

$$\Delta M = \sum_{k=1}^{d} \partial_k M . \Delta X_k = \partial M . \Delta X, \tag{2.1}$$

[6]To compute regulatory capital, banks using internal-ratings based models for credit risk take a forward-looking approach to determine capital ratios. From an accounting perspective, IFRS 9 requires loan allowances based on 12 month expected losses if the credit risk has not increased significantly, and expected lifetime losses for exposures that have deteriorated significantly.

where ∂M denotes the vector of sensitives of balance sheet component M. Similarly we may compute the changes in balance sheet items I, J, N as

$$\Delta I = \partial I.\Delta X, \qquad \Delta J = \partial J.\Delta X, \qquad \Delta N = \partial N.\Delta X. \tag{2.2}$$

These sensitivities may be computed using satellite models linking scenario shocks to credit risk parameters (default shocks), or calculating the impact of risk factors on fair-valued positions using the delta method (market shocks).[7]

Impact on liquidity Liquidity risk arises from the uncertainty to meet payment obligations in a full and timely manner in a stressed environment. In the model, obligations coming due at $t = 2$ include four components.

1. *Unconditional liabilities*: these are liabilities maturing at $t = 2$. Their size corresponds to maturing liabilities and hence is denoted by S_0.

2. *Scheduled cash outflows (SCO)*: these include contractual cash flow obligations (for example, interest payments on interest-bearing liabilities, coupons, operating costs), projected outflows from non-maturing liabilities (for example, sight, operational deposits) and estimated drawdowns of undrawn credit and liquidity lines.[8] Denoting these outflows by SCO, the stable component of short-term liabilities payable at $t = 2$ can then be expressed as

$$S_1 = S_0 + SCO. \tag{2.3}$$

3. *Contingent liquidity risks*: in a derivative transaction or securities financing transaction with no margin payments, although both sides may mark-to-market their position daily, there is no exchange of cash flows: any losses or gains purely affect the solvency of the institution. In this case, capital buffers are an adequate tool to address any risk externalities. On the other hand, if an asset is subject to margin requirements, this creates a liquidity outflow in the form of a variation margin payment. As a result, such shock affects not

[7] See Section 2.3 for more details.

[8] Drawdowns of lines of credit are material for US banks. Levine and Sarkar (2019) document that, together with expected margin calls, they account for around 20 percent of gross outflows in the LCR reported by G-SIBs.

only the solvency of the institution but also its liquidity by drawing on the cash reserves held with an immediate effect (typically within a few days), since all payments are made in cash or liquid assets. Firms post and receive collateral to support or reduce the counterparty credit risk (CCR) relative to derivative transactions or to securities financing transactions, including transactions cleared through a central counterparty (CCP). Here we focus on liquidity needs from changes in the value of collateral posted by the bank (for example, in repo transactions) rather than on collateral received (for example, in reverse repos) to allow an integrated assessment of the solvency and liquidity risk of the firm from valuation shocks to the bank assets. For assets subject to variation margin, negative changes in asset values lead to margin calls that add to maturing liabilities, which we denote by

$$\Delta S = (\Delta I)^- + (\Delta M)^-, \tag{2.4}$$

whereas positive changes generate margin calls to the counterparty, which lead to cash inflows expected at $t = 2$, and which we denote by

$$\Delta C = (\Delta I)^+ + (\Delta M)^+, \tag{2.5}$$

where $(X)^+ = \max(0, X)$ denotes the positive part of a quantity X and $(X)^- = (-X)^+$.

4. *Liquidity needs related to downgrade triggers*: the direct impact of the shocks described above on the firm's equity is given by

$$E_1 = E_0 + \Delta I + \Delta J + \Delta M + \Delta N + SCI - SCO, \tag{2.6}$$

where SCI designate scheduled cash inflows (such as interest payments) and maturing assets that are not reinvested (for example, inflows from performing exposures and secured lending).[9] If these changes result in the firm's equity falling below a threshold, the firm may be subject to a *credit downgrade*. We

[9]For simplicity, we assume that all scheduled cash flows have an equity impact, although most of the equity impact comes from interest that is expected to be received and paid during the horizon.

assume such a downgrade occurs if the leverage ratio exceeds a level δ, that is,

$$\frac{I_1 + J_1 + M_1 + N_1 + C_1}{(E_1)^+} > \delta. \tag{2.7}$$

Such a downgrade may trigger the loss of credit sensitive funding – including institutional and retail deposits that can be withdrawn on demand, and outflows associated with a downgrade in the bank's credit rating. We denote by S_D the increase in maturing liabilities resulting from a downgrade.[10]

As a result, conditional on the stress scenario, maturing liabilities due at $t = 2$ increase to

$$S_2 = S_1 + \Delta S + S_D \mathbb{1}_{\text{downgrade}}. \tag{2.8}$$

On the other hand, the reserve of liquid assets is increased by the scheduled cash inflows:

$$C_1 = C_0 + SCI. \tag{2.9}$$

Mitigating actions At $t = 1$, if liquid assets are not enough to cover conditional cash outflows (expected and unexpected), the bank can undertake mitigating actions (from its contingency funding plan and recovery plan) to cover the liquidity shortfall λ, which we define formally as

$$\lambda = \left(S_2 - \{C_1 + \Delta C\}\right)^+. \tag{2.10}$$

In the short term, a financial institution has access to several sources of funding, stated in the usual order of preference based on cost considerations. This pecking order is consistent with the Senior Financial Officer Survey conducted by the Federal Reserve (2019) for cash management operations of US banks. It has also been documented empirically by Blickle et al. (2019) for the systemic German liquidity crisis of 1931, and is in line with Kapadia et al. (2012) based on information from UK banks' contingency plans and the assessment of defensive actions actively taken during the global financial crisis:

[10]The elasticity of customer deposits to a bank's credit downgrade is parameterised outside the framework.

1. *Unsecuritised borrowing:* we assume the financial institution has access to short-term unsecuritised loans given at an exogenous market interest rate r_U. This access depends on the firm's creditworthiness: we assume that the firm's access to such funding ceases once it has been downgraded.[11] Furthermore, the distance to downgrade leads to an upper bound on the volume of unsecuritised lending available to the firm:

$$v_U = \frac{(E_1\delta - \{I_1 + J_1 + M_1 + N_1 + C_1\})^+}{1 + r_U\delta}. \tag{2.11}$$

   In other words, highly leveraged institutions are considered less creditworthy and hence can access a smaller pool of liquidity than lesser leveraged firms. Subject to this constraint, the amount of money a financial institution will borrow through this channel can be expressed as

$$B_U = \min(\lambda, v_U). \tag{2.12}$$

2. *Repurchase agreements (repo):* banks can raise liquidity by entering a repurchase agreement (repo) with a market counterparty. This requires the provision of liquid marketable (unencumbered) collateral and thus the volume v_R of liquidity the bank can raise through this channel is capped by the size $M_1 + N_1$ of the firm's pool of unencumbered marketable assets, discounted by the corresponding haircut parameter $h \in [0, 1)$, that is

$$v_R = (1 - h)(M_1 + N_1). \tag{2.13}$$

   Consequently, the amount of cash that the financial institution can raise through the repo market is then given by

$$B_R = \min\{\lambda - B_U, (1 - h)(M_1 + N_1)\}, \tag{2.14}$$

   with an associated borrowing cost given by the (exogenous) repo rate r_R.

[11] The firm interacts with other financial institutions through the leverage constraint: the ability of a firm to tap interbank funding decreases when other banks choose not to roll over or grant new funding over solvency concerns. The propagation of funding stress through the interbank market is, however, outside the scope of the thesis.

3. *Central bank repo:* one may also consider the possibility of raising short-term funding against collateral through a repo agreement with the central bank – a possibility available to banks in many jurisdictions as a backup source of funding. An example of this is the Eurosystem collateral framework that has played a key role during the financial and sovereign debt crisis to help prevent large-scale liquidity-driven defaults in Europe (Bindseil et al., 2017). The central bank typically accepts a wider range of collateral than repo markets; this corresponds in our notation to a fraction $\widetilde{\theta}_J \in [0,1]$ of (unencumbered) illiquid assets. Compared to the repo market, a higher haircut $H > h$ is typically applied to the collateral pledged to secure such funding. The amount of funding that is raised through this channel is given by

$$B_C = \min\{\lambda - B_U - B_R, (1-H)\widetilde{\theta}_J J_1\} \tag{2.15}$$

and is capped by the value of the unencumbered collateral available net of the haircut applied, that is $(1-H)\widetilde{\theta}_J J_1$.

4. *Liquidation of assets (fire sales):* we assume that over the short-term horizon a liquidity-stressed financial institution can only sell a fraction $\theta_J \in [0,1]$ of its illiquid assets (that cannot be accepted as a collateral by the central bank) in a fire sale with a price discount $\psi \in [0,1)$. Note that only unencumbered illiquid assets can be monetised in a fire sale. In other words, the maximum amount of liquidity that can be raised during the short-term can be expressed as

$$v_F = (1-\psi)\theta_J J_1. \tag{2.16}$$

The fraction θ_J depends for example on the available market liquidity and the length of the sales horizon. Consequently, we expect θ_J to be small in a stress test scenario. Similarly, we usually think of the associated fire sale discount as large (in excess of 50 percent).[12]

These mitigating actions increase the liquidity buffer of the bank at $t = 2$ to

$$C_2 = C_1 + \Delta C + B_U + B_R + B_C + \omega v_F, \tag{2.17}$$

[12]This is consistent with calibrated parameter values in the literature (Cont and Schaanning, 2017).

where B_U represents the amount of new unsecuritised borrowing, the total amount of borrowing through repo is $B_R + B_C$, and $\omega \in [0, 1]$ is an endogenous fraction of liquidated assets in a fire sale for a price discount of $\psi \in [0, 1)$ such that

$$\omega = \min \left\{ \frac{(S_2 - (C_1 + \Delta C + B_U + B_R + B_C))^+}{(1 - \psi)\theta_J J_1}, 1 \right\}. \tag{2.18}$$

The amount of other liabilities rises by the amount of new liabilities from unsecured and secured funding and declines by the cash flow amount due to the run of credit risk sensitive funding, that is,

$$L_2 = L_0 + (1 + r_U)B_U + (1 + r_R)(B_R + B_C) - S_D \mathbb{1}_{\text{downgrade}}. \tag{2.19}$$

As a consequence of these mitigating actions, the value of equity reduces to

$$E_2 = E_1 - r_U B_U - r_R(B_R + B_C) - \omega\psi\theta_J J_1. \tag{2.20}$$

The associated cost of liquidity management activities means that the interaction between solvency and liquidity risk through margin requirements and creditor runs may lead to a severe amplification of losses in a stressed environment (see Section 2.2.4).

Insolvency and illiquidity A financial institution is deemed insolvent when the equity falls below a certain threshold, here taken without loss of generality to be zero. That is, a firm fails due to insolvency when $E_2 < 0$. On the other hand, it is said to be illiquid when maturing liabilities exceed the firm's capacity to raise liquidity, that is, $C_2 < S_2$, where C_2 is the available liquidity, given by Equation (2.17) and S_2 are the maturing liabilities due at $t = 2$, given by Equation (2.8). It is possible for a firm to be illiquid without being insolvent, as it is possible to be insolvent without being illiquid.

The dynamics of balance sheet components are summarised in Figure 2.4.

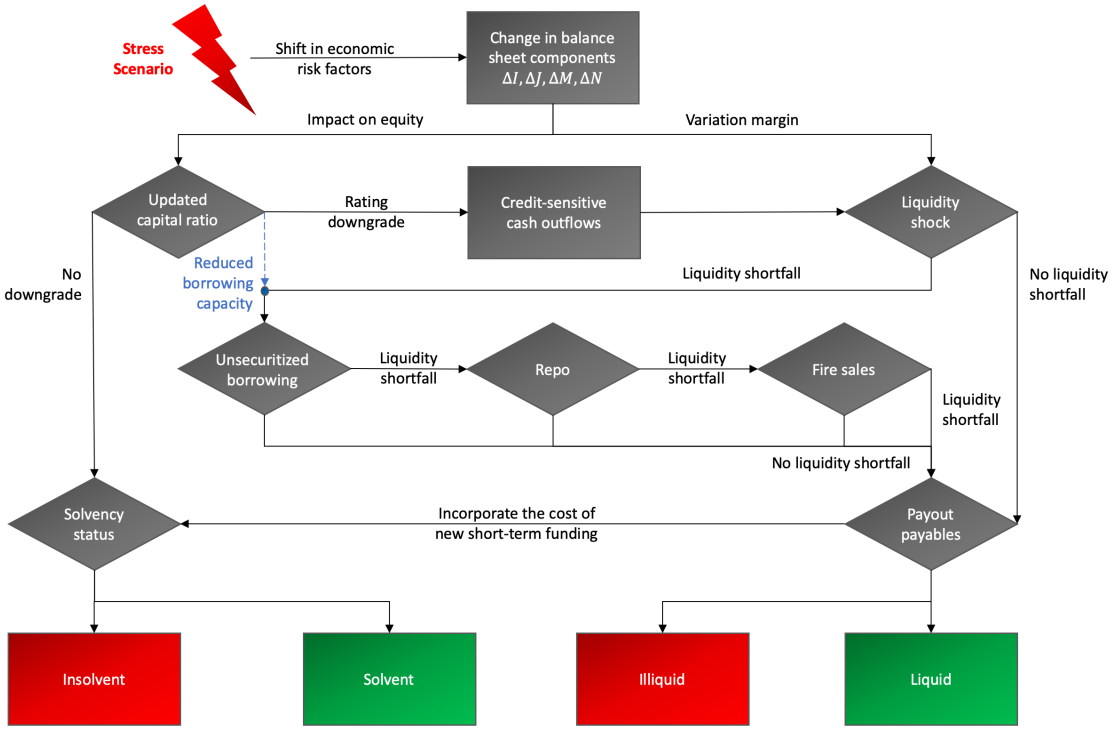


Figure 2.4: Joint stress test of solvency and liquidity. In a given stress scenario, an endogenous liquidity shock arises as a response to the change in the value of balance sheet components. Faced with a liquidity shortfall, a financial institution will attempt to increase its cash reserves through liquidity management activities. Since the available new funding is limited in volume, the firm may still default on payment. Furthermore, the cost of new funding exacerbates the solvency shock, and thus may lead to a failure due to insolvency.

2.2.3 Solvency-Liquidity Diagrams

The balance sheet dynamics in a stress scenario may be visualised in the form of a *solvency-liquidity diagram* in which the financial institution's equity is represented on the horizontal axis and its liquidity resources on the vertical axis (see Figure 2.5).

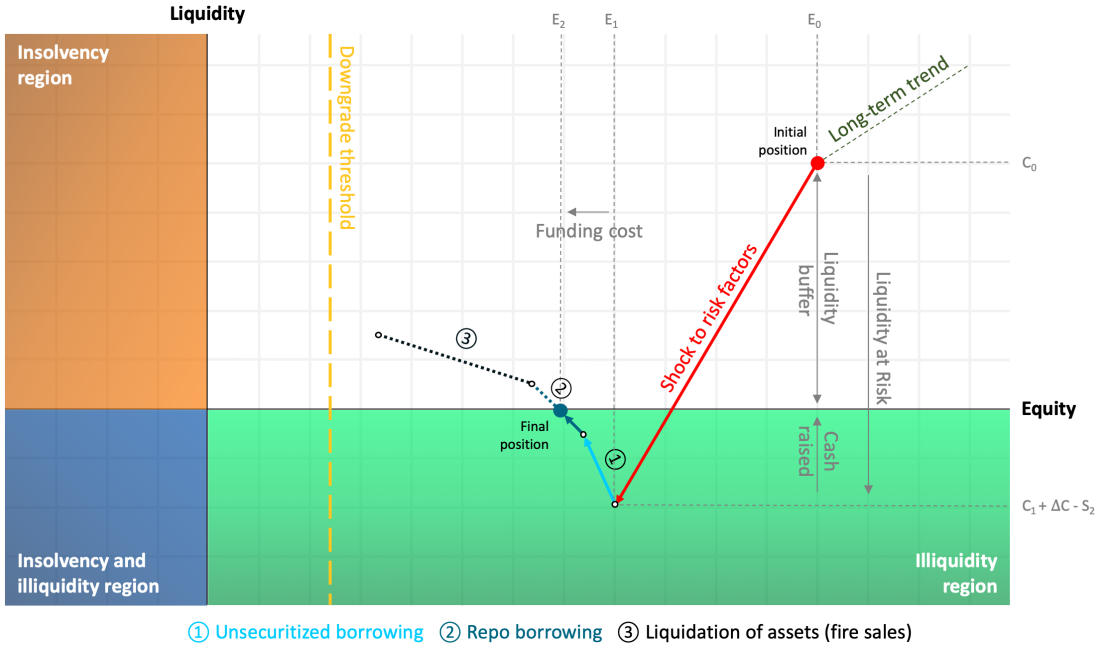


Figure 2.5: Solvency-liquidity diagram describing the behaviour of a balance sheet in a stress scenario. At every point in time, the balance sheet may be summarised by the held liquidity buffer and its equity. An adverse shock will usually reduce these quantities, resulting in respective downward and leftward movements. Although the liquidity buffer may be increased through borrowing and liquidation of assets, resulting in an upward movement, the associated cost will further deteriorate the solvency condition, moving the firm further to the left. Firms with their final position at time $t = 2$ that are found below the horizontal axis are in a default, while those with their final position to the left of the vertical axis have failed due to insolvency.

A solvent and liquid institution corresponds to a point in the upper right quadrant (first quadrant). The vertical coordinate corresponds to its liquidity buffer while the horizontal coordinate correspond to the firm's equity.

A loss in asset values in a stress scenario moves this point to the left. Depending on the cash flows arising in the stress scenario, we may also have a vertical displacement upwards (if there is net incoming cash, for example due to the variation margin and interest received) or downwards (if there are net outflows, for example from margin and interest payments).

Failure occurs when the institution exits this first quadrant. If it crosses the horizontal axis (see Figure 2.6), this corresponds to an illiquidity induced default, and if it crosses the vertical axis (see Figure 2.7) this corresponds to failure due to insolvency. The distance to the axes represents the capital and liquidity buffers (see Figure 2.5).

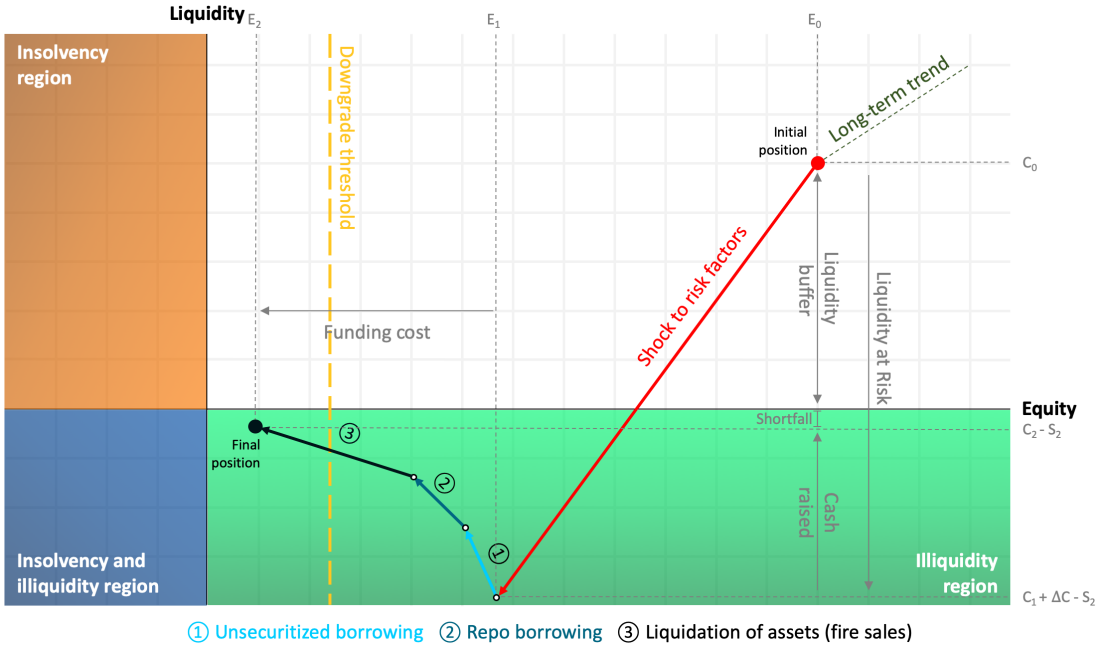


Figure 2.6: Example of a stress scenario leading to illiquidity. A severe liquidity shock results in outflows significantly beyond the firm's liquidity buffer. Despite mitigating actions that include fire sales, the firm is unable to raise sufficient new short-term funding to cover this liquidity shortfall, and thus defaults on payment.

An adverse stress scenario leads to a "south-west" shift on the diagram: the precise direction of the shift depends on balance sheet sensitivities, while the size of the shift corresponds to the severity of the shock. A pure solvency shock draws on the capital buffer without affecting the firm's liquidity reserves, and hence

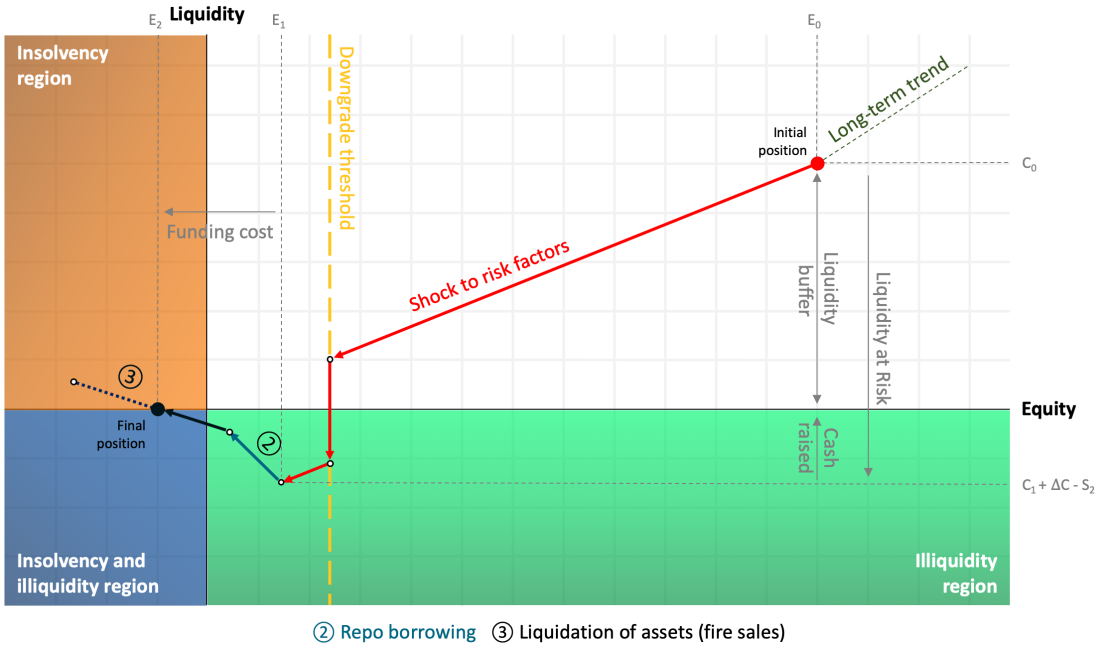


Figure 2.7: Example of a stress scenario leading to insolvency. Following a severe shock, liquidity stress is exacerbated by the downgrade. Having lost its creditworthiness, the firm cannot get new unsecuritised loans and thus is forced to enter fire sales to cover its liquidity shortfall. Such costly liquidation of assets pushes the firm to insolvency.

corresponds to a horizontal shift on the solvency-liquidity diagram. On the other hand, a pure liquidity shock caused by a run of creditors or a failure to rollover short-term debt due to a downgrade corresponds to a vertical shift on the diagram.

For a fixed adverse market scenario, the initial loss in equity due to the shock is independent of the balance sheet composition in terms of margin requirements. However, as the proportion of assets subject to variation margin increases, the reduction in the liquidity position of a financial institution (unencumbered assets) and its liquidity risk (sensitivity to market shocks) also increases. In that case, it becomes more likely that the firm becomes illiquid while still solvent as the shock severity increases.

2.2.4 Loss Amplification through Liquidity-Solvency Interaction

The mechanisms described above may result in an amplification of the initial shock to the firm's equity.

In addition to the impact on equity due to the initial shock

$$E_1 - E_0 = \Delta I + \Delta J + \Delta M + \Delta N + SCI - SCO, \tag{2.21}$$

funding costs lead to an additional loss in equity given by

$$E_2 - E_1 = -r_U B_U - r_R(B_R + B_C) - \omega\psi\theta_J J_1. \tag{2.22}$$

This is represented by a horizontal shift in the solvency-liquidity diagram (Figure 2.5). This amplification effect may be quantified by the ratio of new funding costs to the initial shock to equity:

$$\text{Loss amplification} = \frac{E_2 - E_1}{E_1 - E_0} \times 100\%. \tag{2.23}$$

Loss amplification increases in both the volume of new funding and the cost of raising new liquidity. Therefore, this effect becomes especially prominent for stressed institutions that are forced to liquidate a large amount of their assets in a fire sale.

In the model, the amplification effect occurs for shocks that lead to cash outflows larger than the current liquidity buffer held by the bank. However, if for example the firm aims to sustain a certain level of liquidity buffers, this effect can occur in practice for any adverse shocks.

We illustrate the significance of the loss amplification mechanism on numerical examples shown in Section 2.4.2.

2.3 Mapping of Balance Sheet Variables and Liquidity Templates

The purpose of this section is to show how balance sheet information – especially in the format of templates available to regulators – may be mapped to the format shown in Table 2.1 used as an input for our stress testing approach. In this section we describe how to use various data sources to generate the inputs required in our

framework. We then provide a numerical illustration using publicly available data for a global systemically important bank (G-SIB).

2.3.1 Data Requirements

Our stress testing approach requires three types of inputs:

1. *Balance sheet data*, with sufficient granularity in order to extract the categories displayed in Table 2.1.

2. *Risk parameters*, including credit scores, internal risk reports, and market risk sensitivities to be used for estimating the profit and loss (P&L) and other comprehensive income (OCI) of various portfolio components in the stress scenario.

3. *Liquidity data*, to estimate the amount of available unencumbered assets, contractual maturity cash in-/outflows, and the potential liquidity generation capacity of securities over different time horizons.

These requirements are not very different from the inputs of current solvency stress tests but require the data to be formatted in a slightly different way, as discussed in Section 2.2. Central banks and regulators typically have access to data on portfolio positions, risk parameters, pricing models and methodologies to assess sensitivities to stress. For instance, in the European reporting framework, financial data are collected in Financial Reporting Framework (FINREP) templates, while risk data are submitted in Common Reporting Framework (COREP) templates. The reporting requirements, defined by the European Banking Authority (EBA) via the implementation of technical standards or guidelines, are complemented with short-term exercise ad-hoc data requests. These correspond to additional granular data on complex portfolios including sensitivities to moves in market risk factors. Our stress testing framework requires the data to be available at a sufficiently granular level to derive the above information for each component of the balance sheet.

Table 2.2 summarises the mapping of asset categories observed in regulatory and accounting templates to balance sheet components required in the model. Assets are classified as "marketable" or "illiquid". *Marketable* refers to the availability

of the assets for raising short-term funding in a stress scenario, either through a repurchase agreement or sale. Such assets therefore need to be unencumbered by other repurchase agreements. Because we are interested in behaviour of the balance sheet under stress, we restrict marketable assets to those that can generate liquidity through monetisation at stressed haircuts over the relevant time horizon. Illiquid assets that can be subject to fire sales include loans, investments in associates, and physical assets. Assets that are not available to raise funding and cannot be pledged for repo transactions include complex hard-to-value assets (Level 3 in the fair value hierarchy), goodwill, and deferred tax assets.

	Not subject to Variation Margin	*Subject to Variation Margin*
Illiquid assets	Loans Non-financial investments Physical assets	Non-standard OTC derivatives Encumbered assets
Marketable assets	Unencumbered GC: • Assets held for trading • Financial investments Equity	Exchange-traded derivatives Standardised OTC derivatives
Liquid assets	Cash (unencumbered) Reverse repos	

Table 2.2: Mapping of common asset classes to the model input format.

Once the balance sheet data has been mapped to the format shown in Table 2.2, the stress test requires estimating the variations in each component in the stress scenario considered. The estimation of P&L may be done either through full revaluation in a pricing model, which requires granular data on fixed-income and derivatives positions; or through a linear approximation, using sensitivities to risk factors. In the latter case, one would only require sensitivities to risk factors aggregated at the level of the balance sheet components shown in Table 2.1.

Projection of losses in stress scenarios typically involves two types of risk: credit risk and market risk.

For a credit risk assessment, the loss related to default events on lending positions, traded credit risk positions, and counterparty exposures like OTC derivatives needs to be projected. Impact on P&L through newly created adjustments for loan

loss provisions can be estimated using satellite models based on internal ratings-based models or standardised approaches using stressed credit risk parameters. Under IFRS 9 accounting standards,[13] losses are generated from obligor grade migration using an expected loss, forward-looking approach. Similarly, the current expected credit loss (CECL) impairment model under US GAAP, will require banks to estimate expected credit losses over the contractual life of an instrument, before incurred losses materialise.

To assess market risk, we need to measure the impact of the shocks on the fair values of the underlying positions. Shocks include shifts to risk factors across asset classes including benchmark rates, credit spreads, foreign exchange, equities, and commodities. Accounting data serves to classify exposures at fair value (mark-to-market) relative to exposures at amortised cost. While shocks to financial assets held for trading and financial assets designated at fair value through P&L impact directly, shocks to available-for-sale financial assets affect regulatory capital through OCI. By contrast, shocks to held-to-maturity assets affect bank capital through an increase in provisions. The sensitivities with respect to the relevant (market) risk factors can be calculated using portfolio valuation models or requested from banks through usual regulatory submissions. These sensitivities report the impact of a risk factor move on the fair value of the position. Fire sale risk is reflected in the discount rate applied to the endogenous sale of illiquid assets. Raising liquidity by selling assets in a fire sale is the most costly management action. Therefore it is considered the last course of action.

Basel Liquidity Monitoring Templates provide a granular decomposition of cash outflows and inflows by time horizon (Pohl, 2017), which can be exploited to estimate liquidity needs arising from an adverse scenario over a defined time horizon. To populate the cash flow equation, maturing liabilities according to current contractual conditions include securities issued, unsecured funding by retail and wholesale counterparties, liabilities from secured funding, and additional outflows from derivative transactions and other contingent obligations.

[13]Under IFRS 9 implementation, credit risk is based on the categorisation of exposures in three stages: S1 (credit risk has not increased significantly since initial recognition, and provisions are based on a 12-month expected loss); S2 (credit risk has increased significantly, so the loss allowance should equal lifetime expected credit losses); and S3 (exposure is considered credit-impaired with lifetime allowance and non-recognition of interest accrual).

The stress test needs to project scheduled net cash outflows over the time horizon using the contractual maturity mismatch template, and including estimated values on behavioral flows basing on banks' modelling assumptions or relying on Basel LCR-prescribed scenario assumptions. Contingent liabilities from a downgrade in the bank's credit rating can be estimated using the bank's reported outflows in the liquidity templates, and applying stressed run-off rates on credit sensitive contractual outflows (for example, uninsured deposits, unsecured wholesale funding) based on historical experience.

Contingent liabilities from assets subject to margin requirements can be calculated by applying scenario shocks to risk factors on the value of collateral posted for counterparty credit risk exposure in derivative transactions and Securities Financing Transactions. The data is reported in the contractual mismatch and asset encumbrance submission of the Liquidity Monitoring Templates. While contingent outflows also can be triggered by financial instruments' price changes related to own securities issued, or unsecured funding instruments, these are typically not material.

2.3.2 Example of a G-SIB Balance Sheet

We give an example of such a mapping based on publicly available data for a European G-SIB at end 2017. Public data sources include the bank's annual report containing information on balance sheet, asset encumbrance, fair value hierarchy, and financial assets available for sale and held to maturity; Pillar 3 disclosures containing information on composition of collateral for CRR exposure, and liquidity coverage ratio; and the supplementary information on the balance sheet, income statement and regulatory ratios from the Fitch database. We now illustrate how balance sheet variables are mapped to the portfolio components of the balance sheet using portfolio data on credit risk and market risk positions.

Illiquid assets subject to margin requirements (asset class I) are mapped to encumbered assets pledged as collateral in derivative and securities financing transactions including trading portfolio assets,[14] loans, and financial assets designated at fair value. These positions amount to a value of 64,021 M EUR.

[14]This excludes financial assets for unit-linked investment contracts.

Illiquid assets not subject to margin requirements (asset class J) represent 514,550 M EUR. These include three categories of assets:

1. Encumbered assets, not pledged as collateral, but restricted and not available to secure funding: this category includes mainly financial assets for unit-linked investment contracts, and some lending positions. They reach 23,573 M EUR.

2. Assets that cannot be pledged as collateral: this category covers some loans, cash collateral on securities borrowed, reverse repurchase agreements, and other assets including cash collateral receivables, goodwill, and deferred tax assets. Assets in this category represent 167,445 M EUR.

3. Other realisable assets. These assets include most lending positions (that is, loans in the banking book, due from banks, and financial assets designated at fair value), some trading portfolio assets, property investment, and investment in associates. The amount of realisable assets reaches 323,532 M EUR.

Marketable assets subject to margin requirements (asset class M) denote the fair value of derivative transactions including Level 1 and Level 2 assets of the fair value hierarchy. These contrast with Level 3 instruments that do not have quoted prices in active markets and rely on valuation models where significant inputs are not based on observable market data (for example, long-dated complex derivatives). The latter are considered non-marketable and cannot be monetised over a short time horizon. For the G-SIB considered in the example, derivative instruments include mainly interest rate and foreign exchange contracts, and to a lower extent equity contracts. Less significant are credit derivative and commodity contracts. The value of this category reaches 118,227 M EUR.

Finally, marketable assets not subject to margin requirements (asset class N) include unencumbered instruments available to secure funding. These marketable assets include financial assets at fair value for 45,117 M EUR, trading portfolio assets for 68,369 M EUR, financial assets available for sale for 8,419 M EUR, and held-to-maturity instruments for 9,166 M EUR. Overall, category N represents 131,071 M EUR.

To complete the mapping of balance sheet assets, liquid assets (asset class C), including unencumbered cash and balances with central banks, amount to 87,775 M EUR. Equity reaches 51,271 M EUR. Maturing liabilities represent outflows on retail deposits according to the bank modelling assumptions, and outflows on maturing unsecured debt. The remaining obligations are denoted as other liabilities. Scheduled cash outflows include contractual funding obligations for 13,000 M EUR, outflows from secured wholesale funding for 79,000 M EUR, and estimated drawdowns of committed credit and liquidity facilities for 9,000 M EUR. Scheduled cash inflows include inflows from reverse repurchase agreements for 83,000 M EUR, inflows from fully performing exposures for 33,000 M EUR, and other cash inflows for 10,000 M EUR.

The result of the mapping is shown in Table 2.3.

<table>
<tr><th>Assets</th><th>Liabilities and equity</th></tr>
<tr><td rowspan="2">Illiquid assets:
• Subject to VM, $I_0 = 64021$
• Not subject to VM, $J_0 = 514550$</td><td>Maturing liabilities, $S_0 = 37000$</td></tr>
<tr><td rowspan="2">Other liabilities, $L_0 = 827373$
(incl. deposits of 409000)</td></tr>
<tr><td>Marketable unencumbered assets:
• Subject to VM, $M_0 = 118227$
• Not subject to VM, $N_0 = 131071$</td></tr>
<tr><td>Liquid assets, $C_0 = 87775$</td><td>Equity, $E_0 = 51271 \quad (5.6\%)$</td></tr>
</table>

Table 2.3: Simplified balance sheet of a European G-SIB for year 2017 (in millions of EUR).

2.4 Liquidity at Risk

The framework introduced above enables us to move beyond a liquidity risk analysis purely based on exogenous *expected* cash flows: we define a concept of liquidity stress *conditional* on a stress scenario, which we baptise *Liquidity at Risk*.

2.4.1 A Conditional Measure of Liquidity Risk

Definition 2.4.1 (*Liquidity at Risk*)**.** *Consider a stress scenario defined in terms of shocks to asset values. We call* Liquidity at Risk *associated with this stress*

scenario the net liquidity outflows resulting from this stress scenario:

$$\textit{Liquidity at Risk} = \textit{Maturing Liabilities} + \textit{Net Scheduled Outflows} + \textit{Net Outflow of Variation Margin} + \textit{Credit-Contingent Cash Outflows}$$

The liquidity shortfall in a stress scenario is thus given by the difference between the Liquidity at Risk associated with the stress scenario and the liquid assets available at the point where the scenario occurs.

Liquidity at Risk is easy to read off from the solvency-liquidity diagrams introduced in Section 2.2.3: it corresponds to the vertical shift (that is, the liquidity shock) induced by the stress scenario. In terms of the model variables defined in Section 2.2, we have

$$\text{Liquidity at Risk} = S_2 - (C_1 - C_0 + \Delta C). \tag{2.24}$$

We note that:

- Liquidity at Risk is a conditional concept: it quantifies the expected total draw on liquidity resources of the bank *conditional* on the stress scenario being considered. In particular, the evolution of liquid balances and maturing liabilities constitute a part of this measure.

- Liquidity at Risk measures a *net outflow* corresponding to the stress scenario considered. This can be compared to the liquidity resources potentially accessible by the bank in the stress scenario, including feasible mitigating actions, to assess the potential for default.

- In contrast to the Liquidity Coverage Ratio (LCR), which is estimated based on historical data on margin calls or average run-off rates, Liquidity at Risk is a portfolio-specific and forward-looking concept: it quantifies the liquidity stress for a specific portfolio conditional on a scenario defined in terms of co-movements in risk factors.

The concept of Liquidity at Risk does not refer to a specific statistical model for generating risk scenarios. It may be applied to historical risk scenarios as well as hypothetical stress scenarios generated from a stochastic model for risk factors. In the case where one starts from such a statistical model for risk scenarios, one can

define a corresponding notion of Liquidity at Risk given a certain confidence level (for example, 99 percent Liquidity at Risk). However, in this work, we do not refer to a specific statistical assumptions about risk factors, and thus do not elaborate further in this direction.

2.4.2 Examples

We now illustrate the concept of Liquidity at Risk using two examples: a synthetic balance sheet and the balance sheet of a G-SIB.

2.4.2.1 A Synthetic Bank Balance Sheet

We consider a synthetic example of a bank balance sheet given in Table 2.4. Our example is representative of a typical balance sheet of a large commercial bank,[15] with a leverage ratio of 17.6. A large portion of the assets is allocated in a form of illiquid assets not subject to variation margin (mostly loans). Deposits are assumed to amount to 130,000 M EUR, which corresponds to 60 percent of other liabilities.

<table>
<tr><th>Assets</th><th>Liabilities and equity</th></tr>
<tr><td rowspan="2">Illiquid assets:
• Subject to VM, $I_0 = 16000$
• Not subject to VM, $J_0 = 134000$</td><td>Maturing liabilities, $S_0 = 18000$</td></tr>
<tr><td rowspan="2">Other liabilities, $L_0 = 215000$
(incl. deposits of 130000)</td></tr>
<tr><td>Marketable unencumbered assets:
• Subject to VM, $M_0 = 43000$
• Not subject to VM, $N_0 = 16000$</td></tr>
<tr><td>Liquid assets, $C_0 = 38000$</td><td>Equity, $E_0 = 14000$ (5.7%)</td></tr>
</table>

Table 2.4: A synthetic example of balance sheet for a representative large commercial bank (in millions of EUR).

The balance sheet is assumed to be sensitive to changes in interest rates and the equity market, as shown in Table 2.5. We consider two stress scenarios: a mild scenario of a +200 bps increase in interest rates and a −750 bps decrease in the equity market, and a severe scenario of a +100 bps increase in interest rates and a

[15] The synthetic balance sheet is motivated by the publicly available data on the 2016 JPMorgan Chase & Co. balance sheet.

−1,500 bps decrease in the equity market. Loans are taken to be highly sensitive to changes in interest rates, while changes in the equity market are taken to have little to no effect on the illiquid assets. Consequently, in our example, an increase in interest rates is mostly a solvency-type shock, whereas changes in the equity market leads to large cash outflows due to the presence of margin requirements.

	Risk factor	**Shift**	ΔI	ΔJ	ΔM	ΔN
Scenario I	Interest rates	+200 bps	400	4800	160	640
	Equity market	−750 bps	90	0	2150	400
Scenario II	Interest rates	+100 bps	200	2400	80	320
	Equity market	−1500 bps	180	0	4300	800

Table 2.5: Sensitivities of the synthetic balance sheet shown in Table 2.4. Values represent a decrease in the value of balance sheet components (in millions) in response to a shift in the risk factor under a mild (I) and severe (II) stress scenario.

In both scenarios, we assume that only 5 percent of unencumbered illiquid assets can be readily liquidated in a fire sale at a price discount of 50 percent, and no illiquid assets are eligible for a repo with the central bank. Furthermore, we assume a repo haircut of 32 percent with associated repo rate of 5 percent, and unsecuritised borrowing rate of 1 percent is available to the bank as long as its leverage ratio does not exceed the threshold $\delta = 20$. The money market benchmark rates are given in the scenario. The scheduled cash inflows in the example are taken to be 12,000 M EUR, while outflows are set to 10,000 M EUR. Finally, we assume a credit downgrade to trigger a severe depositor run-off of 58,000 M EUR (45 percent of total deposits).

Scenario I Consider a scenario defined by a +200 bps move of interest rates and an equity market drop of −750 bps. As a result of the initial shock, the bank's leverage ratio exceeds the creditworthiness threshold $\delta = 20$ and hence becomes downgraded. The Liquidity at Risk in this scenario is 76,800 M EUR: net cash outflows are comprised of 18,000 M EUR in maturing liabilities, reduced by 2,000 M EUR in net scheduled inflows, 2,800 M EUR in variation margin and 58,000 M EUR run-off due to downgrade. With an initial liquid assets buffer of 38,000 M EUR, the bank faces a liquidity shortfall of 38,800 M EUR, which needs

to be covered by 37,842 M EUR in new repurchase agreements with an associated cost of 1,892 M EUR and 958 M EUR in fire sales with an associated equity impact of 958 M EUR. As a result, the initial equity of 14,000 M EUR (5.7 percent) is reduced by the adverse shock to 7,360 M EUR (3.0 percent), and drops further to 4,510 M EUR (1.9 percent) due to incurred funding costs (Figure 2.8). Interactions between solvency and liquidity thus lead to a 43 percent loss amplification effect.

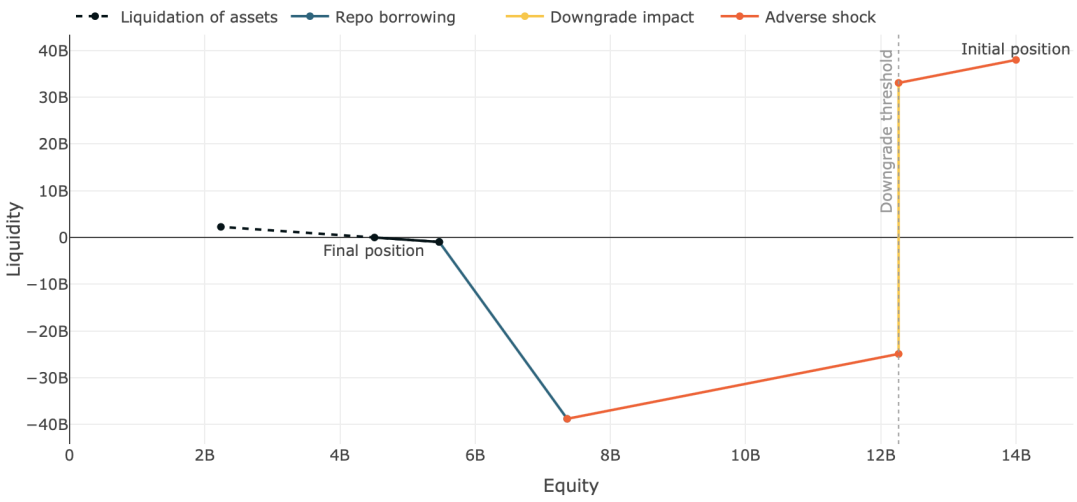


Figure 2.8: Solvency-liquidity diagram for the synthetic balance sheet shown in Table 2.4 in a stress scenario with +200 bps increase in interest rate and −750 bps equity market move.

Scenario II Consider now a severe market stress scenario defined by an interest rate increase of +100 bps and an equity market move of −1,500 bps. Similarly to the previous scenario, the bank is in a breach of the creditworthiness threshold of $\delta = 20$ that results in its downgrade. Consequently, the Liquidity at Risk for this scenario is 78,760 M EUR: the increase from the previous scenario is attributed to a larger variation margin outflow of 4,760 M EUR. This results in a liquidity shortfall of 40,760 M EUR, which exceeds the maximum funding capacity of 39,670 M EUR (36,380 M EUR in repo and 3,290 M EUR from assets liquidation), subsequently leading to a default. On the solvency side, the initial equity of 14,000 M EUR (5.7 percent) is reduced by the adverse shock to 7,720 M EUR (3.2 percent), and

drops further to 2,611 M EUR (1.1 percent) due to funding costs: 1,819 M EUR from repo and 3,290 M EUR from asset liquidation (see Figure 2.9). Interactions between solvency and liquidity thus lead to a 81 percent loss amplification effect.

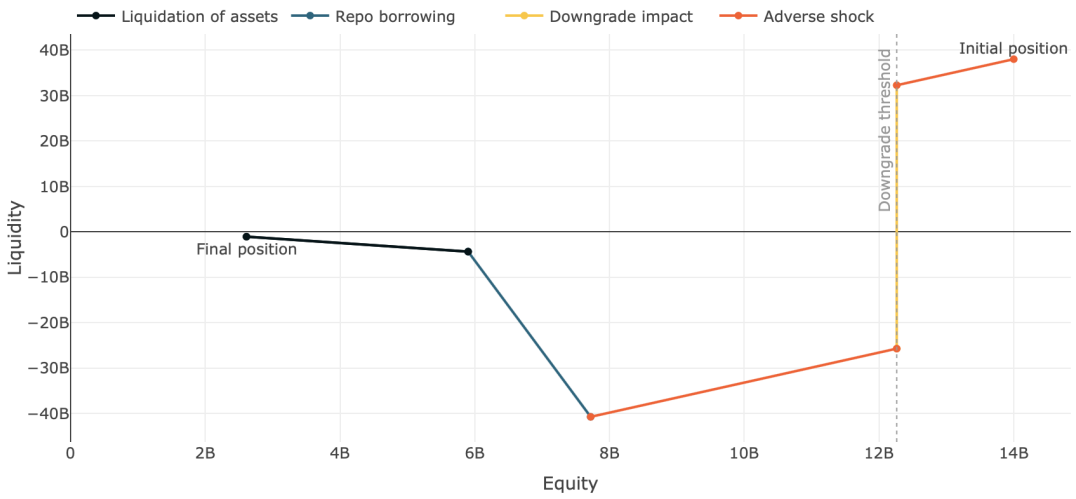


Figure 2.9: Solvency-liquidity diagram for the synthetic balance sheet shown in Table 2.4 in a stress scenario with +100 bps increase in interest rate and −1,500 bps equity market move.

Reverse stress testing So far we have discussed Liquidity at Risk in a single stress scenario. We can also use our approach to quantify solvency and liquidity impact across a range of adverse scenarios, parameterised by the severity of shocks to risk factors and identify "critical" shock amplitudes that potentially lead to insolvency or illiquidity. This "reverse stress testing" approach requires revaluation of balance sheet components under each scenario considered, but we may simplify this calculation considerably using a sensitivity-based approach, that is, by assuming a linear impact of risk factors on the portfolio components.

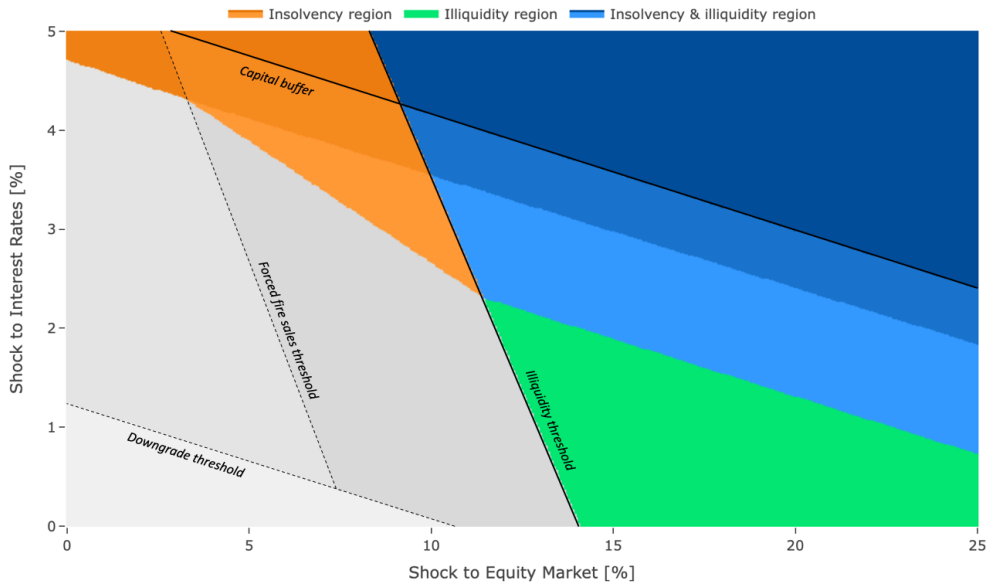
In the above example, this corresponds to using the sensitivities given in Table 2.5 to scale the impact of risk factor shocks on portfolio components across a range of amplitudes. Figure 2.10a summarises the impact of a shock on interest rates and equity of up to 5 percent and 25 percent, respectively. Our example illustrates a crucial point: the interaction of solvency and liquidity risk matters when

modelling default risk. Failure to incorporate it into a stress testing framework can significantly underestimate the total risk of a financial institution. An approach solely based on solvency risk would distinguish two regions in Figure 2.10a: a region of sufficient capital buffer (no failure) and a region of failure where loss of equity in a shock scenario exceeds the available buffer. Liquidity stress tests focus on sufficient liquidity buffers and the bank's ability to access sources of short-term funding in order to withstand adverse liquidity shocks. Consequently, independently conducted solvency and liquidity stress tests will fail to identify the regions where failure arises through the interaction of solvency and liquidity rather than through one channel alone, and thus will underestimate the risk of failure. These results are consistent with the observations in Schmitz et al. (2019), but push their conclusions further, showing that neglecting the solvency-liquidity nexus not only leads to the underestimation of solvency risk, but also of liquidity risk. The degree to which the credit risk is underestimated depends on the model parameters, balance sheet composition and sensitivities to risk factors.

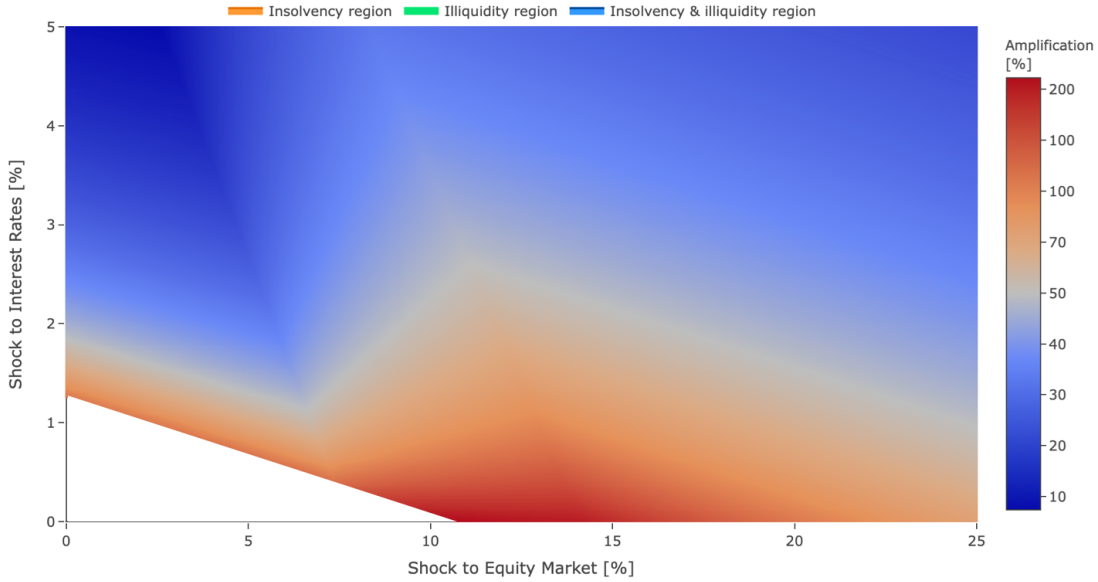
Joint modelling of solvency and liquidity also leads to a more accurate representation of the aggregate impact of a stress scenario on the firm's equity loss. Figure 2.10b illustrates the additional equity loss due to funding costs as a percent of the direct equity loss resulting from the shock to risk factors. For the balance sheet shown in Table 2.4, the initial equity loss may be amplified by up to two times through solvency-liquidity interactions. Consequently, our methodology is not simply a juxtaposition of two stress tests: it provides a consistent joint stress testing framework for solvency and liquidity, taking their interactions into account. Appropriate modelling of the solvency-liquidity nexus is essential to capture the effect of these interactions to provide a more accurate stress testing framework.

The amplification effect is non-linear. When a financial institution defaults due to illiquidity, the cost of new funding is at its maximum: a firm will attempt to repo and liquidate all its eligible assets to cover the liquidity shortfall, incurring a new, higher funding cost. Beyond this point, the amplification effect decreases for larger shock sizes because the funding cost no longer increases, as the bank no longer can increase the volume of its new funding, while the initial equity shock grows. In fact, for larger shocks the funding cost can decrease, since a shock can reduce the mark-to-market value of assets eligible for repo and sale, effectively reducing the

total volume, and hence also the cost of new funding. On the other hand, for small shock sizes, the amplification remains small, as the firm can obtain new funding for a relatively low price. The amplification effect increases significantly beyond the fire sale threshold, which tends to be an extremely costly way of managing liquidity. Furthermore, the amplification effect increases around the downgrade threshold with an increase in the size of credit-sensitive funding.



(a) Insolvency and illiquidity regions.



(b) Equity loss amplification due to funding costs (on a $\log_{10}$ scale).

Figure 2.10: Multiple stress test scenarios for a synthetic balance sheet, shown in Table 2.4, using linear extrapolation of sensitivities shown in Table 2.5.

2.4.2.2 A G-SIB Example

Let us return to the G-SIB example from Section 2.3.2. Recall that Table 2.3 shows a simplified view of the consolidated balance sheet data for a G-SIB with a leverage ratio of 17.9, and whose sensitivities to two key risk factors are shown in Table 2.6.

Similarly to the synthetic balance sheet example, we assume that in a stress scenario only 5 percent of unencumbered illiquid assets can be liquidated in the short-term with an associated 50 percent fire-sale discount, and no illiquid assets are eligible for a repo with the central bank. Funding through repo at a 5 percent rate requires a 32 percent haircut, while unsecuritised borrowing at a 1 percent rate is available up to the downgrade threshold of $\delta = 20$.

Risk factor	**Shift**	ΔI	ΔJ	ΔM	ΔN
Interest rates	+200 bps	1250	17000	2100	3600
Equity market	−750 bps	3900	0	4200	4600

Table 2.6: Balance sheet sensitivities for the balance sheet shown in Table 2.3. Values represent a decrease in the value of balance sheet components (in millions EUR) in response to a shift in the risk factor.

We subject this balance sheet to a stress scenario (Scenario I in Section 2.4.2.1) defined by an interest rates move of +200 bps, and an equity price move of −750 bps. We estimate stressed run-off rates contingent on a downgrade scenario calibrated on real crisis cases in line with the 2019 ECB Sensitivity Analysis of Liquidity Risk under the extreme scenario. We apply differentiated run-off rates across different types of deposits to the liability structure of the G-SIB. This results in an aggregate 55 percent run-off rate of the aggregate customer deposit base (224,950 M EUR). The impact of this stress scenario can be represented through a solvency-liquidity diagram, shown in Figure 2.11.

The effect of the scenario on the bank's net worth raises financial leverage, leading to an increase in the probability of default, and triggering a credit rating downgrade. Liquidity at Risk conditional on this scenario equals 248,400 M EUR: 11,450 M EUR payable in variation margin, 37,000 M EUR due to maturing liabilities, 101,000 M EUR due to SCO reduced by 126,000 M EUR from SCI, and

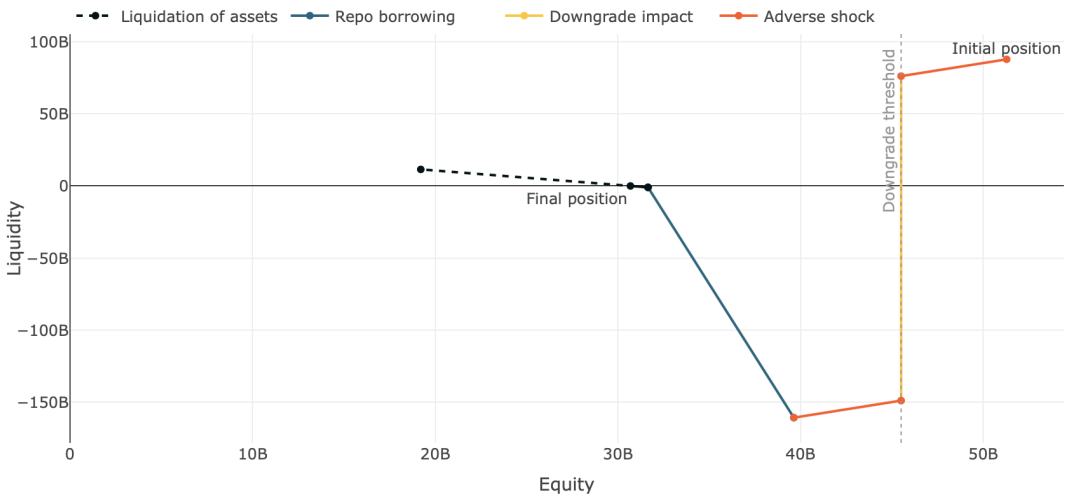


Figure 2.11: Solvency-liquidity diagram for the G-SIB balance sheet shown in Table 2.3 in a stress scenario with +200 bps interest rates and −750 bps equity market moves.

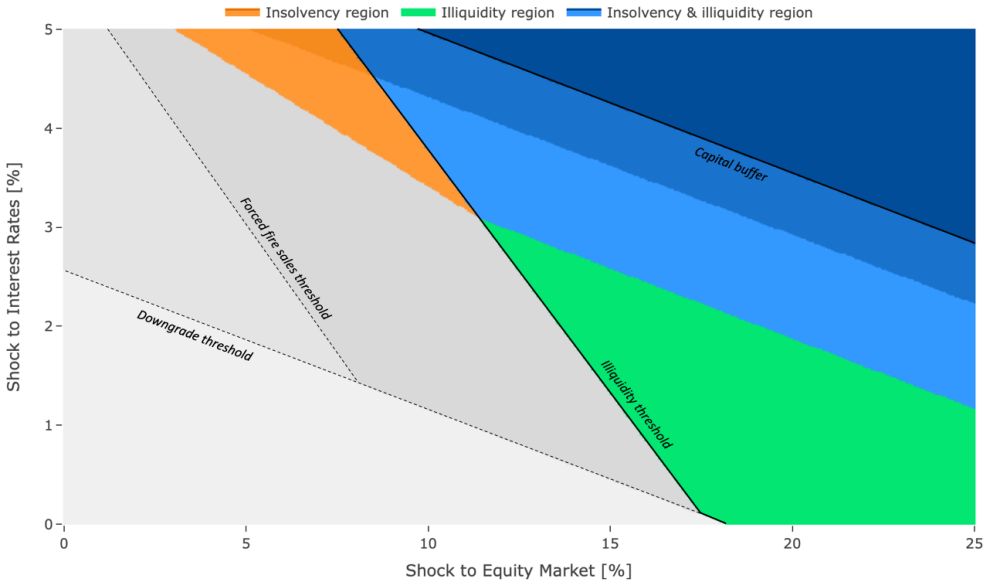
224,950 M EUR due to the run-off on deposits. This exceeds the bank's liquidity buffer of 87,775 M EUR, and results in a liquidity shortfall of 160,625 M EUR, which can be fully covered by borrowing in secured markets (159,662 M EUR for a cost of 7,983 M EUR) and through liquidation of assets (962 M EUR with an equity impact of 962 M EUR). As a result, the initial equity of 51,271 M EUR (5.6 percent) is reduced by the adverse scenario to 39,621 M EUR (4.4 percent), and drops further to 30,675 M EUR (3.4 percent) due to incurred funding costs, leading to a 77 percent loss amplification effect.

Comparing this example with the synthetic portfolio in Section 2.4.2.1, we observe that the same stress scenario applied to risk factors leads to *different* liquidity shocks to the balance sheet. The resulting liquidity shocks are endogenous and strongly dependent on balance sheet composition, funding structure, and bank resilience. This is quite different from the current practice of applying exogenous liquidity stress scenarios in liquidity stress tests (European Central Bank, 2019).

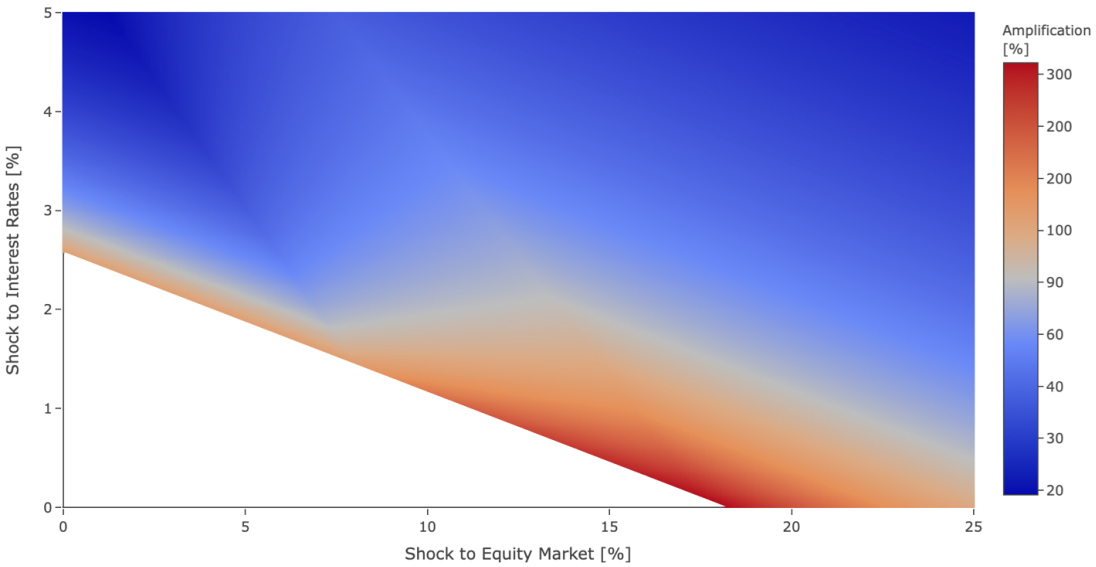
Reverse stress testing Using the same linear impact assumptions as in the previous example, we can extrapolate this analysis to other scenarios obtained by scaling the shocks to risk factors. The corresponding outcomes are represented in

Figure 2.12a. Under the severe depositor run-off assumption of 55 percent we see that liquidity risk becomes a major component of the default risk for large shocks.

A non-linear amplification effect emerges due to the interaction between solvency and liquidity, as discussed previously. The loss amplification becomes most significant for large equity market moves of about 1,800 bps with no moves in interest rates – in this case, the initial loss is amplified by more than 350 percent due to the high cost of new funding. On the other hand, for small shocks sizes that do not lead to a downgrade, the bank holds a sufficient amount of liquidity buffer to cover its liquidity at risk in full, and thus we do not see any loss amplification effect. It should be noted that the loss amplification effect crucially depends on the funding sources that a bank is able to tap under stress.



(a) Insolvency and illiquidity regions.



(b) Equity loss amplification due to funding costs (on a $\log_{10}$ scale).

Figure 2.12: Multiple stress test scenarios for a G-SIB balance sheet, shown in Table 2.3, using linear extrapolation of sensitivities shown in Table 2.6.

2.5 Concluding Remarks

Financial crises have repeatedly confirmed that the lack of liquidity is an inherent risk throughout the banking sector (Pohl, 2017). Liquidity and solvency are two interrelated dimensions of credit risk that cannot be modelled, or stressed, separately. Nonetheless, the interaction between liquidity and solvency tends to be omitted in stress testing practices. In response to calls from regulators to develop integrated liquidity and solvency stress tests (Basel Committee on Banking Supervision, 2015), we have developed a coherent framework for the joint stress testing of solvency and liquidity risk.

In our framework, solvency shocks affect liquidity through margin requirements, via a firm's ability to raise short-term funding, and through credit risk sensitive outflows. Consequently, this leads to endogenous liquidity shocks. In turn, solvency stress is exacerbated through the cost of new funding resulting from a liquidity shortfall, and fire sales. We distinguish between two types of failure: financial institutions can become illiquid without being insolvent, insolvent while remaining liquid, or – in the case of extreme shocks – both illiquid and insolvent. The model illustrates the danger of underestimating credit risk by models that do not account for the solvency-liquidity nexus. As shown by our examples, balance sheet composition has a significant effect on the solvency-liquidity nexus. In particular, our insights show that structural solvency risk models are insufficient to capture this dependency and we advocate the use of a more granular balance sheet view by the regulators when conducting a stress test.

Our proposed framework provides a more realistic stress test framework which establishes coherence between the design of solvency and liquidity stress tests. It also includes mitigating actions that can be extracted from the bank's contingency funding plan and recovery plan. By defining the concept of Liquidity at Risk, we provide a tool to quantify the total draw on liquidity resources of the bank conditional on the stress scenario defined directly in terms of an adverse shock to risk factors. Sudden liquidity stress can result in the inability to obtain sufficient funding in due time and can lead to insolvency.

The tool is calibrated using available regulatory templates on financial data, risk data, and liquidity monitoring templates. The model is amenable to reverse

stress testing and naturally permits a range of sensitivity tests around crucial inputs including changes to the classification of fair valued instruments, the liquidity generation capacity of unencumbered securities, the evolution of market haircuts and funding costs, and fluctuations in creditors' risk appetite framework.

The model yields useful policy implications for central banks and supervisory authorities. It helps supervisors to identify whether managerial options to fend off liquidity risk are helpful in avoiding insolvency or illiquidity in plausible stress scenarios. It also enables the identification of sources of systemic spillovers, that is, shocks to risk factors that can become a conduit of systemic risk propagation and can threaten financial stability. Crucially, it helps authorities to make better decisions regarding the provision of central bank emergency liquidity assistance (ELA) to "illiquid but solvent" financial institutions. Ultimately, it serves to quantify the amount of resolution funding, which remains perhaps the key likely impediment in banking resolution.

Chapter 3

Network Analysis of the UK Reinsurance Market

Chapter based on:
Artur Kotlicki, Andrea Austin, David Humphry, Hannah Burnett, Philip Ridgill, Sam Smith. Network Analysis of the UK Reinsurance Market. Staff Working Paper, Bank of England, forthcoming.

In this chapter, we provide an empirical analysis of the network structure of the UK reinsurance sector based on Solvency II regulatory data. We examine counterparty credit risk originating from reinsurance contracts as a source of financial contagion in the insurance industry. The granularity of the Solvency II data provides a new opportunity for detailed analysis of the actual exposures in the system, detection of potential systemic vulnerabilities, and reinsurance spirals. In our multi-layered network approach, we incorporate information on reinsurance contract risk types and ownership structure for both life and non-life insurers.

Our findings suggest that the UK reinsurance sector exhibits the 'small-world' property with a scale-free, core-periphery structure and topological characteristics common to other financial networks. These characteristics of risk dispersion from the periphery to the core make

the network 'robust-yet-fragile' to financial shocks. We explore the robustness of the network to adverse shocks through a stress-simulation exercise, where we find it robust to system wide shocks affecting the value of total investments, and to idiosyncratic shocks applied to large, highly interconnected reinsurers.

3.1 Introduction

The financial crisis of 2008 sparked a plethora of regulatory responses and new research on systemic risk in financial systems. Despite the important role played by reinsurers in the economy, the topic of contagion and systemic vulnerabilities in the reinsurance sector has received relatively little attention in literature. This is especially surprising given how the level of contagion in the system is highly sensitive to the modelling assumptions and network topology, as corroborated by numerous studies on simulated network structures (Klages-Mundt and Minca, 2020; Roukny et al., 2013; Nier et al., 2007; Battiston et al., 2012a). In particular, conclusions based on empirical studies of banking systems and their specific counterparty exposures (Cont et al., 2013; Caccioli et al., 2015; Boss et al., 2004) may not hold for insurance markets. The recently enacted Solvency II regulatory framework, which in part promotes enhanced data disclosures, thus provides us with a unique opportunity to explore the interconnectedness of the UK insurance sector in much detail, and to study its inherent systemic vulnerabilities under stress scenarios.

In this chapter, we aim to utilise network analysis to assess the potential for contagion and systemic risk in the UK insurance market stemming from reinsurance contracts. In particular, we use the unique and confidential Solvency II data set from 2016,[1] provided by the Bank of England, to examine the network structure of the UK insurers' reinsurance contracts, and in consequence to cast light on the intricate nature of the interconnections, whether risk is dispersed or concentrated in the system, and the resulting implications for financial stability using a simulation-based stress testing approach.

[1]The Solvency II Directive was adopted in 2016. As such, the data featured in this thesis comprises of the first Solvency II data submission to the Bank of England.

Our study builds on the growing literature of network analysis applied to the financial sector which gained impetus following the 2008 financial crisis. Network models provide an adequate framework for assessing the potential for losses to spread in a financial system following counterparty default (Cont et al., 2013). In particular, the topology of networks affects their vulnerability to the risk of contagion (Allen and Gale, 2000; Caccioli et al., 2012; Roukny et al., 2013). A nonlinear relationship between the interconnectedness and the stability of financial markets, characterised as the 'robust-yet-fragile' property, has been well documented in the literature on financial contagion (Gai and Kapadia, 2010; Gai, 2013; Acemoglu et al., 2015; Caccioli et al., 2015) – while increased network connectivity and diversification of exposures can reduce the likelihood of contagion, it can also amplify losses when contagion occurs. In other words, financial networks exhibit a phase transition with respect to adverse shocks: below a certain threshold, losses are attenuated through risk-sharing and diversification practices, while beyond this tipping point, interconnectedness provides the means for the contagion to spread, amplifying the financial stress experienced in the system.

The literature on financial networks highlights the importance of key network characteristics for network stability and the global level of systemic risk (Amini et al., 2012; Roukny et al., 2013; Nier et al., 2007; Battiston et al., 2012a). Notably, financial networks can often be characterised by a heavy-tailed power law degree distribution of connections (Boss et al., 2004; Cont et al., 2013; Caccioli et al., 2015), and a core-periphery structure (Chen et al., 2020; Fricke and Lux, 2015; Craig and von Peter, 2014; Caccioli et al., 2018). As such, the network structure is far from being random, where strong hierarchical relations are indicative of firms' preferences when choosing their counterparties to conduct business. In particular, a core-periphery structure is characterised by a set of highly connected hub nodes in the core that intermediate connections between peripheral nodes, where nodes in the periphery tend not to connect with each other. Barabási and Albert (1999) show that the emergence of a core-periphery structure in real-world networks can be attributed to preferential attachment, where new nodes in the network are more likely to form a connection with already well-connected connected nodes in the system.

Nodes in the core of the network play an important role from the systemic risk perspective: through their diversification and risk-sharing practices, hub nodes aid in the mitigation of losses following a counterparty default. However, they also render the system more vulnerable to their own default, as the resulting financial distress can propagate to a large part of the system, triggering systemic defaults. Consequently, networks exhibiting the core-periphery structure are shown to be resilient to the failure of a random node, but are vulnerable to defaults of systemically important hub nodes (Caccioli et al., 2012; Roukny et al., 2013). In contrast to the 'too-big-to-fail' theory, which considers only the absolute size of an institution, network theory is more suited for an identification of systemically important nodes that are 'too-central-to-fail' (Battiston et al., 2012b). In particular, simulation studies provide useful tools in assessing the robustness and stability in financial markets, as well as identifying systemically important nodes whose default can trigger a cascade of losses.[2]

Attention has only turned in recent years to the role of the insurance sector in creating or propagating systemic risks. Insurance risk is generally considered 'uncorrelated' by design. This means that the occurrence of an insured event does not increase the likelihood of a separate insured event happening. This is in strong contrast to issues related to credit risk, for instance in bond insurance[3] and CDS contracts, where the inability of an entity to pay their liabilities makes it more likely that other entities within the system will also be unable to pay theirs. This results in defaults being much more correlated across the financial system, and therefore rendering the system more susceptible to a systemic event. Furthermore, insurers enable risk transfer in the economy through accepting and pooling risks, and through providing long-term savings products. The subsequent investment of premiums by insurers in financial assets can make them susceptible to systemic risk – risk management practices such as balance sheet diversification are thus a core activity for insurers. An example of which involves purchasing reinsurance,

[2]Refer to Upper (2011) for a detailed survey on the work in this topic.

[3]Insurance of credit risk, such as the risk of municipal bonds is a notable exception. Prior to the financial crisis in 2018 several municipal bond insurers also provided credit guarantees for mortgage-backed securities, creating an exposure to the housing market; see Saporta (2016).

where the reinsurer accepts part of the insurance losses of the insurer in return for a premium.[4]

In 2016 the International Monetary Fund described two channels through which systemic risk might spread through the insurance sector. One channel is through counterparty default – a 'domino' view, where the failure of one insurer triggers the failure of others (International Monetary Fund, 2016). Factors that affect how this channel operates in practice include the size of the insurer, its interconnectedness, whether its activities can be substituted by other insurers, its leverage, its funding liquidity risk, and the complexity of its operations. An often cited example, albeit caused by products not traditionally associated with insurance, is that of the near-failure of American International Group (AIG), which received support from the US government due to the potential default of its CDS counterparties (McDonald and Paulson, 2015). Insurers face counterparty credit risk from their reinsurers, creating the potential for one 'domino' – a reinsurer – to knock over several others if it is unable to pay its reinsurance claims.

The other channel for systemic risk, which does not rely on firm failure, is the 'tsunami' view – capital weaknesses can stop insurers from performing the role of taking on risk during crises. This could impact economic activities such as the availability of insurance cover and the availability of funding to the economy through investments,[5] which could further contribute to financial market turbulence through pro-cyclical investment behaviours (Ellul et al., 2015).

Insurers carrying out the traditional role of risk transfer from perils or mortality tend to be regarded as not posing systemic risk, at least from the perspective of the domino view. However, reinsurance stands out within the insurance business model as a source of counterparty credit risk. This type of transaction creates the risk of direct contagion – a domino effect – between insurers. The generally accepted view in the literature is that the use of reinsurance in practice does not,

[4]For instance, Bäuerle and Glauner (2018) present a model in which a single reinsurer providing excess of loss reinsurance contracts is able to reduce the value at risk for a group of insurers.

[5]In a recent study, Malik and Xu (2017) examine the interconnectedness among global systemically important banks and global systemically important insurers for US, European and Asian regions in the period of 2007 to 2016.

however, give rise to systemic risk (IAIS, 2011, 2012; French et al., 2015). This is because of five main factors:

1. Insurers only cede a small proportion of their total reinsured risks to each reinsurer, so losses from individual failures are more likely be absorbed by the insurer.

2. Collateral may be posted by reinsurers to mitigate the effect of their failure.

3. Insurers make conscious choices to cede certain risks that allow for greater diversification benefits and improved capital management (therefore not contributing to the build-up of risk in the system).

4. The links between insurers are hierarchical and do not give rise to feedback loops. For example, the International Association of Insurance Supervisors noted in 2011 that 'while primary insurers link to reinsurers, interlinkages among primary insurers are comparatively limited. In other words, links between entities in the insurance market are almost entirely hierarchical, and there is no network-like inter-insurance market similar to the interbank market. [...] As a result, there are fewer feedback mechanisms to create non-linearity and a potential for systemic risk within the insurance sector.'

5. The timing difference for liquidity between insurance and banking makes counterparty exposures less problematic. Insurance claims are paid out over a much longer period of time, possibly years, than, say, the interbank loan market, which may require settlement overnight. This gives reinsurers in distress more opportunity to recover before its ceding insurers are affected.

This consensus is supported by historical experience, where less than 4% of all impaired insurers failed as a result of the failure of a reinsurer (IAIS, 2012). There is, however, one notable exception – the London Market Excess of Loss spiral, which took place in the late 1980s and inadvertently saw larger losses arise than anticipated because of retrocession (Bain, 1999). In this case, the reinsurance contracts between insurers and syndicates within the Lloyd's of London Market saw syndicates cede risk only to accept that same risk through another contract. Such retrocession spirals aggregate losses from multiple parts of the market in an

opaque manner, potentially leaving the dampening reinsurer with extreme losses and thus destabilising the entire system (Klages-Mundt and Minca, 2020). Since then, the use of reinsurance and retrocessions by syndicates has changed a great deal, with lower proportions of reinsurance, higher levels of retained risk, and other measures put in place by the Society of Lloyd's to identify these risks.

Empirical research on financial contagion from reinsurance, whilst limited in scope,[6] finds no evidence of systemic risk due to counterparty defaults in the insurance market (that is, the 'domino' view). Given the opacity of the reinsurance industry and scarcity of data, most recent studies focus on the property-causality reinsurance market in the United States (Park and Xie, 2014; Chen et al., 2020; Klages-Mundt and Minca, 2020).

In particular, Park and Xie (2014) consider two potential contagion mechanisms in the period of 2003–2009: the direct counterparty risk due to failure of top reinsurers, and an indirect information-based effect of a reinsurer's downgrade. The latter effect contributes to loss spill-overs even to insurers with no direct exposure to the downgraded reinsurer. However, the worst-case scenario in the simulation study, comprising of a failure of the top reinsurer group, had a minor effect on the solvency of insurance market participants. Therefore, despite the strong interconnectedness of the system, the likelihood of systemic risk triggered by reinsurance contracts is found to be small.

Chen et al. (2020) extend the study of interconnectedness within the US property-casualty reinsurance market with more extensive and granular data for both insurers and reinsurers. The contagion mechanism includes both counterparty exposures due to reinsurance premiums paid and reinsurance recoverable amounts outstanding. In a simulation analysis, the authors consider both the failure of individual reinsurers – the domino view – and the potential amplification of systemic stresses emanating from asset markets via the default of reinsurers because of deterioration in their financial assets. Similar to the previous studies,

[6]Research in this field is often hindered by data limitations: parts of the market remain opaque to both regulators and market participants (Davison et al., 2016). In particular, disclosures on retrocession agreements are scarce in contract details, making estimation of counterparty exposures difficult in practice. We refer to Section 3.3.3 for a discussion on Solvency II data limitations in the context of identification of network cycles.

they do not observe widespread insolvencies in the industry even during very severe stress scenarios.

An empirical analysis by Lelyveld et al. (2011) examines the resilience of Dutch insurers to failing reinsurance covers in the period of 2003–2005. On the individual institution level, no default cascade occurred in their scenario analysis. However, there was a risk to entities within a group from intra-group reinsurance contracts.

Kanno (2016) assesses the systemic importance of insurers in the global non-life insurance market during 2006–2013. Using the Eisenberg and Noe (2001) approach for the allocation of losses, the author finds systemic risk to be relatively restricted: only a few contagious defaults were observed under a severe initial shock to the economy. The network analysis of contagious defaults was limited, however, due to the use of aggregate data on reinsurance transactions.

A noteworthy exception is the work of Klages-Mundt and Minca (2020), in which the authors caution that the risk of contagion may be underestimated because of the potential for unexpectedly large losses from non-proportional reinsurance. They emphasize the existence of complex interactions between insurance losses and counterparty default, which are not captured adequately by the existing contagion models (Eisenberg and Noe, 2001; Acemoglu et al., 2015; Elliot et al., 2014). In particular, non-linearities from excess of loss contracts are shown to obfuscate risks, rendering the system vulnerable to excess costs from network effects. However, the methodology of Klages-Mundt and Minca (2020) is extremely sensitive to model parameters. In their simulations made on the 2012 US property-casualty reinsurance data, small perturbations in the estimated parameter set are shown to affect key players of the market in a substantial manner. As a result, this methodology is not well suited to the current data regime, where the required contract details are not available and instead need to be estimated using common rules of thumb.

Similarly, Davison et al. (2016) emphasise the role of market opacity in a possible emergence of reinsurance spirals that can increase the risk of contagion due to concentration of losses. However, their simulations of the frequency, severity, and patterns of exposure to loss find that reinsurance networks are robust to plausibly large losses. That said, different patterns of retrocession and large haircuts to recoverable amounts can influence the levels of contagion. These results should be

taken with caution however, as the data on the size and connections of the network is not available to the authors, who resort instead to testing several plausible networks based on a sample of large global reinsurers.

The differentiating feature of our network analysis of the insurance market is our ability to measure the interconnectedness directly through the use of Solvency II regulatory returns that contain details on reinsurance contracts between insurers, including premiums, sums insured, and amounts recoverable. It is the first study of the UK insurance sector of this kind, and unlike previous studies that have focused on property-causality underwriters, our work incorporates both life and non-life insurance contracts. In addition, our analysis includes syndicate-level data from the Lloyd's of London market to give a detailed view of the interconnectedness of the London insurance market. We use line of business data to identify the prevalence of cycles of risk transfer within the network.

Our simulation-based stress test analysis focuses on a direct contagion through exposures via reinsurance contracts, and incorporates bankrupcy costs. Similarly to Chen et al. (2020), we consider both the contagion impact of a network wide shock and the impact of default of the most connected insurers. Our results show the UK reinsurance market to be robust even under extreme stress scenarios. We also show that the network topology shares common characteristics with other financial networks, and in particular can be characterised by a scale-free, core-periphery structure.

The remainder of this chapter is organised as follows: in Section 3.2 we introduce data sets considered in our analysis, we describe the adjustments made to them, and provide detailed description of the methodology used in our network analysis. Section 3.3 then presents the main results of our the network analysis, including the topological characterisation of the UK reinsurance market, and the observed nuances which affect its vulnerabilities to contagion risk. We also discuss the contribution of this study as a tool for detection of possible retrocession spirals. Section 3.4 presents simulation analysis to assess the degree of systemic risk and potential for network contagion due to direct counterparty risk. We present our conclusions in Section 3.5.

3.2 Data Sets and Methodology

We use for our study Solvency II data on reinsurance contracts submitted to the Prudential Regulation Authority (PRA) by authorised insurance companies and Lloyd's of London syndicates, covering a single year, recorded at year-end 2016.

We note that for simplicity, throughout the study we use the term *insurer* to refer to the company that is ceding risk, but also to refer to insurance companies generally when we do not want to specifically refer to reinsurers. We use *reinsurer* to refer to the company that is accepting the risk. The same company may be referred to as either an insurer or a reinsurer within this study depending on the context of the discussion. We also do not make an explicit distinction between reinsurance and retrocession (that is, reinsurance of reinsurance business).

The data submitted to the PRA is non-public, and thus in this thesis we only present anonymised or aggregated results. We identify an insurer with the following characteristics where relevant for the analysis: whether the insurer is an insurance company or a Lloyd's of London syndicate, where the insurer is a life or non-life insurer, and whether the insurer is an insurance group.

3.2.1 Data Overview

For the purposes of network analysis, we construct two data sets for year-end 2016: one on the nature of contracts in place (referred to as the *treaty and facultative data*), and the other on the amounts recoverable (referred to as the *recoverables data*), with metadata about line of business and risk description added to allow for layering of the network and more in depth analysis of the results.

Treaty and facultative data The first set of data includes information about facultative and outgoing treaty reinsurance programmes by contract as of year-end 2016. *Facultative reinsurance* is transacted on an individual risk basis, where the ceding company has the option to offer individual risks to the reinsurer and the reinsurer retains the right to accept or reject the risk. *Treaty reinsurance* on the other hand is a transaction encompassing a block of the ceding company's book of

business. The reinsurer must accept all business included within the terms of the reinsurance contract.[7]

Our data on facultative and treaty reinsurance contracts (henceforth 'treaty and facultative data') captures reinsurance contracts with individual insurers, where a UK insurer has ceded risk to another (not necessarily UK-based) insurer. It covers the identity of the reinsurer, the sum insured, the premium, the line of business, and a description of the risk insured. Furthermore, the data records whether the contract is proportional, such as a quota share, where the reinsurer has a liability for a percentage of the loss, or non-proportional, such as excess of loss, where the reinsurer has a liability for losses that exceed a certain amount. Other information in the data includes the geographic location of the reinsurer, and whether the reinsurer is external to the insurer's group of companies, or internal to it, such as a subsidiary or a captive reinsurer.

Concerning the line of business, Solvency II requires insurers to identify contracts as one of eight lines of business for life and one of 28 lines of business for non-life. We aggregate the treaty and facultative data into fewer lines of business based on similarity of the risks they cover.[8] 'Fire and damage to property' is the most common type of non-life reinsurance, while 'multiline' contracts, which cover multiple risks, are the ones with the highest premium ceded on average (in terms of median value). The 'other life' category, which includes reinsurance contracts against longevity risk, is the most frequent type of life reinsurance contract. 'Unit-linked or index-linked' reinsurance, which combines insurance and investment into a single integrated plan, and 'other life' reinsurance are the lines of business that have the highest premium ceded for the life market. Figure 3.1 provides a quantitative summary of premiums ceded by each line of business in the treaty and facultative data set. Reinsurance contracts are seen to be highly heterogeneous, with a small number of contracts reporting extreme values of exposures. Such heterogeneity in counterparty exposures is a common feature of financial networks,

[7]See Munich RE (2010) for a concise introduction to facultative and treaty reinsurance.

[8]The aggregated lines of business are: fire and other damage to property; marine, aviation and transport, general liability, motor, non-life annuities; credit and suretyship, health, medical expense; multiline; other non-life; life; with-profit participation; unit-linked or indexed-linked; other life.

as emphasised in the past studies on banking systems (Caccioli et al., 2012; Cont et al., 2013; Boss et al., 2004).

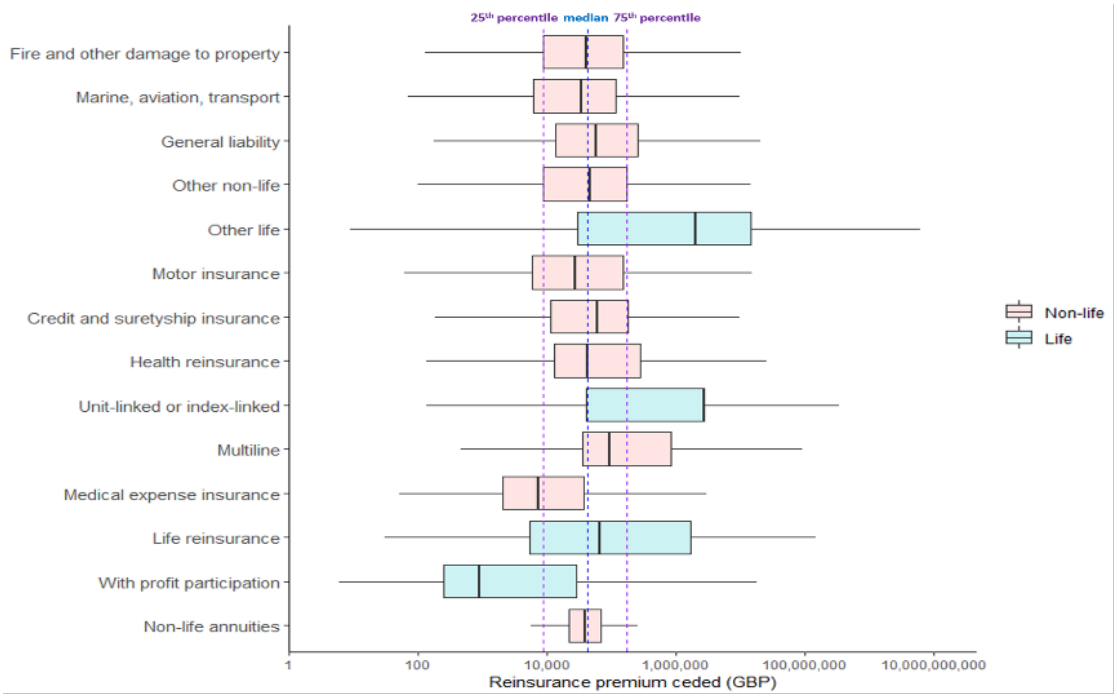


Figure 3.1: Box-plot of treaty and facultative premiums by line of business and by contract type. Blue dashed lines reflect distribution of all UK reinsurance premiums ceded. Values in GBP on a $\log_{10}$ scale.

The treaty and facultative data set includes 799 reinsurers and 41,883 contracts (82% of UK contracts). We do not have sight of all facultative contacts because as part of Solvency II reporting insurers are only required to disclose the 10 largest contacts by exposure for each line of business.[9] Consequently, over 93% of the treaty and facultative data come from treaty contracts (by number of contracts). Treaty contracts are in general more valuable than facultative contracts in terms of premiums ceded, with more than 99% of UK premiums attributed to treaty reinsurance.

Whilst there are fewer life contracts than non-life contracts, they account for a much higher proportion of premiums ceded – out of the observed £107 billion

[9]The number of disclosures is sometimes lower than 10 per insurer, indicating that the insurer in this case has less than 10 contracts in total per line of business.

in premiums ceded, £69 billion is life and £38 billion is non-life. This is because unit-linked life insurance, which is mainly a type of investment product, can be organised as a reinsurance contract by the insurer. Non-proportional contracts, such as excess of loss, are more common than proportional contracts, accounting for more than 81% of all contracts in the data set. This feature tends to follow the life and non-life distinction – life contracts tend to be proportional (87% of contracts), while non-life contracts tend to be non-proportional (84% of contracts). As a result, although non-proportional contracts are more common, they account for a lower share of premiums ceded. Most reinsurance contracts are external to the insurance group (93% of the insurance contracts) but internal contracts account for a large share of premiums ceded (81% of UK premiums). In particular, our data contains a few intra-group contracts with exceptionally large premiums ceded that significantly skews the data. Notably, values of these contracts, although unusually large, are confirmed on an individual contract basis with industry experts to be valid and hence are not excluded from the analysis.

We refer to Appendix A.1 for a detailed breakdown of the number and value of reinsurance contracts by their type and line of business.

In general, premiums ceded by contract follow a positively skewed distribution. The mean is greater than the median: that is, there is a minority of high-value contracts that pull-up the mean. This characteristic is maintained whether we look at all contracts together, or by other categorisations, such as life and non-life contracts separately (see Table 3.1). We again remark that extreme values of premiums ceded were put under scrutiny through consultation with subject matter experts, and here we only consider reinsurance contracts that have been verified. Consequently, the observed variability is a result of inherent heterogeneity of the insurance market.

	All	**Life**	**Non-life**
Mean	2,571,399	34,816,994	951,842
Standard deviation	62,452,612	254,020,332	28,320,323
25^{th} *percentile*	8,634	17,589	8,437
Median	43,125	1,383,006	40,893
75^{th} *percentile*	174,137	7,950,774	152,164

	Proportional	**Non-proportional**	**Groups**
Mean	12,253,819	306,867	2,153,463
Standard deviation	143,282,276	2,939,960	51,695,191
25^{th} *percentile*	25,937	7,297	8,696
Median	282,799	34,125	42,000
75^{th} *percentile*	2,331,247	113,755	164,801

Table 3.1: Descriptive statistics for treaty and facultative contracts based on premium ceded. Values in GBP.

Recoverables data The second set of data, referred to as the 'recoverables data', includes current levels of reinsurance amounts recoverable from individual reinsurers by UK insurers as of year-end 2016. In particular, the data contains the total amount of recoverables per individual reinsurer, and may encompass the recoverables from multiple contracts and lines of business. By usual convention, we classify a recoverable amount from a reinsurer under life insurance if more than 50% of the technical provisions (that is, the liability amount) is for the life insurance business, and we classify as non-life otherwise.

This data set covers 22,713 separate recoveries by UK insurers from their reinsurers (88% of the UK total recoveries) and 3,560 reinsurers. The total technical provisions recoverable comes to £240 billion in 2016, of which life technical provisions account for 86% of this total. Netting off collateral, total recoverables amounted to £116 billion. As in the case of the treaty and facultative reinsurance data set, the distribution of recoverables is skewed, with some very high values of recoverables, mainly relating to life insurance (see Table 3.2).

	All	**Life**	**Non-life**	**Groups**
Mean	5,746,898	133,882,490	1,729,200	4,109,034
Standard deviation	159,434,728	865,804,686	30,237,584	118,965,451
25^{th} *percentile*	289	48,097	1,834	273
Median	13,761	1,287,026	26,567	13,186
75^{th} *percentile*	214,800	18,104,041	275,338	203,139

Table 3.2: Descriptive statistics from individual insurers of amounts of reinsurance recoverable per reinsurer (net of collateral). Values in GBP.

Data limitations Our data corroborates the view that reinsurance is a global industry. According to our treaty and facultative data, eight countries accounted for 95% of the reinsurance premiums ceded. The life market was more concentrated geographically with six countries accounting for 95% of the premiums ceded, while the non-life market was less concentrated (13 countries).

This brings to light a limitation of using these data sets for network analysis – each insurer submission only contains the reinsurance ceded by UK insurers, and does not identify where non-UK insurers have ceded risks to UK (re)insurers.

Quantifying bilateral exposures Instead of considering counterparty risk on an individual reinsurance contract basis, we are interested in quantifying the total bilateral exposures in the UK reinsurance market. In light of this, we aggregate multiple instances of unique reinsurance contracts between an insurer and the same counterparty.

In particular, we quantify the total bilateral exposure using the available additional data on the reinsurance contracts and reinsurance recoverables as follows.

- For the treaty and facultative data set, we use the information on the outgoing premium ceded and the amount of exposure ceded or sum reinsured to compute the aggregate value of risk transfer from an insurer to its specific reinsurer as of year-end 2016. When discussing insurance risks for a specific line of business (a particular network layer), we quantify the aggregate exposure using only the relevant contracts, and disregard other reinsurance contracts.

- For the recoverables data set, we use the value of the total reinsurance recoverable net of collateral to compute the aggregate counterparty default exposure from a reinsurer to its primary insurer. In particular, this value thus represents the maximal (short-term) loss to the insurer in case of an immediate default of its reinsurer counterparty as of year-end 2016.

We perform adjustments to the data prior to assigning counterparty exposure size to each market participant. We exclude missing weight values in the treaty and facultative data set and in the recoverables data (that is, blank fields). Contracts with a specified zero value may be contracts that start in a neutral position and whose value can change over time;[10] or may be the result of reporting error. Such contracts are excluded from our main network analysis as they do not represent a current exposure.[11]

The reporting templates for the treaty and facultative data allow insurers to record the sum reinsured or exposure ceded as -1 to represent unlimited liability, which are a potential source of high losses incurred by the reinsurer. In our study, we convert unlimited contracts amounts to a right-tail value (97.5%) of the reinsurance recoverables distribution, representing an extreme but plausible scenario. We also remove contracts with premiums or sums insured that are significantly negative from the treaty and facultative data set, as we were unable to verify the validity of these contracts.

We do not exclude negative values from the recoverables data set as these are likely to represent reinsurance contracts containing additional performance clauses, which indicate that the insurer owes the reinsurer a payment. From the perspective of counterparty risk contagion, which relates to the failure of payment made by a reinsurer to its counterparty, it is important to include these contracts in the network. When computing the aggregate exposure, we include the absolute value of these contracts, and for any reinsurance contract with negative values we reverse the role of the insurer and its counterparty.

[10] For instance, some longevity reinsurance contracts have this feature.

[11] In our network analysis, we also computed unweighted topology statistics of the network that includes contracts with a value of £0, and find that these statistics are not materially different to those for the weighted network.

In particular, let e_{ij}^r represent the counterparty exposure (the value of the total reinsurance recoverable net of collateral) of insurer v_j to reinsurer v_i specified by a contract $r \in \mathcal{R}$, where $\mathcal{R}$ represents the set of all reinsurance contracts in the recoverables data. The aggregate exposure w_{ij} of insurer v_j to reinsurer v_i is then given as

$$w_{ij} = \sum_{r \in \mathcal{R}} e_{ij}^r \, \mathbb{1}(e_{ij}^r > 0) - \sum_{r \in \mathcal{R}} e_{ji}^r \, \mathbb{1}(e_{ji}^r < 0), \tag{3.1}$$

where $\mathbb{1}(\cdot)$ is the indicator function. An analogous formula follows for computing the aggregate value of risk transfer using the treaty and facultative data set. In particular, denote by $\mathcal{S}$ the set of all reinurance contracts in the treaty and facultative data set. For a contract $s \in \mathcal{S}$, let e_{ij}^s represent the the outgoing premium ceded and the amount of exposure ceded or sum reinsured. Since in the observed data we have $e_{ij}^s \geq 0$, for all $s \in \mathcal{S}$, the aggregate value of risk transfer from insurer v_i to reinsurer v_j can be defined as

$$w_{ij}^* = \sum_{s \in \mathcal{S}} e_{ij}^s. \tag{3.2}$$

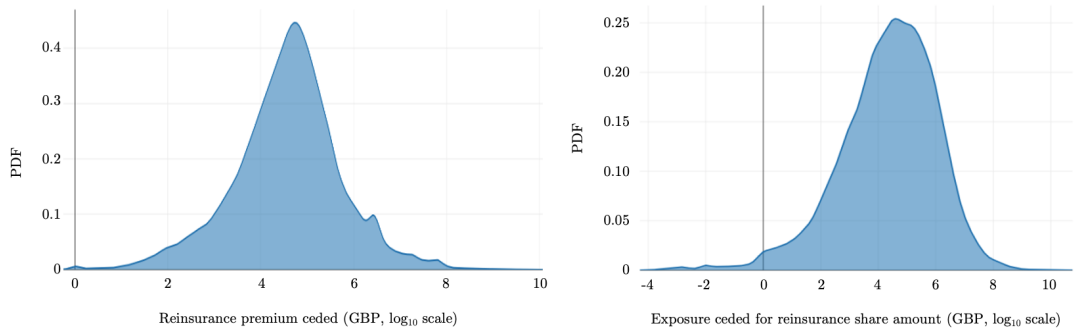
In total we include 5,349 unique contracts with a negative recoverable value, of which 159 are life contracts and 5,076 are non-life contracts.[12] We include further details on the number of contracts and performed data adjustments in Appendix A.2.

Figures 3.2a and 3.2b display the respective density plots of the aggregate counterparty exposures in the treaty and facultative data set,[13] and the recoverables data set. In both data sets there is a small proportion of very high value counterparty exposures, highlighting the skew to the data seen previously in the descriptive statistics. Figure 3.3 provides further detail for the treaty and facultative data: the complementary cumulative distribution of counterparty exposures exhibits a linear decay on a logarithmic scale, suggesting a Pareto tail – emphasising there are few very large exposures, and many smaller ones. These results corroborate the view that the UK reinsurance market exhibits strong heterogeneity in reinsurance exposures; and are consistent with the results for other financial

[12] In 114 cases it not possible to identify if the contracts are life or non-life.

[13] The density plot excludes approximately 3,000 treaty and facultative contracts for which the sum insured is unlimited.

networks studied in literature (see, for example, Caccioli et al. (2015) and Cont et al. (2013)).



(a) Counterparty exposures in the treaty and facultative data set.

(b) Counterparty exposures in the recoverables data set.

Figure 3.2: Density plots of the aggregate counterparty exposures in the reinsurance recoverables (left) and treaty and facultative (right) data sets. Values in GBP on a $\log_{10}$ scale.

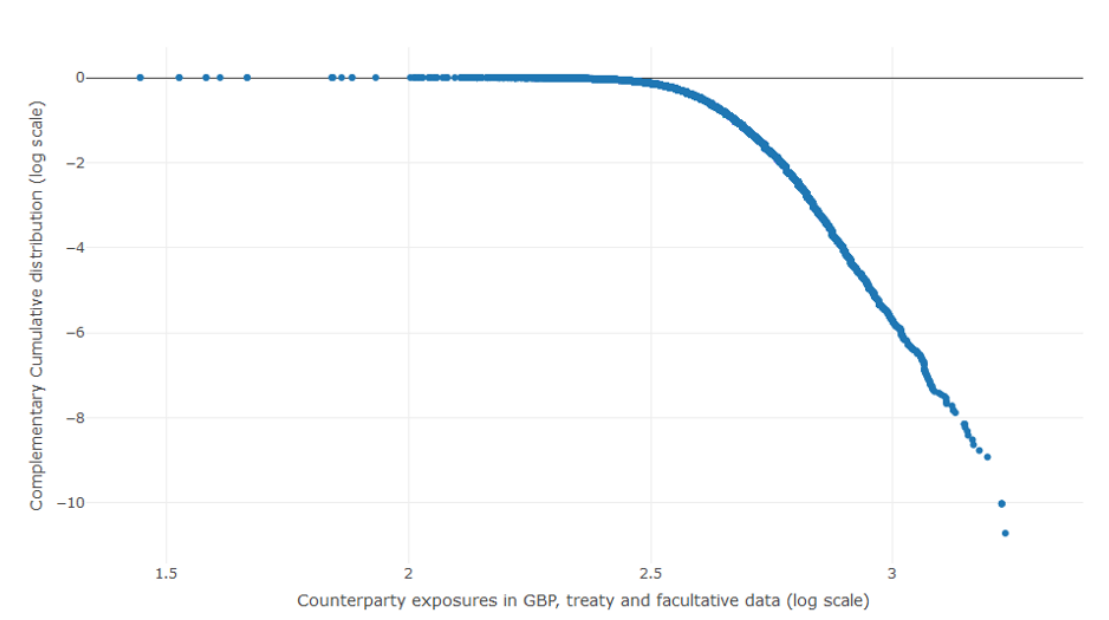


Figure 3.3: Complementary cumulative distribution plot of the aggregate counterparty exposures in the treaty and facultative data set. Values in GBP on a $\log_{10}$-$\log_{10}$ scale. Reinsurance exposures are strongly heterogeneous and exhibit a Pareto tail.

3.2.2 Identification of Insurance Market Participants

Under the Solvency II data regime, insurers are required to identify their reinsurance counterparty using both their name and a code – either the Legal Entity Identifier (LEI), or a Specific Code (SC) created by the insurer.[14] Across the treaty and facultative data and recoverables data there were approximately 14,900 unique names and code identifiers submitted. Of these, 3,100 had LEIs whilst the remaining 11,800 were submitted with SCs. Through a process of standardising insurer names – using LEIs to verify names, and names to verify, or identify, LEIs – we are able to reduce the final number of reinsurance counterparties identified to approximately 5,100. If not for this standardisation of the list of reinsurance counterparties, three variants of the same name, for instance, would be counted as three separate reinsurers.

Insurer group structure Some insurers were also allocated to an insurance group where one could be identified. For UK insurers this was done using other information on group structure reported under Solvency II. For other insurers this was carried out primarily by name, combined with the use of public information, where possible. Where an insurance group could not be identified, the solo legal entity was treated as an insurance group. In total 1,276 solo insurers with non-zero reinsurance contracts were allocated to an insurance group in the recoverables data set, and 500 were identified for the treaty and facultative data set.

Treatment of Lloyd's of London One unique aspect of the UK insurance market is Lloyd's of London. Although often spoken of as a single entity, in reality it is a marketplace made up of many investors (members), which can include insurance groups, who bear the insurance risk that is allocated through syndicates underwriting insurance (see Figure 3.4), and which fall under centralised oversight by Lloyd's. Typically, Lloyd's syndicates subscribe to underwrite risks jointly. Syndicates of underwriters are managed on behalf of the members by managing agents, which employ the underwriters, and provide other services essential to the running of syndicates.

[14]Specific Codes are permitted where an LEI is not available.

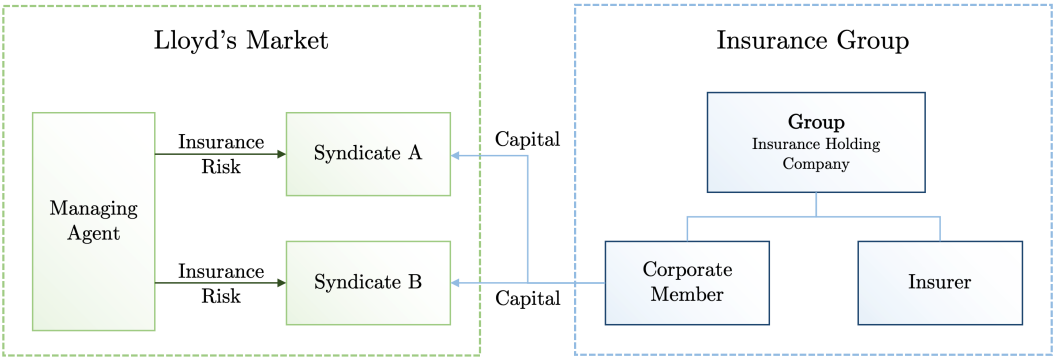


Figure 3.4: Simple representation of the Lloyd's of London market, and an example how it may be connected to an insurance company. Members of Lloyd's provide capital to underwrite policies through syndicates – formed by one or more members joining together to accept insurance risks. Corporate members include insurance groups that provide the majority of the capital for the Lloyd's market. A managing agent is a company set up to manage and oversee daily operations of one or more syndicates on behalf of the members.

If the capital of individual investors is insufficient to bear losses, then there are sources of mutual capital to protect policyholders in the form of capital called from other investors at Lloyd's as well as a Central Fund, subject to approval by the Council of Lloyd's. Risk underwritten and capital are overseen by the Corporation of Lloyd's.[15]

Our network analysis identifies individual syndicates as individual entities in the market and, using supervisory and public information, we allocate them to the insurance groups that own them. This approach enables us to capture risks within an insurance group related to both the risk transfer through their insurance companies and the business they conduct through the Lloyd's market.

3.2.3 Reinsurance Networks

We represent the interconnectedness of the reinsurance market using a weighted directed network, where market participants are represented as nodes and their interactions are defined by edges between them. We begin the discussion on rein-

[15]For further information see: `https://www.lloyds.com/about-lloyds/what-is-lloyds`.

surance networks with a formal introduction of basic network theory concepts.[16]

In the context of the reinsurance market, we define a network or a graph, denoted by $\mathcal{G} = (V, E)$, through a non-empty set of vertices (or nodes) $V = \{v_1, \ldots, v_N\}$ representing insurance market participants, and the set of edges $E = \{(v_i, v_j)\}$, with $v_i, v_j \in V$, representing bilateral exposures originating from reinsurance contracts. In particular, if (v_i, v_j) is in the set E, then we say that v_i is *adjacent* to v_j, and vertices v_i, v_j are said to be neighbours. As a result, every graph may be characterised by its *adjacency matrix*, defined as follows.

Definition 3.2.1 (Adjacency matrix)**.** *The adjacency matrix of a graph $\mathcal{G}$ is the $N \times N$ matrix $A = A(\mathcal{G})$ whose entries a_{ij} are given by*

$$a_{ij} = \begin{cases} 1, & \text{if } v_i \text{ is adjacent to } v_j; \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

To represent the direction of risk transfer in the insurance network, we represent bilateral exposures in the form of a directed graph. The graph is said to be *directed* if the set of edges E is defined by ordered pairs of nodes. By convention, the adjacency matrix of a directed network has element $a_{ij} = 1$ if there is an edge from j to i (and zero otherwise). In that case, the adjacency matrix does not need to be symmetric. Consequently, in a directed network each edge (v_i, v_j) has associated a source node (the initial vertex), $v_i \in V$, and a target node (the terminal vertex), $v_j \in V$, specifying the nature of the interaction. By convention, for risks ceded under the treaty and facultative contracts, we draw an edge from the insurer to a reinsurer to highlight the insurance risk transfer from the primary insurer to its counterparty. On the other hand, when discussing the bilateral exposures due to counterparty default risk from existing reinsurance recoverables, we consider an edge originating from an reinsurer to the primary insurer that is owed money in the form of reinsurance recoverables.

When studying the reinsurance network, we are also interested in quantifying the size of counterparty exposures. We achieve this by considering a weighted version of the adjacency matrix A in (3.3), denoted by $W = W(\mathcal{G})$, where we

[16] See, for example, Diestel (2017); Newman (2010) and Jackson (2008).

give the elements of the adjacency matrix values equal to the weights of the corresponding connections. In particular, for the recoverables network $\mathcal{G}^r$, we replace each non-zero entry a_{ij} of the adjacency matrix A with the respective size of the aggregate counterparty default exposure from node v_j to node v_i, given by w_{ij} in Equation (3.1). Similarly, for the treaty and facultative network $\mathcal{G}^s$, non-zero entries a_{ij} are replaced with the aggregate value of risk ceded from v_i to node v_j, given by w^*_{ij} in Equation (3.2). We remark that by construction matrices $W(\mathcal{G}^r)$ and $W(\mathcal{G}^s)$ are non-negative.

Basic concepts in network theory Often an indirect connection between two nodes in the network is of interest. The concept of a (directed) path from node v_i to node v_j relates to an ordered sequence of nodes, starting from v_i and finishing at v_j, such that there exist a (directed) edge between each consecutive pair of nodes. Formally, a path is a non-empty graph $P = (V^*, E^*)$ of the form

$$V^* = \{v_0, v_1, \ldots, v_k\}, \quad E^* = \{(v_0, v_1), (v_1, v_2), \ldots, (v_{k-1}, v_k)\}, \tag{3.4}$$

where the v_i are all distinct. For simplicity, we may refer to a path by the natural sequence of its nodes, where in the above example we would write $P = v_0 v_1 \ldots v_k$. The vertices v_0 and v_k are said to be *linked* by P, and the number of edges of a path is its *length*, k. Two nodes may have more than one distinct path connecting them, and the shortest path is said to be the *geodesic* path. The length of the shortest path defines the distance between the two nodes. The *diameter* of a graph can then be defined as the length of the longest geodesic path between any pair of nodes in the network for which a path actually exists. Studying these properties proves useful in identifying how risk is transferred through the reinsurance market and how losses could spread (or be contained).

A related idea to geodesic paths in a network is the one of betweenness centrality. In network analysis, centrality measures aim to characterise important nodes in the network.[17] In particular, the betweenness centrality of a node v_l consid-

[17]As the notion of node importance is often ambiguous in a general context, there are many definitions of centrality and the associated measures. A notable example includes the DebtRank measure of systemic impact of a financial intuition, introduced in Battiston et al. (2012b). In our analysis, however we do not attempt to assess the systemic importance of insurers based on some of these measures. Instead, we use a simulation-based approach to assess the systemic vulnerabilities in different stress scenarios, including a default of the most prominent insurers.

ers the number of shortest paths linking any two nodes in the network that pass through the given node v_l. Formally, let b_{ij}^l be 1 if node v_l is in the geodesic path from v_i to v_j, and 0 otherwise (including when such path does not exist). Then, define the betweenness centrality B_l of vertex v_l as

$$B_l = \sum_{i,j} b_{ij}^l. \tag{3.5}$$

Retrocession spirals, such as the London Market Excess of Loss spiral observed in the late 1980s, reveal how global interconnectedness can make the reinsurance market vulnerable to contagion. We can formalise the notion of a spiral, in which insurers cede risk only to accept the same risk through another contract, with a concept of a network *cycle*. Consider a path $P = v_0 v_1 \dots v_{k-1}$ such that $k \geq 3$. Then, we define a cycle by the graph $C := P + v_{k-1}v_0$, and its length k is given by the number of edges (or vertices).

The number of paths of a given length k on a network can be easily computed using the adjacency matrix. Noticing that the product $a_{il}a_{lj}$ is 1 if a path of length 2 from v_j to v_i via v_l exists (and zero otherwise); the total number of paths of length 2 from v_j to v_i is given by $N_{ij}^{(2)} = \sum_{l=1}^{N} a_{il}a_{lj} = [A^2]_{ij}$. This concept can be generalised to an arbitrary path length k, $N_{ij}^{(k)} = [A^k]_{ij}$. Similarly, the total number of cycles of length k is given by $L^{(k)} = \sum_{i=1}^{N} [A^k]_{ii} = \mathrm{Tr}(A^k)$.

If there exists a path between each node in the graph, the said network is *connected*. Otherwise, it is a disconnected network, which can be partitioned into distinct connected components. A *connected component* of an undirected network is defined by a subgraph in which any two nodes are connected to each other by paths, and also is such that no further nodes can be added to the subgraph preserving the above property. In a directed graph, components can be classified into two types: a strongly connected component in which every node needs to be connected with other nodes by a directed path, and a weakly connected component in which existence of an undirected path between every node in the subgraph is sufficient. Component analysis provides an alternative measure of interconnectedness of the network; and in the case of the reinsurance network, strongly connected component may be used to identify the active risk sharing community subjected to a potential contagion risk (Chen et al., 2020).

An important property of a node is its degree. We define the *degree* (or valency) $d(v_i)$ of a vertex v_i to be the number $|E(v_i)|$ of edges at v_i. In a directed graph, we can extend the notion of a degree to an in-degree of a vertex v_i, denoted by $d^{in}(v_i)$, to be the number of edges with v_i as the terminating vertex. Similarly, the out-degree of a vertex v_i, denoted by $d^{out}(v_i)$, is the number of edges with v_i as the initial vertex. Recalling that the adjacency matrix of a directed network has element $a_{ij} = 1$ if there is an edge from j to i, in- and out-degrees can be written as

$$d^{in}(v_i) = \sum_{j=1}^{N} a_{ij}, \qquad d^{out}(v_j) = \sum_{i=1}^{N} a_{ij}. \tag{3.6}$$

By definition, we have that $d(v_i) = d^{in}(v_i) + d^{out}(v_i)$. In a reinsurance network, in-degree is the number of incoming links to the reinsurer, and thus corresponds to the total number of counterparties that are exposed to its default risk. Similarly, the out-degree computes the number of outgoing links from the insurer, which represents the total number of reinsurance contracts held by the insurer. The degree is the sum of an insurer's in- and out-degree components, and measures the connectivity of the insurer. Analogously, the *strength* of an insurer in the network incorporates the weight of each link when computing this measure. That is, we can write the respective in- and out-strength as

$$s^{in}(v_i) = \sum_{j=1}^{N} w_{ij}, \qquad s^{out}(v_j) = \sum_{i=1}^{N} w_{ij}, \tag{3.7}$$

where W is the weighted adjacency matrix with entries w_{ij}. Similarly, we define the strength of node v_i by $s(v_i) = s^{in}(v_i) + s^{out}(v_i)$.

Naturally, the number m of links present in the network can be related to the sum of the degree of all vertices, that is $m = \frac{1}{2}\sum_{i=1}^{N} d(v_i) = \frac{1}{2}\sum_{i,j} a_{ij}$. Consequently, we may be interested in expressing the global level of network connectivity through its density measure. In particular, we define the connectance or *density* ρ of a graph as a ratio of links that are present and the maximum possible number of edges, that is

$$\rho = \frac{2m}{N(N-1)}. \tag{3.8}$$

When considering the maximum possible number of edges we do not allow for multi-edges (that is, at most one edge between any pair of vertices) or self-edges

(that is, no edges between a vertex and itself) in the network; in which case, we can have at most $\binom{N}{2} = \frac{1}{2}N(N-1)$ links. A network with no links has density 0, and a complete network has network density 1.

Another measure often of interest in financial networks is the *assortativity* of a network, which relates to the Pearson correlation coefficient of degree between pairs of linked nodes. Positive values indicate a relationship between nodes of similar degree (that is, highly connected nodes are connected to other highly connected nodes and vice versa), while negative values indicate relationships between nodes of different degree. Financial networks are often found to be disassortative (Caccioli et al., 2015; Cont et al., 2013), where market participants exhibit specific preferences when selecting their business counterparty.

Similarly, the nature of interconnectedness may be studied though the notion of a clique. Formally, in an undirected version of a graph, a *clique* is defined as a maximal subset of the vertices in which every member of the set is connected by an edge to every other. Here, a maximal subset means that no other node can be added to the subset while preserving the above property. A related concept is that of a local clustering coefficient, which measures the probability of neighbours of node v_i being themselves neighbours. In particular, we define a clustering coefficient c_i for a vertex v_i as the ratio of the number of pairs of neighbours of v_i that are connected and the number of pairs of neighbours of v_i (Newman, 2010). Following Watts and Strogatz (1998), we can define a clustering coefficient c for an entire network as the mean of the local clustering coefficients for each vertex, that is $c = \frac{1}{N}\sum_{i=1}^{N} c_i$.

The definition of a clique, requiring that every possible edge between its nodes is present, is a very stringent one, and often found to be impractical when studying real-world financial networks. Instead, it may be useful to study a *community* structure in these networks. In particular, we may be interested in dividing the graph into groups, clusters, or communities according to the pattern of edges. Commonly, the practice of community detection attempts to partition the vertices in such a way that there are many edges inside each group and only a few edges between groups. As the number of groups is not fixed beforehand, community detection includes a wide variety of different algorithms that can provide general

insight on the intricate nature of interconnectedness within a network, which is not easily captured through the raw network topology.

Network layers: distinguishing reinsurance contracts For some types of insurance risk, the pattern of counterparty exposures may differ depending on the type of risk or its line of business. To explore these differences, we incorporate a multi-layered network approach in certain parts of the analysis, where we assign network layers to describe different types of interactions between the insurer and reinsurer. For example, at the level of lines of business, while the set of nodes for each layer remains unchanged, a given layer is defined to contain only a subset of links for a particular business line. Other works have found taking a multi-layer approach useful to illustrate certain characteristics observed in real world systems; for example, see Kivelä et al. (2014).

This approach allows us to assess the systemic risk beyond counterparty exposures; for example, by looking at the insurance risk transfer and conducting an analysis of potential retrocession spirals. Similarly, this approach enables us to perform an analysis of sub-networks that only consist of certain type of nodes, such as life or non-life insurance networks. In principle, a multi-layered framework allows for an added level of granularity, and enables testing of the resilience of the system to particular types of shock scenarios and risks.

3.3 The Network Structure of the UK Insurance Market

Characterising the structure of the network using topological measures provides insights about the interactions among the market participants, and important economic information about contagion risk and the stability of the system (Gai and Kapadia, 2010; Roukny et al., 2013; Caccioli et al., 2012; Amini et al., 2012).

Our empirical findings on the structure of the UK insurance network are in general quantitatively and qualitatively consistent for the networks that arise from both the treaty and facultative data set (see Table 3.3), and the recoverables data set (see Table 3.4). Therefore, we often do not make an explicit distinction between

the data source but instead discuss topological properties of 'the reinsurance network' more generally. Furthermore, when discussing the network characterisation of the UK insurance market, we implicitly refer to the single (monoplex) network that captures all aggregate risk exposures. We also note that in general, network characteristics and topology remain consistent across different layers.[18]

Network	**All**	**Life**	**Non-life**	**Group**
Density	1.08%	1.17%	1.23%	2.40%
Diameter	9	3	8	9
Average path length	2.752	1.501	2.690	2.552
Average clustering	0.302	0.030	0.306	0.482
Assortativity	-0.093	-0.216	-0.111	-0.208
Average betweenness	0.019	0.003	0.021	0.024
Average degree	19.18	3.77	20.21	18.42
Degree: power law exponent estimate	1.680	2.294	5.559	1.699
In-degree: power law exponent estimate	1.711	2.964	1.697	8.027
Out-degree: power law exponent estimate	7.780	2.063	7.781	5.098
Link fraction (weighted) top 5% connected nodes	60.6% (81.4%)	49.8% (67.5%)	59.5% (66.0%)	59.5% (84.5%)

Table 3.3: Summary of network statistics based on the treaty and facultative dataset. Results shown for the full network including all reinsurance contracts, as well as network layers considering life contracts, non-life contracts, and contracts on a group-level only.

3.3.1 Core-Periphery Structure

We find core-periphery relationships in the reinsurance data sets considered, with a small core of densely connected reinsurers, and other reinsurers in the periphery dispersing risk (see Figure 3.5 for a visual representation of our networks). Some UK insurers have a significant number of reinsurance contracts with reinsurers that no other UK insurer is using, which gives rise to a strong hierarchical relationship

[18]We omit presentation of more detailed results so as to not disclose information that could lead to identification of particular firms in the case of smaller network layers.

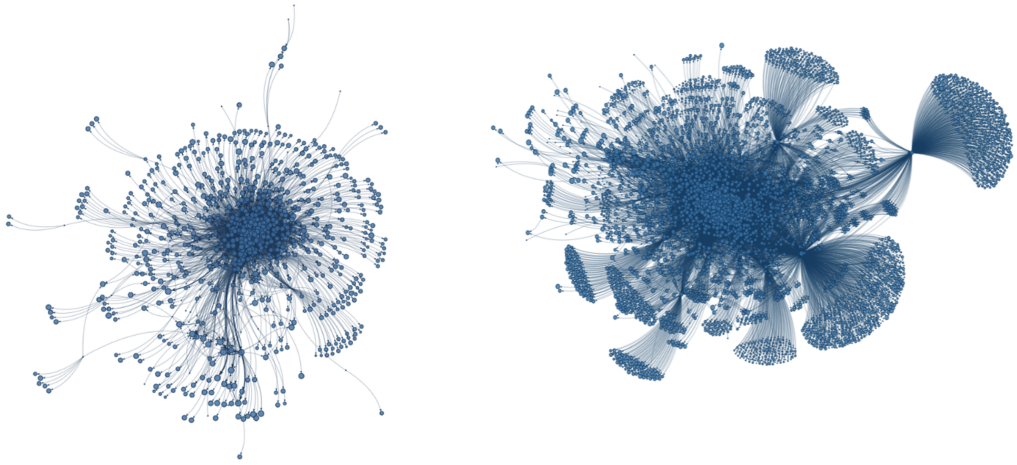
Network	**All**	**Life**	**Non-life**	**Group**
Density	0.12%	0.45%	0.14%	0.13%
Diameter	9	9	8	6
Average path length	3.271	3.684	3.304	2.842
Average clustering	0.141	0.051	0.163	0.110
Assortativity	-0.346	-0.291	-0.381	-0.399
Average betweenness	0.110	0.018	0.109	0.117
Average degree	10.38	3.85	10.39	7.89
Degree: power law exponent estimate	1.798	3.630	11.761	1.914
In-degree: power law exponent estimate	1.673	2.086	2.048	2.248
Out-degree: power law exponent estimate	1.956	3.340	1.910	2.077
Link fraction (weighted) top 5% connected nodes	98.0% (96.0%)	68.0% (90.0%)	97.8% (90.0%)	100% (99.0%)

Table 3.4: Summary of network statistics based on the recoverables dataset. Results shown for the full network including all reinsurance contracts, as well as network layers considering life contracts, non-life contracts, and contracts on a group-level only.

between market participants. These UK insurers act to extend the network by interacting with nodes outside the UK insurance market and disperse risk further.[19] This is in contrast to other financial networks where the periphery tends to be connected to a single common set of hubs (Fricke and Lux, 2015; Barucca and Lillo, 2016).

The degree distribution sheds more light on the heterogeneous structure of exposures and the network's resilience to financial contagion. Indeed, visual inspection of the network plots in Figure 3.5 reveals that some reinsurers have very many connections, while others have very few, suggesting the presence of hierarchical relationships. This observation is further supported by the double logarithmic plots of the empirical complementary cumulative distributions of degree and strength

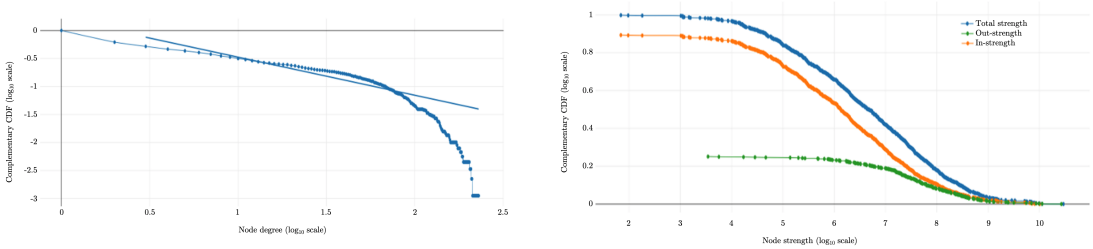
[19]The nature of the scope of the data set, by not including the reinsurance contracts of these reinsurers, and therefore not knowing if risk is then ceded back into the UK insurance sector means we cannot draw too strong a conclusion here. However, at least for the purposes of understanding how UK reinsurance risk is ceded, these insurers do appear to extend the network of UK insurance risks and increase the dispersion of risk.



(a) Treaty and facultative data set. (b) Reinsurance recoverables data set.

Figure 3.5: Visualisation of UK reinsurance networks, year-end 2016.

for the treaty and facultative data set, shown in Figure 3.6. We observe similar results for the recoverables network.



(a) Total degree distribution with a power law fit on a $\log_{10}$-$\log_{10}$ scale. (b) Node strength distribution plot on a $\log_{10}$-$\log_{10}$ scale.

Figure 3.6: Degree (left) and strength (right) distribution for the network of facultative and treaty reinsurance contracts.

Moreover, the reinsurance network is characterised by heavy tailed degree distributions and negative degree correlations. We observe linear decay in the tails of the in-degree, out-degree and total degree distributions suggesting a heavy Pareto tail. In the treaty and facultative network for example the maximum likelihood estimates of the exponent parameter are 1.7, 1.7 and 7.8 for the distribution of

total degree, in-degree and out-degree respectively.[20] We recall that a network whose degree distribution $P(k)$ follows a power law, that is $P(k) \sim k^{-\gamma}$, is known as a scale-free network. A smaller estimate of exponent γ relates to a heavier tail in the distribution of connectivity. Our results are indicative of a structure that includes a relatively small number of hubs – that is, nodes that are very highly connected in the network – that form the main core of the network.

To assess the goodness-of-fit of our power law estimates, we use a one-sample Kolmogorov-Smirnov test at a 1% significance level. We do not find enough evidence to reject the null hypothesis that the original data could have been drawn from the fitted power-law distribution,[21] and hence conclude the UK insurance market forms a scale-free network. Our results are consistent with the previous study of Chen et al. (2020) on the US property-casualty insurance, as well as with degree distributions observed in other financial networks (Boss et al., 2004; Degryse and Nguyen, 2004; Iazzetta and Manna, 2009; Cont et al., 2013; Caccioli et al., 2015).

We also observe that highly connected insurers tend to have larger exposures, as evidenced by the scatter plot in Figure 3.7. The statistical dependence is supported formally by the Kendall tau test at a 1% significance level. For the treaty and facultative data we have a tau statistic (and the corresponding p-value) of 0.57 (< 0.01) for total degree against strength, 0.61 (< 0.01) for in-degree against in-strength, and 0.94 (< 0.01) for out-degree against out-strength.

Our findings of a heterogeneous degree distribution and negative degree correlations reinforce our view that the network has a core-periphery structure. Highly connected reinsurers at the core of the graph play the role of central hubs that mediate risk transfer between lesser connected insurers in the market. Our empirical analysis finds a small core (less than 5% of all insurers in the network), consisting of hub reinsurers, interacting with a large part of the network and mediating a

[20] We implement the method of Clauset et al. (2009) to calculate the parameters of the fitted distribution. A summary of maximum likelihood estimates in the case of the recoverables data set and other network layers across both data sets is presented in Tables 3.3 and 3.4. In general, we find estimates in the usual range of 2–3.

[21] For the treaty and facultative data for example the reported p-values for the test statistic are 0.32, 0.39 and 0.96 for total, in- and out-degree, respectively.

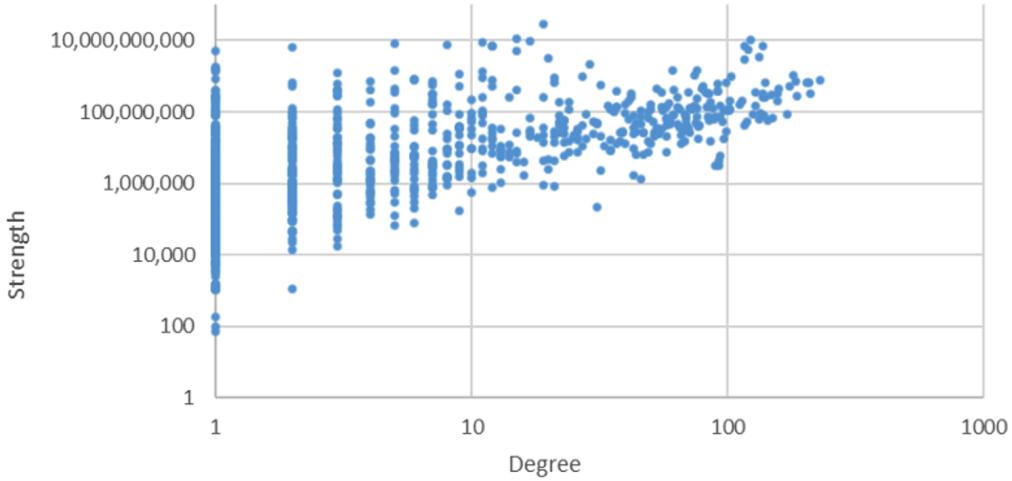


Figure 3.7: A positive relation between node degree and strength distribution for the facultative and treaty network.

significant portion of risk transfer activity in the market.[22] For the facultative and treaty data set, the top 5% of connected insurers hold 60% of links in the network, amounting to 81% of the total value of reinsurance in the market; while for the recoverables data set, the hub insurers hold 98% of links, representing 96% of the total reinsurance amount.

The distribution of connectivity has important implications on systemic vulnerability (Allen and Gale, 2000; Battiston et al., 2012b). In particular, Caccioli et al. (2012) show that conditional on observing a contagion event the size of the loss cascade increases with the average degree: higher connectivity implies losses can be transmitted faster to a larger number of market participants. On the other hand, the probability of observing a contagion event is a concave function of the average degree: while highly connected networks allow for higher diversification of exposures and thus increase the robustness of the system to reinsurer default, poorly connected networks do not have a sufficient number of links to trigger a cascade. Consequently, networks with a medium level of connectivity tend to have the highest probability of observing a default cascade, as there are enough links for the losses to spread but insurers do not benefit from much diversification of

[22]Due to data confidentiality, we omit detailed presentation of results on the core reinsurers in the network.

exposures. Our stress-simulation analysis in Section 3.4 provides further insights into the network's resilience.

3.3.2 A Small-World Graph

Links in the UK insurance market are sparse, and with network density of only around 1%, the vast majority of insurers are not connected to one another (see Tables 3.3 and 3.4). The majority of insurers in the periphery of the network have a very small number of connections, and only a small portion of reinsurers in the core are connected to a large number of market participants. Interestingly, with the exception of a few nodes, all insurers belong to the same connected component. In other words, almost all market participants belong to one risk sharing network.

Despite the fact that most insurers belong to a single risk sharing network, we do observe clustering of insurers forming around reinsurers in the core and find high local clustering as measured by the clustering coefficient. In other words, there is a high probability that two insurers with a reinsurance contract linking them also have another common counterparty. This relates to the counterparty risk externality, as described by Acharya and Bisin (2014). That is, the action of a reinsurer underwriting contracts with other insurers increases its default probability, and hence leads to an increased default risk faced by its counterparties. The externality arises in an opaque market, as insurers do not have specific information on the risk taken by its reinsurer.

As a result of the high local clustering, we observe local community structures in the network. These are small sub-networks characterised by a relatively high density of links. In those communities, the likelihood of a connection between its member insurers significantly exceeds the average probability that any two insurers in the network are connected.

Hub reinsurers serve as the common connections for communities, which results in a short path between any two insurers in the system. Consequently, the network diameter, defined as the shortest path length between the two most distant insurers, tends to be low for scale-free networks with a core-periphery structure.

The presence of these hubs, commonly observed in financial networks, distinguish scale-free networks from random networks (Barabási and Pósfai, 2016;

Barabási and Albert, 1999). As suggested by Barabási and Albert (1999), real networks often exhibit a preferential connectivity: the emergence of scale-free topology can be described by the preferential attachment process in which a new node in the system has a higher probability of forming a link with already well connected nodes than with nodes that have a small number of connections. Consequently, insurers are more likely to transfer part of their risk to a large reinsurer in the core of the network than a reinsurer found in the periphery.

We observe that insurers with a smaller number of connections tend to have counterparties with densely connected neighbours (that is, they form a community), while insurers that act as hubs connect insurers that tend not to link with each other. This result is corroborated by Figure 3.8 which displays the relationship between the local clustering coefficient and total degree for the treaty and facultative network. Analogous results were found for the recoverables data set.

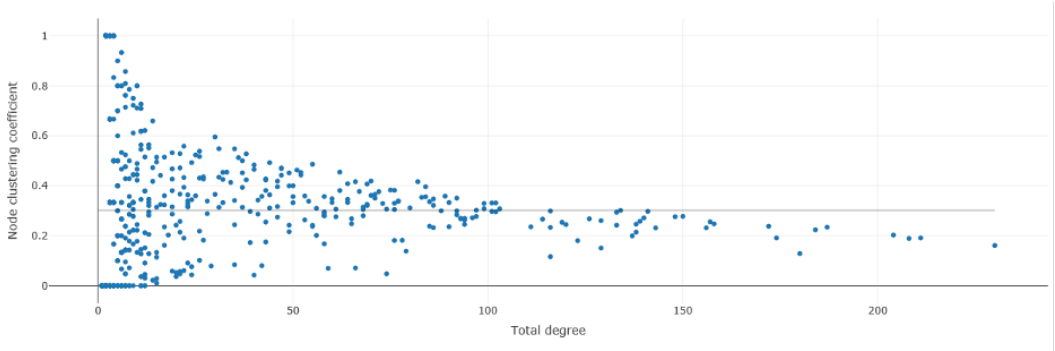


Figure 3.8: The relationship between the local clustering coefficient and node's degree for the treaty and facultative network. The clustering coefficient remains bounded away from zero, which is characteristic of small-world graphs.

This hierarchical relationship between insurers and reinsurers is captured by the assortativity coefficient: a negative value shows the tendency of highly connected insurers (hubs) to enter in a reinsurance contract with insurers that have a small degree (see Tables 3.3 and 3.4). The negative relationship between an insurer's degree and the clustering coefficient is a feature of the core-periphery structure.

Although most of the insurers in the network are not directly connected with one another, they can be indirectly connected by a small number of links through a small set of hubs. Despite the scarcity of links, the average path length between

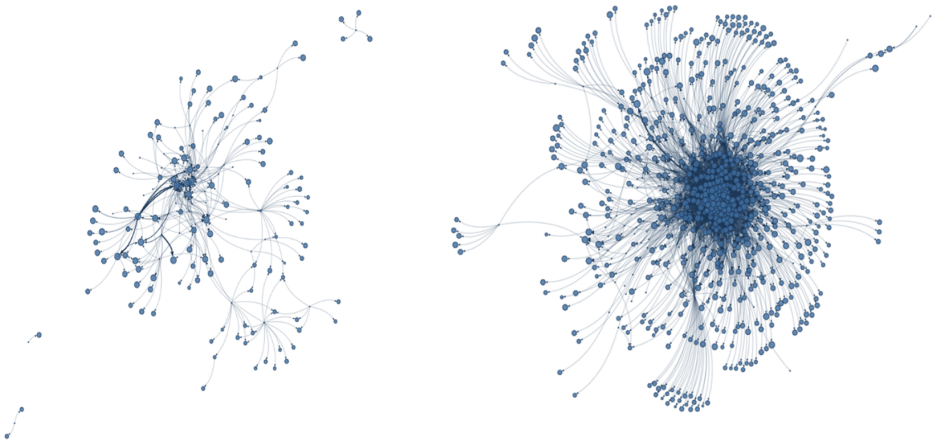
insurers remains small: a pair of insurers is linked by a path with length orders of magnitude lower than the network size. Shorter paths create the conditions for losses to spread more quickly.

Link sparsity, high local clustering and short average path lengths often distinguish real networks from simple theoretical models (Cont and Tanimura, 2008): an empirical observation that inspired the development of a new class of random networks, termed *small-world* graphs (Watts and Strogatz, 1998; Watts, 1999; Newman, 2003).

Our analysis supports the view that the UK reinsurance market forms a small-world graph. A notable feature of this class of graph is that small-world networks are more robust to random shocks in the system than other network architectures: since hubs mediate most connections, the default of a small reinsurer should have minimal effect on systemic stability. On the other hand, default of a large central reinsurer can have a significant adverse effect on the system. Consequently, we expect the commonly described 'robust-yet-fragile' property of financial networks to hold for the UK insurance market as well (Gai and Kapadia, 2010; Gai, 2013; Caccioli et al., 2012). This reinforces our earlier findings on the core-periphery nature of connectivity, and emphasises the importance of stress simulation, conducted in Section 3.4, for assessing the level and source of vulnerability in the UK reinsurance sector.

3.3.3 UK Insurance Market: a Heterogeneous System

Our network statistics (see Tables 3.3 and 3.4 for details) highlight that while core-periphery relationships exist in the reinsurance network there are nuances in how life insurers and non-life insurers share risks. The life reinsurance network is more hierarchical than the non-life network. In particular, it is characterised by a much lower diameter, a lower average path length and a lower betweenness centrality than the non-life network. This suggests greater connectivity, mediated by hubs, between life (re)insurers than is the case for general (re)insurers. Life insurers also tend to cede reinsurance to fewer counterparties than non-life insurers, as can be seen in Figure 3.9.



(a) Life reinsurance network. (b) Non-life reinsurance network.

Figure 3.9: Visualisation of sub-networks with life (left) and non-life (right) contracts using the treaty and facultative data set.

Figure 3.9 also reveals that the network of non-life reinsurance is denser and more connected than its life reinsurance counterpart, suggesting higher levels of insurance risk diversification. There are relatively few reinsurers of non-life insurance that have many connections, as corroborated by estimates of the degree distribution exponents in Tables 3.3 and 3.4, and they are proportionately fewer than comparable reinsurers of life insurance. However, the largest non-life reinsurers account for a higher proportion of contracts, by value, than the largest life insurers.

To examine further sources of heterogeneity within the overall network, we apply community analysis using clustering algorithms to identify cohesive groups known as communities. As explained earlier, these communities consist of densely connected insurers within the wider network, and the number of links between insurers in different communities remains relatively sparse (see Figure 3.10).

We find that clusters of densely connected reinsurers tend not to be organised around individual lines of business. Although we found some particular examples of clusters that were more specialised towards life insurance (see Figure 3.11 for the life and non-life network), most of the communities contained insurers with reinsurance contracts under multiple lines of business. This could simply be a reflection of diversification, particularly for non-life insurers, where insurers

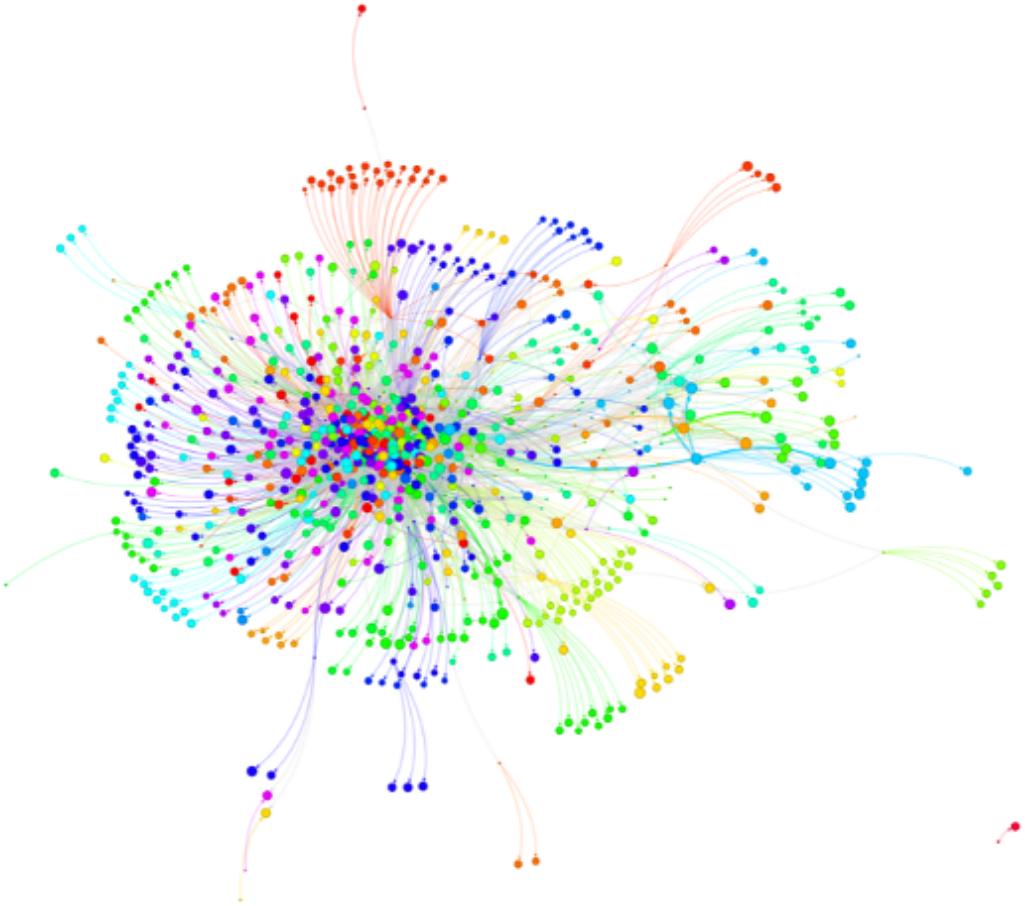


Figure 3.10: Community structure in the treaty and facultative network, detected using a Spinglass algorithm.

underwrite multiple lines of business and have multiple reinsurance contracts in place.

Retrocession spirals The London Market Excess of Loss spiral that affected the Lloyd's of London market participants in the late 1980s is a prime example of how global interconnectedness in the reinsurance market can cause contagion to spread (Bain, 1999). Despite the belief at the time that all parties are properly insured, the intricate structure of retrocession contracts resulted in an unusually large concentration of losses. In particular, retrocession spirals lead to a counter-intuitive non-linear behaviour of losses in the system, where a disproportionate

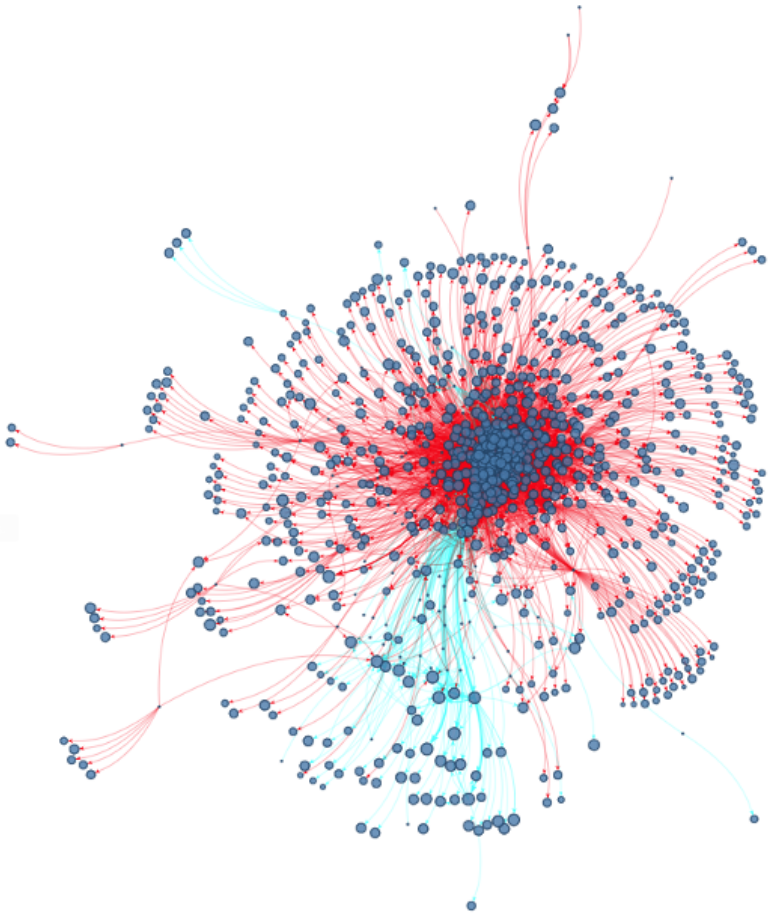


Figure 3.11: Treaty and facultative network visualisation with life edges in blue and non-life edges in red.

amount of excess liability can be left with a single reinsurer (Klages-Mundt and Minca, 2020). In the presence of these network effects that are often invisible to the market participants, insurers face a high level of uncertainty about their risk, and thus may suffer from misspecification of internal risk models. In particular, as recently shown by Klages-Mundt and Minca (2020), the presence of retrocession spirals can have a detrimental effect on financial stability, making the reinsurance system vulnerable to systemic risk.

Network analysis can therefore help us to identify the dangerous retrocession cycles that can result in a severe underestimation of risk by the insurers. In

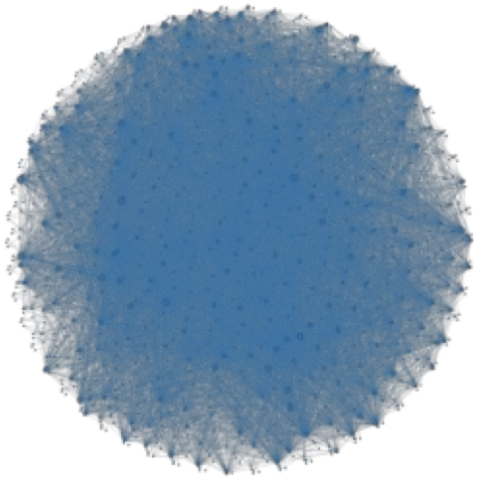
particular, we use the treaty and facultative data to examine the prevalence of network cycles, and hence identify contracts that can lead to such retrocession spirals: that is, identify where risks are potentially retroceded back to the original insurer. Moreover, to focus on the nature of the risk, we layer the network by the different lines of business and identify network cycles of different sizes at each layer. Figure 3.12 shows a network plot for the entire UK insurance market as well as particular layers for the three most common lines of business.

Table 3.5 reports the number of identified network cycles. Note that due to computational complexity, only cycles of up to size five are considered.

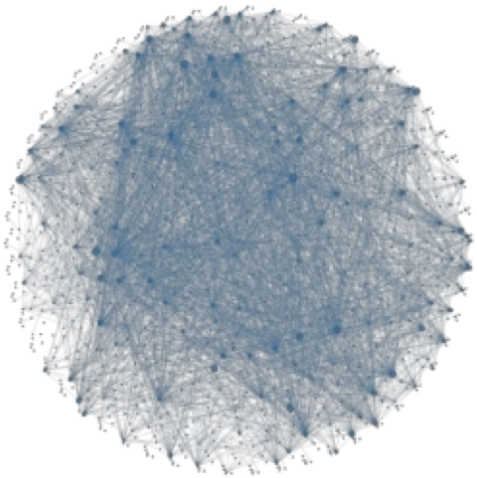
Line of Business	*Cycle Size*			
	2	*3*	*4*	*5*
Full network	335	5488	98831	1849710
Fire and other damage to property	104	888	8747	87286
Marine, aviation, transport	107	929	8915	87454
General liability	28	95	432	1768

Table 3.5: Summary of potential retrocession spirals (network cycles) for major lines of insurance business.

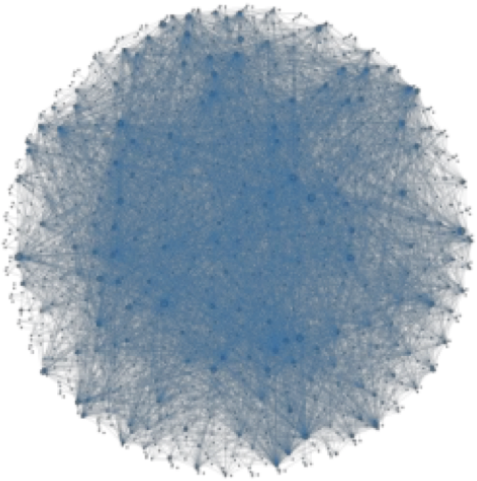
The data, however, imposes limitations on the network cycle detection analysis. Firstly, as we do not have world-wide data on reinsurance contracts, there could be important cycles left undetected due to lack of data for non-UK insurers and their ceded risk. Furthermore, although our results shed light on the potential for retrocession cycles in the UK insurance market, we do not have fine enough detail about the nature of the risk to infer that a loss event could be amplified because of the existence of a network cycle. The lines of business are broad descriptions of risk, and, for instance, damage to property could include residential housing in London and commercial property in Lagos. Consequently, more analysis is needed to detect whether the same risks exist, and given the available data granularity this has to be done on an individual basis. Our analysis only highlights the potential for retrocession cycles to exist, but we cannot be conclusive.



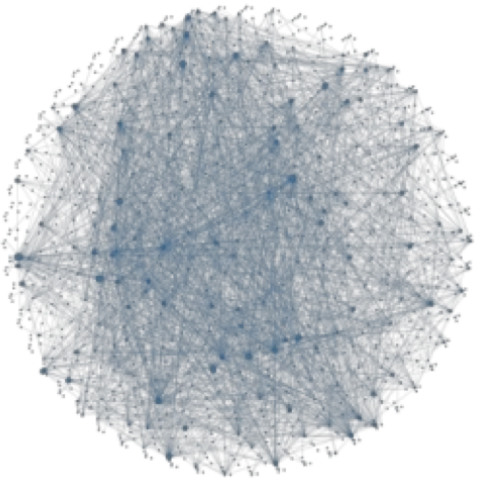
(a) Full network.



(b) Marine, aviation, transport.



(c) Fire and other damage to property.



(d) General liability.

Figure 3.12: Visualisation of major lines of insurance business and potential retrocession spirals (network cycles). Node size proportional to PageRank centrality, and edge width to contract size. PageRank centrality here incorporates both the direction of links and their weight, and was originally developed for website ranking in search engines. The principle of the algorithm is that the vulnerability of a node increases with the number of connections to other vulnerable counterparties.

3.4 Simulation of Stress Scenarios

The analysis of UK reinsurance market structure reveals potential sources of systemic vulnerability. In this section we use a simulation-based approach to assess the risk of counterparty credit contagion. In particular, to test the stability of the reinsurance recoverables network, we examine the effect of a shock to insurers' and reinsurers' financial investments, excluding unit-linked investments,[23] as well as the default of individual large, highly interconnected reinsurer groups within the UK network.[24]

Balance sheet data In our simulation-based approach, we consider the network of reinsurance recoverables consisting of 4,378 individual insurers. To assess the impact of bilateral counterparty exposures on systemic risk, we consider a simple representation of a balance sheet of each market participant.

For UK insurers, we build their balance sheets using the readily available data from Solvency II regulatory returns. For the remaining non-UK insurers, the balance sheet information is obtained manually from SNL, Capital IQ, and US Statutory Insurance Financial information (all from Standard & Poor's). We then construct an approximately equivalent balance sheet to that of Solvency II, for example by removing intangible assets such as goodwill and deferred acquisition costs. Due to time limitations, balance sheet information could be obtained only for the largest non-UK reinsurers, as characterised by the size of their exposures in the UK insurance market.

As a consequence of this data limitation, we possess detailed balance sheet information only for approximately 500 insurers in the network. From the perspective of stress testing, it is however crucial that other insurers are retained in the network. Their removal could significantly alter the graph topology, and thus the way in which default contagion can spread through the system. Therefore, in order to capture the network effects adequately, we retain the remaining insurers

[23] Unit-linked investments are made on behalf of the policyholder, who bears the investment risk (similarly as in the case of mutual funds).

[24] Further layers could be added to the stress simulation exercise, based on retrocession, where reinsurers cede exposure to other reinsurers. That said, our investigations reveal that the data does not contain standardised, precise risk descriptions to enable this analysis, and hence we focus only on the counterparty default risk instead.

in our simulation approach by endowing them with artificial balance sheet data. In particular, we build total assets and liabilities around their observed reinsurance assets and liabilities, assuming a proportional relationship between these items. Our approach is motivated by common reinsurance market risk practices, often influenced by the existing regulatory requirements. We calibrate the proportion of reinsurance to total assets and liabilities for an average insurer using the aforementioned commercial data as well as data taken from the 2016 EIOPA Insurance Stress Test (EIOPA, 2016). Notably, we apply a capital buffer of assets over liabilities of 10% for the artificial balance sheet data, in line with the 2016 EIOPA Insurance Stress Test (EIOPA, 2016).

Stress test methodology For consistency with the EIOPA's exercise, we calibrate the stress to a 4% loss in the value of non-unit-linked investments; we also include much more severe stresses of 10% and 15% of total investments. However, it should be noted that EIOPA's stress test changes the value of assets and liabilities to result in a change in equity. Given that we do not have data on the underlying cash flows and hedging behaviour, we are not able to fully replicate the mechanics of EIOPA's stress scenario on the liability side. Instead, we observe the net impact of the stress scenario on equity and replicate it with a suitable shock on the asset side of the balance sheet.

In our framework, any recoverable amounts that could not be paid were allocated amongst creditors in proportion to their claim. In particular, we use an extension of the Eisenberg and Noe (2001) model, where the obligation of all firms within the system are determined simultaneously and consistently with the priority of debt claims and the limited liability of equity. This extension, introduced by Rogers and Veraart (2013), includes bankruptcy costs through the Greatest Clearing Vector Algorithm (GA) that allows insurers to fail in succession until only solvent firms remain. These bankruptcy costs include, for instance, the costs of insolvency specialists or inefficiencies in the ability to realise asset values. In line with the methodology of EIOPA (2016), we consider in our stress simulation scenarios bankruptcy costs of 10% and 20%. In contrast to other financial sectors, these relatively small bankruptcy costs are motivated by the fact that insurance claims are paid over long time periods (typically years). This allows for a more

orderly recovery of claims by counterparties of the defaulted reinsurer. Moreover, we find our results on loss contagion in general to be robust to varying bankruptcy costs.

Summary of results Applying an adverse scenario in line with the 2016 EIOPA insurance stress test, defined by a loss to total investments of 4% and bankruptcy cost of 10%, we find that whilst 25% of the capital in the system was lost, out of 4,378 firms we observe only 14 insurers initially defaulting due to the adverse shock and then a single default due to network effects. Turning to a much more severe stress scenario with a 15% loss to total investments and bankruptcy cost of 20%, we observe 136 defaults in total, where only 18 (13%) of these stem from the contagion effects. This relatively small number of defaults is despite a severe loss of 93% of capital in the system.

Table 3.6 presents the results based on a range of macroeconomic shocks to total investments (excluding assets backing unit-linked policies) and the assumed bankruptcy cost of 10%. Similar results are observed under a 20% bankruptcy cost assumption, as outlined in Table 3.7.

Shock on total investments	**4%**	**10%**	**15%**
Solvent nodes	4356	4307	4248
Shock defaults	14	62	118
Contagion defaults	1	2	5
Total losses	347 (24.7%)	864 (61.7%)	1297 (92.7%)
Shock losses	344 (99.0%)	859 (99.4%)	1290 (99.5%)
Contagion losses	3.5 (1.0%)	5.0 (0.6%)	7.1 (0.5%)

Table 3.6: Stress simulation results with 90% recovery rate assumed: shock to total investments (excluding assets backing unit-linked policies). Values given in billions of GBP. Results for the UK insurance market consisting of 4,378 nodes and a total initial capital of 1,400 billion GBP. Shock losses result directly from the initial shock of the scenario, whilst contagion losses are a result of network effects.

Our results show a limited impact of default contagion on the stability of the UK reinsurance market. In particular, risk of default remains small even in

Shock on total investments	**4%**	**10%**	**15%**
Solvent nodes	4356	4307	4235
Shock defaults	14	62	118
Contagion defaults	1	2	18
Total losses	348 (24.9%)	867 (61.9%)	1302 (93.0%)
Shock losses	344 (98.8%)	859 (99.1%)	1290 (99.0%)
Contagion losses	4.3 (1.2%)	8.1 (0.9%)	12.4 (1.0%)

Table 3.7: Stress simulation results with 80% recovery rate assumed: shock to total investments (excluding assets backing unit-linked policies). Values given in billions of GBP. Results for the UK insurance market consisting of 4,378 nodes and a total initial capital of 1,400 billion GBP. Shock losses result directly from the initial shock of the scenario, whilst contagion losses are a result of network effects.

the presence of large macroeconomic shocks that erase most of the capital held by the market participants. In contrast to other financial networks, our results suggest that the connectivity induced by reinsurance contracts does not introduce fragility to the system. Secondary losses from counterparty defaults constitute only approximately 1% of total capital lost under the considered range of stress scenarios.

Whilst these results provide a view on how the topology of the reinsurance network affects its stability, they should not be regarded as definitive. In particular, we have had to make important assumptions about entities in the network for whom we lack data. Consequently, our exercise may fail to capture certain idiosyncrasies amongst insurers, for example ones that coincide with a highly connected node in the network. Since hub reinsurers found in the core of the network tend to have a higher level of capitalisation than an average insurer, our market-wide stress scenario does not induce their immediate default. Therefore, even with a reduced capital buffer under a stress scenario, strong risk diversification allows these hub reinsurers to maintain systemic stability by absorbing losses from contagious defaults and thus shielding the firms in the network periphery.

In light of this, we are interested in quantifying the impact of the failure of a given hub node (or a set of nodes related to a single insurance group) on the

stability of the system. In our approach, we consider an idiosyncratic shock to a specific large insurance group resulting in its immediate default.[25] We consider a worst-case scenario in which the shocked reinsurers see the value of their total assets reduced to zero; that is, the immediate counterparties have an assumed 0% recovery rate on their reinsurance exposure. For the remaining contagious defaults, as previously, we assume bankruptcy costs of 20%.

Table 3.8 presents the resulting systemic impact following the default of the most prominent insurance groups. Overall, we find the UK reinsurance market to be robust to default of the largest, most connected insurers. In general, losses are contained within a small part of the network formed by the defaulted insurance group and its direct counterparties. In our simulations, we observe at most four additional contagion-induced defaults, while typically only a single default by contagion is seen. Interestingly, this single default is attributed to the same insurer in the network, which highlights its inadequate capitalisation rather than the systemic impact of its reinsurers.

Defaulted group	**I**	**II**	**III**	**IV**	**V**
Solvent nodes	4376	4377	4377	4372	4376
Shock defaults	2	1	1	3	2
Contagion defaults	1	1	1	4	1
Total losses	10.1 (0.7%)	10.4 (0.7%)	4.3 (0.3%)	21.6 (1.5%)	5.7 (0.4%)
Shock losses	5.9 (58%)	6.2 (60%)	0.1 (2%)	16.2 (75%)	1.5 (26%)
Contagion losses	4.2 (42%)	4.2 (40%)	4.2 (98%)	5.4 (25%)	4.2 (74%)

Table 3.8: Stress simulation results with 80% recovery rate assumed: shock from a default of individual entities within a single large insurance group. Values given in billions of GBP. Results for the UK insurance market consisting of 4,378 nodes and a total initial capital of 1,400 billion GBP. Shock losses result directly from the initial shock of the scenario, whilst contagion losses are a result of network effects.

Our results are in line with the previous conclusions of IAIS (2012) and Geneva Association (2010), which find reinsurance unlikely to contribute or amplify systemic risk due to the relatively small size of premiums ceded and retroceded as

[25]Due to data confidentiality, we omit presenting details on these core reinsurer groups.

compared to the primary insurance market. This view is corroborated by the results from our stress test simulations, both in the case of a macroeconomic shock affecting insurers' portfolios and an idiosyncratic shock leading to the default of prominent insurance groups. We find that the robustness of the system can be attributed to a combination of two factors: the structure and the size of counterparty default risk exposures. In particular, the observed network characteristics, such as the scale-free degree distribution and the core-periphery structure, help to attenuate and disperse disturbances within the system. At the same time, since counterparty exposures due to reinsurance contracts form a relatively small fraction of assets and liabilities on the balance sheet, even large macroeconomic shocks are insufficient to reach the tipping point beyond which interconnectedness intermediates spread of the contagion and thus aids shock amplification. This is in a stark contrast to banking networks, where interbank exposures form a significant part of their business, which renders them robust-yet-fragile to the risk of contagion (Gai and Kapadia, 2010; Acemoglu et al., 2015; Caccioli et al., 2015).

3.5 Conclusions

Our research has sought to review the evidence for the commonly held view that, in and of itself, reinsurance does not pose contagion risk. The hierarchical relationships and limited ceding of insurance risk are believed to provide natural brakes for the spread of counterparty losses experienced by reinsurers.

Using UK data on reinsurance contracts and amounts recoverable, our findings do not contradict this view. We find the reinsurance network consists of densely connected insurers at the core, playing the role of 'hubs' for the UK insurance market, while there are other sparsely connected reinsurers in the periphery. Reinsurers in the core tend not to connect with one another, that is, relationships tend to be strongly hierarchical. The UK reinsurance network exhibits topological properties similar to theoretical scale-free and small-world networks. Relatively few connections are needed to reach each insurer in the network, which gives rise to the 'robust-yet-fragile' property.

We find that life and non-life networks both exhibit the core-periphery structure. In comparison, life reinsurance contracts tend to lead to larger exposures

than non-life contracts, and the non-life reinsurance network is denser than the life network, suggesting greater risk-sharing by non-life insurers. Risk-sharing by life insurers tends to be conducted via central reinsurers to a greater extent than for non-life insurers.

Although there is a small core of hubs that are highly connected in the UK insurance market, these hubs tend to connect with each other in a lesser degree than is the case for other financial networks. Moreover, community structures do not appear to form around lines of business, which indicates a strong diversification of insurance risk.

Finally, we use stress-simulation analysis to gain insight into whether the 'robust-yet-fragile' property of the reinsurance network could lead to instability. Using the EIOPA (2016) stress test as reference, we calibrate a shock to the insurers' financial investments excluding assets held for unit-linked policyholders, and find defaults from contagion to be relatively low. Similarly, to review the risk posed by the failure of hubs we look at the effects of the individual default of highly connected reinsurers within the UK insurance market. Again, we find little contagion risk from these shocks. Our results, however, are based on severe balance sheet assumptions due to data limitations. In particular, our exercise ignores any connections between entities outside of the UK reinsurance market that are not reported under Solvency II. Moreover, the results are based on a large number of non-UK insurers for which we do not have readily available balance sheet data, and thus for which we impose artificial assumptions about their assets and liabilities.

Future work Solvency II creates a rich set of data for analysis, the potential of which we have only partly been able to explore. While we have data on reinsurance contracts held by UK insurers there are further data that would be useful in building a global perspective:

- Group-level data that report transactions by non-UK insurers, including intra-group transactions.
- Other group-level data from EU insurers that are shared within supervisory colleges.

- Data on the insurers' use of special purpose vehicles (SPVs) and insurance linked securities.

- Publicly available data on reinsurance contracts ceded by US insurers.

Besides expanding the global scope of the reinsurance network analysis we could also introduce a temporal dimension by incorporating reinsurance data sets for a wider range of reporting dates. A longer time series of data would help identify whether contractual relationships are stable over time. For example, for reinsurance recoverables in particular there could be some interesting changes between year-end 2016 and year-end 2017 given the numerous catastrophic events that occurred during 2017.

Network effects (and potential retrocession spirals) could be present due to the presence of feedback loops in the system. Further work could take place to standardise risk descriptions in Solvency II reporting to allow a finer identification of risks. A future analysis could also prioritise the identification of more important network cycles – for example, a feedback loop of losses is only possible if links in the cycle correspond to sufficiently large exposures with respect to the insurer's reserve levels. Further work could consider a sub-network that only contains large exposures, where the expected size of losses exceeds a threshold for losses as a proportion of initial capital.

We believe network analysis can be a useful addition to insurance stress tests. For example, it allows supervisors to explore second round effects from contagion following initial shocks. In addition, in order to extend the stress-simulation analysis presented in this chapter it would be useful to:

- Include more data on non-UK insurers, including more detailed and heterogeneous balance sheets.

- Review systemic risk due to an insurance specific shock, for example a major catastrophe. This would create an increase in the value of the reinsurers' liabilities, potentially giving rise to contagion for sufficiently large shocks. It would be particularly interesting to see whether non-proportional, excess of loss contracts can propagate losses as layers of cover are exhausted by losses.

Chapter 4

Modelling COVID-19 Contagion: Risk Assessment and Targeted Mitigation Policies

Chapter based on:
Rama Cont, Artur Kotlicki, and Renyuan Xu. Modelling COVID-19 contagion: risk assessment and targeted mitigation policies. *Royal Society Open Science* 8: 201535, 2021.

We use a spatial epidemic model with demographic and geographic heterogeneity to study the regional dynamics of COVID-19 across 133 regions in England.

Our model emphasises the role of variability of regional outcomes and heterogeneity across age groups and geographic locations, and provides a framework for assessing the impact of policies targeted towards subpopulations or regions. We define a concept of efficiency for comparative analysis of epidemic control policies and show targeted mitigation policies based on local monitoring to be more efficient than country-level or non-targeted measures. In particular, our results emphasise the importance of shielding vulnerable subpopulations and show that

targeted policies based on local monitoring can considerably lower fatality forecasts and, in many cases, prevent the emergence of second waves which may occur under centralised policies.

4.1 Overview

The ongoing 2019 novel coronavirus pandemic has led to disruption on a global scale, leading to more than 1.4 million deaths worldwide at the time of writing. It prompted the implementation of government policies involving a variety of 'non-pharmaceutical interventions' (Ferguson et al., 2020) including school closures, workplace restrictions, restrictions on social gatherings, social distancing and, in some cases, general lockdowns for extended periods. This has led to a range of different public health policies across the world, and the efficiency of specific policy choices has been subject to much debate.

While the nature of these restrictions has been justified by the severe threat to public health posed by the virus, their design and implementation necessarily involves a trade-off, often implicit in the decision-making process, between health outcomes and the socio-economic impact of such social restrictions.

An important feature of the COVID-19 pandemic has been the heterogeneity of epidemic dynamics and the resulting mortality across different regions, age classes and population categories. The importance of these heterogeneities suggests that homogeneous models – often invoked in discussions on reproduction number and herd immunity – may provide misleading insights, and points to the need for more granular modelling to take into account geographic, demographic and social factors that may influence epidemic dynamics.

We propose a flexible modelling framework which can serve as a decision aid to policy makers and public health experts by quantifying this trade-off between health outcomes and social cost. Using a structured population model for epidemic dynamics which accounts for geographic and demographic heterogeneity, we formulate this trade-off as a control problem for a partially observed distributed system and provide a quantitative framework for comparative analysis of various mitigation policies. We illustrate the usefulness of the framework by applying it to the study of COVID-19 dynamics across regions in England and show how it

may be used to reconstruct the latent progression of the epidemic and perform a comparative analysis of various mitigation policies through scenario projections.

Our modelling framework is founded on the pioneering work of Kermack and McKendrick (1927), which introduces a general class of compartmental models that are widely adapted in modern epidemic modelling. The underlying idea behind these models is to partition a population into compartments with a specific structure dictated by the nature of the disease. A notable example includes the basic Susceptible-Infected-Removed (SIR) model describing diseases that confer immunity against reinfection, and in which hosts are removed either due to lasting immunity after recovery or as a result of their death. In general, transitions between these states can be then modelled either in a deterministic manner, via a system of differential equations, or – as in our case – in a probabilistic manner, via a stochastic (point) process.

Deterministic compartmental models are useful in studying a typical dynamic profile of contagious diseases observed in countless epidemics, where the disease invades a population suddenly, grows exponentially in intensity, and then disappears leaving part of the population untouched (Brauer and Castillo-Chavez, 2012). As such, the behaviour of a population is assumed to be determined completely by its history and the rules that characterise the model. By often assuming homogeneity of various modelling parameters, simple deterministic compartmental models allow researchers to effectively focus on a single population or a particular aspect of the infection, providing insights into key factors that influence the spread of disease.

In particular, an often considered quantity is that of the basic reproduction number $\mathcal{R}_0$ (Diekmann et al., 1950; Anderson and May, 1992). It defines the number of secondary infections caused by a single infected individual over their entire infectiousness period in a completely susceptible population. Similarly, we can also consider the effective reproduction number $\mathcal{R}_e$, which considers a population with both susceptible and immune individuals instead. The basic reproduction number represents the critical threshold required for an epidemic to emerge: if $\mathcal{R}_0 > 1$, there is an epidemic, while if $\mathcal{R}_0 < 1$, the infection dies out. Therefore, the critical fraction of the population that needs to be immunised for the disease to not become an endemic – that is, for the population to reach herd immunity – is given by $1 - 1/\mathcal{R}_0$ (Brauer, 2008). In general, the basic reproduction number $\mathcal{R}_0$

is an important parameter that governs the shape of the typical dynamic profile of an infection.

In consequence, a number of recent studies on the COVID-19 pandemic focus on the impact of governmental intervention on the basic reproduction number $\mathcal{R}_0$. For example, Pindyck (2020) uses the SIR model to provide a basic illustration on how control of the contagion can affect pandemic progression and policy trade-off. In particular, a reduction in $\mathcal{R}_0$, which leads to fewer fatalities, is shown to extend the duration (and thus the economic cost) of the pandemic, as well as the probability of a second wave of infections. Lourenco et al. (2020) calibrate the SIR model to early epidemic data from the UK and Italy and estimate the duration of the pandemic to be around 2–3 months in the absence of any interventions. On the other hand, Roques et al. (2020b) use the SIR model to show that lockdown in France is successful in hindering the growth of the pandemic by significantly reducing the effective reproduction number R_e by a factor of seven. By considering the total fraction of infected individuals, the authors caution against an early relaxation of the lockdown, which can lead to an uncontrolled second wave of infections. However, a prolonged period of large-scale intervention measures aimed at inhibiting transmission of the infection can have a detrimental effect on the economy and social welfare. In light of this, Rowthorn and Maciejowski (2020) consider a cost-benefit analysis of government intervention in a SIR model setting. By quantifying the value of a human life in monetary terms, the authors present a simple methodological framework that determines the optimal path for lifting the restrictions.

Several recent studies on the COVID-19 pandemic also use refined versions of the SIR model that incorporate more complicated compartmental structures, which are better suited for describing the particular characteristics of the disease. Notably, the class of SEIR models allows researchers to take into the account an exposed period between being infected and becoming infectious. In addition, SEIAR models take also into the consideration the fact that infectious individuals may be asymptomatic by decomposing the infectious compartment into symptomatic and asymptomatic sub-compartments. For example, Rawson et al. (2020) use an adapted SEIR model to investigate the efficacy of potential lockdown release strategies in the UK. Their approach is formulated as an optimal control

problem, which aims to maximise the number of people who are able to return to work, subject to the constraint that the number of infectious cases does not exceed the maximum carrying capacity of the health services at any time. A gradual release strategy, in which only a part of the population is released initially, is found to be more effective than the strategy of releasing everyone, followed with another lockdown if infections become too high. Aguilar et al. (2020) consider the SEIAR model to quantify the effect of risk mitigation and presence of asymptomatic carriers on community transmission of COVID-19. By considering the effective reproduction number $\mathcal{R}_e$, the authors show that in the presence of undetected asymptomatic carriers outbreak control becomes more challenging. In particular, although asymptomatic infections may accelerate the transmission of the disease, their presence can also accelerate the emergence of herd immunity in the population.

To account for varying infection, hospitalisation and fatality rates between age groups, several other studies also consider an age-stratified version of compartmental models to analyse the dynamics and impact of the COVID-19 epidemic in various countries (Davies et al., 2020; Prem et al., 2020; Singh and Adhikari, 2020). Notably, Acemoglu et al. (2020) develop a multi-risk SIR model with three age groups for a quantitative analysis of intervention policies. Intuitively, control policies targeting the most vulnerable group are found to significantly reduce the number of fatalities while allowing for a more relaxed measures for the lower-risk groups. Furthermore, a combination of targeted policies with measures that reduce interactions between different age groups, and increase testing and isolation of the infected, is shown to minimise both the economic losses and the number of fatalities. This view is corroborated by several other studies. For example, Lipton and Lopez de Prado (2020) use an age-stratified SEIR model calibrated to early epidemic data from China and the US. The authors advocate the use of targeted lockdowns of the high-risk population. Similarly, Chikina and Pegden (2020) show that strict age-targeted mitigation strategies are effective in reducing the number of fatalities and the strain on ICU utilisation.

Deterministic compartmental epidemic models assume that the size of compartments is sufficiently large such that the mixing of members is homogeneous. However, at the early stage of the disease outbreak, there is a very small number of

infective individuals whose specific pattern of contacts with the susceptible population influences the transmission of infection in a stochastic manner (Brauer, 2008). To address this caveat, some recent studies use instead a probabilistic approach to estimate the effective reproduction number $\mathcal{R}_e$. For example, Kucharski et al. (2020) fit a stochastic SEIR model to publicly available data sets on early cases in Wuhan, China and internationally exported cases from Wuhan. Based on the estimated median daily reproduction number, the authors show that COVID-19 transmission declined in Wuhan during late January 2020, which coincides with the introduction of travel control measures. The study of Donnat and Holmes (2020) emphasises that stochastic models provide a more complete description of the epidemic situation across countries than most SEIR-based models. In particular, probabilistic approaches are shown to detect more refined nuances between scenarios, such as a potential reduction in 'superspreader events'.

Our framework, while compatible with such homogeneous models at the aggregate level, accounts for demographic and spatial heterogeneity in a more detailed manner, leading to regional outcomes which may substantially deviate from homogeneous models. Similar, though somewhat less detailed, heterogeneous models have been recently used to study COVID-19 outbreaks by Danon et al. (2020) for the UK, Birge et al. (2020) for New York City, and Roques et al. (2020a) for France.

We first present below an overview of the main features of our methodology and the key findings, before going into more detail on the approach and results.

4.1.1 Methodology

We formulate a stochastic compartmental (SEIAR) epidemic model with spatial and demographic heterogeneity (age stratification) for modelling the dynamics of the COVID-19 epidemic and apply this model to the study of COVID-19 dynamics across regions in England.

The model takes into account:

- Epidemiological features estimated by previous studies on COVID-19;
- The lack of direct observability of the total number of infectious cases and the presence of a non-negligible fraction of asymptomatic cases;

- The demographic structure of UK regions (age distribution, density);
- Social contact rates across age groups derived from survey data;
- Data on inter-regional mobility; and
- The presence of other random factors, not determined by the above.

We first demonstrate that this model is capable of accurately reproducing the early regional dynamics of the disease, both pre-lockdown[1] and a month into lockdown, using a detailed calibration procedure that accounts for demographic heterogeneity across regions, low testing rates, and the existence of asymptomatic carriers. The calibration reveals interesting regional patterns in social contact rates before and during lockdown.

Underlying any public health policy is a trade-off between the health outcome – which may relate to mortality or hospitalisations – and the socio-economic impact of measures taken to mitigate the magnitude of the impact on public health. We present an explicit formulation of this trade-off and use it to perform a comparative analysis of various 'social distancing' policies, based on two criteria:

- The benefit, in terms of reduction in projected mortality; and
- The cost, in terms of restrictions on social contacts.

The goal of our analysis is to make explicit the policy outcomes for decision-makers, without resorting to (questionable) concepts such as the 'economic value of human life' used in some actuarial and economic models (Acemoglu et al., 2020; Pindyck, 2020; Rowthorn and Maciejowski, 2020).

In our comparative analysis, we consider a broad range of policies and pay particular attention to population-wide versus *targeted* mitigation policies, feedback control based on the number of observed cases and the benefits of broader testing. We introduce the concept of an *efficient* policy, and show how this concept allows the identification of decision parameters which lead to the most efficient outcomes for each type of mitigation policy. The granular nature of our model, together with validation based on epidemiological data, provide a more detailed picture of the relative merits of various public health policies.

[1]Here we refer to the March 2020 lockdown in the UK.

4.1.2 Summary of Findings

Our first set of results concerns the reconstruction of the pandemic progression in England, in particular its latent spread through asymptomatic carriers.

- Using a baseline epidemic model consistent with epidemiological data and observations on fatalities and cases reported in England up to June 2020, we estimate more than 17.8 million people in England (31.7% of the population) have been exposed to COVID-19 by August 1, 2020. These estimates are much higher than the numbers discussed in media reports, based on the number of *reported* cases.

- Based on a comparison of fatality counts and reported cases, we infer that less than 5% of cases in England had been detected prior to June 2020. This low detection probability implies in particular that the number of reported cases may severely underestimate the latent progression of the epidemic.

- We observe significant differences in epidemic dynamics across regions in England, with higher fatality and contagion levels in northern regions than southern regions, both before and during the lockdown period. This points to the importance of demographic and geographic heterogeneity for modelling the impact of COVID-19.

After calibrating the model to replicate the regional progression of COVID-19 in England for the period March 1 to May 31, 2020, we use it for scenario projections under various mitigation policies. Comparative analysis of mitigation policies reveals that measures targeting subpopulations – such as regions with outbreaks – are more efficient than population-wide measures in terms of the trade-off between health outcomes and socio-economic cost. More specifically:

- Shielding of senior populations is by far the most effective single measure for reducing the number of fatalities.[2]

[2]We do recognise that the implementation of such shielding measures may be extremely challenging in practice.

- By contrast, school closures and workplace restrictions are seen to be less effective than social distancing measures outside of school and work environments.

- Adaptive policies ('feedback control') which trigger measures when the number of daily observed cases exceed a threshold, are shown to be more effective than pre-planned policies, leading to a substantial improvement in health outcomes. Since such policies are based on the monitoring of new cases, broader testing significantly improves their outcome.

- A decentralised policy which triggers regional confinement measures based on regional daily reported cases is found to be more efficient than centralised policies based on national indicators, resulting on average in an overall reduction of 20,000 in fatalities and, in many cases, significant damping of a 'second wave'.

- Comparative analysis of policies (Table 4.10) shows a wide range of health outcomes. The most effective policy in terms of reducing fatalities involves decentralised triggering of regional confinement measures based on the monitoring of new cases, coupled with the shielding of senior populations.

The present work should be seen as an illustration of what may be done using our methodology, rather than an exhaustive analysis of different policy options and scenarios. We have made available an online implementation of the model, which may be used to explore other scenarios and policies than those presented below:

`http://covid19.kotlicki.pl`

4.1.3 Outline

The remainder of this chapter is organised as follows. The modelling framework is described in Section 4.2. Data sources and parameter estimations are detailed in Section 4.3. Section 4.4 highlights the implications of partial observability of state variables and the associated model uncertainty. Section 4.5 discusses the

counterfactual scenario of no intervention, which serves as a benchmark to evaluate the impact of social distancing policies.

The outcomes of various epidemic control policies are then discussed in Sections 4.6 and 4.7. Pre-planned policies are discussed in Sections 4.6.1 and 4.6.2, while Section 4.7 discusses *adaptive* ('feedback') control policies, in which measures are triggered when the daily number of new reported cases exceeds a threshold, and concludes with a comparative analysis of health outcomes and social cost of various types of mitigation policies.

4.2 Modelling Framework

To take into account the role of geographic and demographic heterogeneity, we use a stochastic compartmental (SEIAR) model with age stratification, mobility across sites, social contact across age stratification, and the impact of asymptomatic infected individuals. For general concepts on deterministic and stochastic compartmental models we refer to Anderson and May (1992); Brauer and Castillo-Chavez (2012); Britton et al. (2019); Lloyd and Jansen (2004).

4.2.1 State Variables

We consider a regional meta-population model with K regions labeled $r = 1, \ldots, K$. Each region r has a population $N(r)$ which is further subdivided into G age classes labeled $a \in \{1, 2, 3, 4, \ldots, G\}$. We denote $N(r, a)$ the population in region r in age category a, with $\sum_{a=1}^{G} N(r, a) = N(r)$.

Individuals in each region and age group are categorised into six compartments:

1. Susceptible (S) individuals who have not yet been exposed to the virus;

2. Exposed (E) individuals who have *contracted* the virus but are not yet infectious. Exposed individuals may then become *infectious* after a certain incubation period;

3. Infectious (I) individuals who manifest symptoms;

4. Asymptomatic (A) infectious individuals;

5. Recovered (R) individuals. In line with current experimental and clinical observations on COVID-19, we shall assume that individuals who have recovered have temporary immunity, at least for the horizon of the scenarios considered, and cannot be re-infected (Bao et al., 2020); and

6. Deceased (D) individuals.

The progression of the disease in the population is monitored by keeping track of the respective number of individuals in each compartment, denoted by

$$S_t(r,a), \quad E_t(r,a), \quad I_t(r,a), \quad A_t(r,a), \quad R_t(r,a), \quad D_t(r,a).$$

As the model focuses on the dynamics of the epidemic over a short period (1,000 days), we neglect demographic changes over this period and assume that the population size $N(r,a)$ in each location and age group is approximately constant, that is

$$S_t(r,a) + E_t(r,a) + I_t(r,a) + A_t(r,a) + R_t(r,a) + D_t(r,a) = N(r,a)$$

is constant.

4.2.2 A Metapopulation SEIAR Model

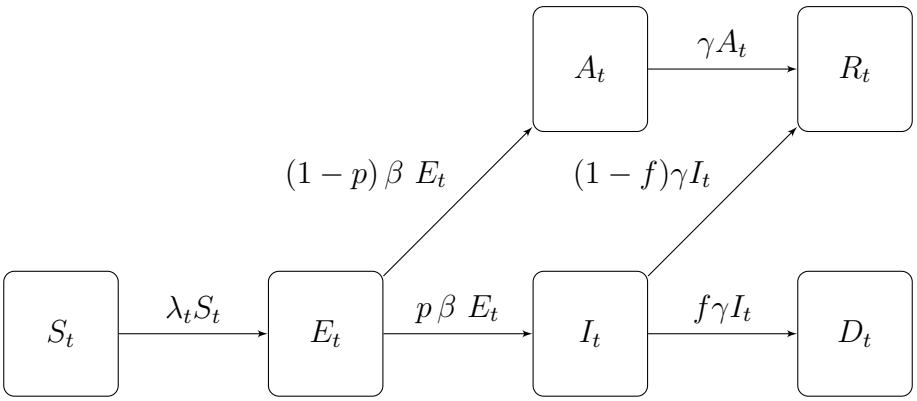


Figure 4.1: Epidemic dynamics in a compartmental model.

When each subpopulation (r,a) is large and homogeneous, the dynamics of state variables may be described through the following system of equations, represented

in Figure 4.1:

$$\begin{cases} \dot{S}_t(r,a) = -\lambda_t(r,a)\, S_t(r,a), \\ \dot{E}_t(r,a) = \lambda_t(r,a)\, S_t(r,a) - \beta E_t(r,a), \\ \dot{I}_t(r,a) = p_a \beta E_t(r,a) - \gamma I_t(r,a), \\ \dot{A}_t(r,a) = (1-p_a)\beta E_t(r,a) - \gamma A_t(r,a), \\ \dot{D}_t(r,a) = \gamma f_a I_t(r,a), \\ \dot{R}_t(r,a) = \gamma(1-f_a) I_t(r,a) + \gamma A_t(r,a) \\ N(r,a) = S_t(r,a) + A_t(r,a) + E_t(r,a) + I_t(r,a) + R_t(r,a) + D_t(r,a). \end{cases} \tag{4.1}$$

where

- β is the incubation rate, and $1/\beta$ is the average incubation period;
- γ is the rate at which infectious individuals recover;
- $0 < p_a < 1$ is the probability for an infected individual in age group a to develop symptoms;
- f_a is the infection fatality rate for age group a, representing the probability that an infected individual in age group a dies from the disease; and
- $\lambda_t(r,a)$ is the *force of infection*, which measures the rate of exposure at location r for age group a, and is given by

$$\begin{aligned} \lambda_t(r,a) = \alpha \sum_{a' \notin \mathcal{W}} \sigma^r_{a,a'}(t) \frac{\kappa I_t(r,a') + A_t(r,a')}{N(r,a')} \\ + \alpha \sum_{a' \in \mathcal{W}} \sigma^r_{a,a'}(t) \sum_{r'=1}^{K} M_{r,r'}(t) \frac{\kappa I_t(r',a') + A_t(r',a')}{N(r',a')}, \end{aligned} \tag{4.2}$$

  where $0 < \alpha < 1$ is the infection rate per contact (that is the probability of infection conditional on contact), and $\mathcal{W}$ represents the set of age groups with a working population. We introduce a reduction of κ in a contact rate of infectious symptomatic individuals due to (partial) self-isolation.

The force of infection in each subpopulation (r,a) depends on the rate of contact with (infectious) individuals in other subpopulations, which differentiates this model from a homogeneous model. These interactions occur through:

- *Contacts across age groups in the same region*: the term $\sigma^r_{a,a'}(t)$ represents the average number of persons from age class a' encountered per day by a person from age class a in region r on a day t. For infectious individuals with symptoms, we assume a lower contact rate $\kappa\sigma < \sigma$ due to (partial) self-isolation. This leads to the first term in (4.2).

- *Inter-regional mobility*: the second term in (4.2) corresponds to contacts between individuals in region r and age class a and those in the working population (age classes $a' \in \mathcal{W}$) commuting from other regions $r' \neq r$. $M_{r,r'}(t)$ represents the proportion of individuals from region r' among the population of adults at a location r on date t.

4.2.3 Stochastic Dynamics

The deterministic dynamics (4.1) ignore the variability of outcomes (Isham, 1991) due to random factors not being taken into account in the model. To account for this variability of outcomes we model the variables $\{S(t), E(t), I(t), A(t)\}$ as a continuous-time Markov point process (Allen, 2017; Britton et al., 2019) defined through its transition rates conditional on the history $\mathcal{H}_t$ up to date t:

$$\begin{cases} \mathbb{P}(\Delta S_t(r,a) = -1|\mathcal{H}_t) = -\lambda_t(r,a)\, S_t(r,a)\Delta t + o(\Delta t) \\ \mathbb{P}(\Delta E_t(r,a) = 1|\mathcal{H}_t) = \lambda_t(r,a)\, S_t(r,a)\Delta t + o(\Delta t) \\ \mathbb{P}(\Delta E_t(r,a) = -1|\mathcal{H}_t) = \beta E_t(r,a)\Delta t + o(\Delta t) \\ \mathbb{P}(\Delta I_t(r,a) = 1|\mathcal{H}_t) = p_a \beta E_t(r,a)\Delta t + o(\Delta t) \\ \mathbb{P}(\Delta I_t(r,a) = -1|\mathcal{H}_t) = \gamma I_t(r,a)\Delta t + o(\Delta t) \\ \mathbb{P}(\Delta A_t(r,a) = +1|\mathcal{H}_t) = (1-p_a)\beta E_t(r,a)\Delta t + o(\Delta t) \\ \mathbb{P}(\Delta A_t(r,a) = -1|\mathcal{H}_t) = \gamma A_t(r,a)\Delta t + o(\Delta t) \\ \mathbb{P}(\Delta D_t(r,a) = 1|\mathcal{H}_t) = f_a \gamma I_t(r,a)\Delta t + o(\Delta t) \end{cases} \tag{4.3}$$

where $\mathcal{H}_t$ is formally defined as the natural filtration generated by the state variables and their intensities.

The stochastic dynamics (4.3) are consistent with the deterministic dynamics of (4.1) for large populations, in the sense that the population fractions represented by each compartment converge to those represented by the solution of (4.1) as $\min_r N(r)$ increases. However, even when the overall population is large, the

stochastic dynamics (4.3) can substantially deviate from the deterministic model (4.1), especially in small subpopulations and in the early phases of the epidemic when the number of infectious individuals in each region may be small, leading to random flare-ups and breakouts not present in the deterministic model. In the sequel we use the stochastic model (4.3) for the dynamics of the state variables.

4.2.4 Policies for Epidemic Control

Social distancing policies (and lockdowns) affect epidemic dynamics by influencing (lowering) the social contact rates σ_{ij}^r and the inter-regional mobility $M_{r,r'}$. To discuss targeted policies which may influence social contact rates at different locations differently, we decompose the baseline social contact matrix σ^r as

$$\sigma^r(0) = \sigma^{r,H} + \sigma^{r,W} + \sigma^{r,S} + \sigma^{r,O}, \tag{4.4}$$

where the components correspond respectively to contacts at home ($\sigma^{r,H}$), at work ($\sigma^{r,W}$), school ($\sigma^{r,S}$) and other locations ($\sigma^{r,O}$). Social distancing policies are then parameterised in terms of their impact on various components of the social contact matrix:

$$\sigma_{ij}^r(t) = u_{ij}^{r,H}(t)\sigma_{ij}^{r,H} + u_{ij}^{r,S}(t)\sigma_{ij}^{r,S} + u_{ij}^{r,W}(t)\sigma_{ij}^{r,W} + u_{ij}^{r,O}(t)\sigma_{ij}^{r,O} \leq \sigma_{ij}^r(0), \tag{4.5}$$

where $0 \leq u_{ij}^{r,X}(t) \leq 1$ are modulating factors that measure the impact of the policy on social contacts between age groups i and j at a location X in region r. In absence of social distancing or confinement measures, we have $u_{ij}^{r,X}(t) = 1$; the value of $u_{ij}^{r,X}(t)$ reflects the fraction of social contacts between age groups i and j at location X in region r when the policy is applied.

This parameterisation allows us to consider policies targeted towards subpopulations or specific regions. For example, school closure in region r during time period $[t_1, t_2]$ corresponds to setting $u_{ij}^{r,S}(t) = 0$ for $t \in [t_1, t_2]$, while $0 < u_{ij}^{r,S} < 1$ corresponds to social distancing in schools, with lower values of $u_{ij}^{r,S}$ corresponding to stricter enforcement of measures.

In most cases $u_{ij}^{r,X}(t)$ does not explicitly depend on the age groups i, j as it is unfeasible to discriminate between age groups when implementing social distancing requirements. Dependence on age groups arises when certain types of contacts are primarily related to certain age groups:

- Shielding of senior populations: such policies affect the contact rates between seniors and other age groups.

- Work restrictions, which affect contacts between age groups of the working population (denoted $\mathcal{W}$): $u_{ij}^{r,W}(t) = u^{r,W}\mathbb{1}(i \in \mathcal{W})\mathbb{1}(j \in \mathcal{W})$.

Regarding the inter-regional mobility matrix M, following the interpretation discussed in Section 4.3.2, we modulate its value according to the fraction $u^{r,W}$ of the population who continue to commute, that is

$$M_{r,r'}(t) = u^{r,W}(t)M_{r,r'} + (1 - u^{r,W}(t))\mathbf{I}.$$

Here, $M_{r,r'}$ is the fraction of population in region r whose habitual residence is in region r', and $\mathbf{I}$ denotes the identity matrix.

The modulating factors $u_{ij}^{r,X}$ may be chosen in advance or expressed as a function of the state of the system. We distinguish:

- *Pre-planned* (also called 'open-loop') policies, in which target values of modulating factors $u_{ij}^{r,X}(t)$ are decided in advance; and

- *Adaptive* policies (also called 'closed loop' or feedback control), in which actions are decided and updated as a function of observed quantities such as the number of daily reported cases or number of daily fatalities.

Comparative analysis of mitigation policies To perform comparative analysis across different policies, we need to evaluate policy outcomes across two dimensions: health outcome and socio-economic impact.

We quantify the health outcome of each policy by the total number of fatalities during a reference period, taken to be $t_{\max} = 1000$ days after the reference date of March 1, 2020. The length of this reference period is chosen such that it takes into account an eventual 'second wave' of fatalities. We denote this outcome by $D_{t_{\max}}(u)$, which represent the total fatalities at date $t_{\max}$ associated with policy u.

To quantify the socio-economic impact of a policy, we use as metric the reduction in social contact resulting from the policy over the horizon $[0, t_{\max}]$, that is

$$J(u) = \sum_{t=1}^{t_{\max}} \sum_{r=1}^{K} \sum_{i,j=1}^{G} \Big(\sigma_{ij}^{r}(0) - \sigma_{ij}^{r}(t)\Big) N(r,i), \tag{4.6}$$

defined in terms of person×day units. The definition of our measure $J(u)$ encompasses a wide range of important welfare factors that ought to be taken into consideration when designing intervention policies. For example, reduction in social contacts among the working population has a direct impact on productivity and economic growth. Tourism and hospitality industries are prime examples of businesses that rely on customer interactions, and for which recent pandemic-related restrictions have resulted in significant declines in revenue, insolvencies and job losses. In general, metric $J(u)$ may be treated as a proxy for reduction in economic activity. Furthermore, social interaction has also been proven to play an important role in education (Okita, 2012). Another example stems from literature on social psychology that highlights the importance of a wide range of social interactions in maintaining mental well-being (Sandstrom and Dunn, 2014). In consequence, we argue that $J(U)$ may be used as a suitable proxy for quantifying the adverse effects of lockdowns and control policies on an entire population, which thus demonstrates the inherit trade-off between health outcomes and socio-economic impact.

The range of policies examined below lead to different outcomes in terms of fatalities $D_{t_{\max}}(u)$ and social cost $J(u)$. A policy v *dominates* (or improves upon) a policy u if it leads to a similar or better health outcome at an equal or lower cost:

$$J(v) \leq J(u) \qquad \text{and} \qquad D_{t_{\max}}(v) \leq D_{t_{\max}}(u),$$

with at least one inequality being strict. A policy u is *efficient* among a class of policies U if it cannot be improved upon by any policy in this class. Given a set of policies U, the subset of efficient policies forms the *efficient frontier* of U.

Some recent economic models (Acemoglu et al., 2020; Pindyck, 2020; Rowthorn and Maciejowski, 2020) formulate the trade-off in different terms, by introducing a concept of the monetary *value of human life* in order to build a (monetary) welfare function combining both terms. Aside from ethical issues linked to the very concept of monetisation of human life, there is no consensus on its actual value, which is a key determinant of the trade-off in this approach. Our approach avoids specifying such a value and aims at identifying the range of efficient policies, leaving the final choice of the trade-off to policymakers.

In what follows, the goal is to determine the set of efficient policies and describe the characteristics and outcomes of such policies. Pre-planned policies are discussed in Sections 4.6.1 and 4.6.2, while adaptive policies are discussed in Section 4.7.

4.3 Data Sources and Parameter Estimation

In this section, we detail the model inputs as well as the methodology used in the parameter estimation. Table 4.5 contains a summary of model parameters.

4.3.1 Data Sources

The basic inputs of the model are panel data on number of cases and fatalities reported at the level of Upper Tier Local Authorities (UTLA) in England, provided by the Public Health England and NHSX (2020). This defines the geographic granularity of the model: we partition the population of England into 133 regions as defined by the Nomenclature of Territorial Units for Statistics at level 3 (NUTS-3) (Eurostat, 2020b).

For the purpose of our study we distinguish $G = 16$ age groups, as shown in Table 4.1, which is the maximum granularity allowed by the available estimates of age-dependent social contact rates and fatality rates. The size $N(r, a)$ of age group a in region r is retrieved using the population data set provided by Eurostat (2020a). Appendix B.1 provides the list of UK regions used in this study and outlines the performed mapping procedure from UTLA to NUTS-3 regions to ensure consistency across data sources.

Age group	[0,5)	[5,10)	[10,15)	[15,20)	[20, 25)	[25, 30)	[30, 35)	[35, 40)
Size (millions)	3.3	3.5	3.3	3.1	3.5	3.8	3.8	3.7
Fraction	5.9%	6.3%	5.9%	5.5%	6.2%	6.8%	6.8%	6.6%
Age group	[40,45)	[45,50)	[50,55)	[55,60)	[60, 65)	[65, 70)	[70, 75)	[75, 100)
Size (millions)	3.4	3.8	3.9	3.6	3.1	2.8	2.8	4.7
Fraction	6.0%	6.7%	7.0%	6.5%	5.5%	5.0%	4.9%	8.4%

Table 4.1: Age group distribution for England, 2019. Source: Eurostat (2020a).

4.3.2 Modelling of Inter-Regional Mobility

For our baseline estimate of inter-regional mobility we use the 2011 Census data on location of usual residence and place of work in the United Kingdom, provided by the Office for National Statistics (2020b). The data set classifies people aged 16 and over in employment during March 2011 and shows the movement between their area of residence and workplace, defined in Local Administrative Units at level 1 (LAU-1) terms. We then map this data onto NUTS-3 regions using the lookup table between LAU-1 and NUTS-3 areas provided by the Office for National Statistics (2020c).

The data is represented in the model through the inter-regional mobility matrix M, where elements $M_{r,j}$ represent the fraction of population working in region r whose habitual residence is in region j. Denote by $\Pi(r, j)$ the population with residence registered in region j and workplace registered in region r for $r \neq j$. In addition, we denote by $\Pi(r, r) = \sum_{a \in \mathcal{W}} N(a, r)$, where $\mathcal{W} = \{5, 6, 7, 8, 9, 10, 11, 12\}$ represents the total population at location r in the age category $[20, 60)$ years. Then, we estimate the coefficients of $M_{r,j}$ by

$$\widehat{M_{r,j}} = \frac{\Pi(r, j)}{\sum_{i=1}^{K} \Pi(r, i)}. \tag{4.7}$$

4.3.3 Epidemiological Parameters

Epidemiological parameters were either estimated from publicly available sources (Adhikari et al., 2020; Mossong et al., 2017) or set to values consistent with recent clinical and epidemiological studies in COVID-19 (Dorigatti et al., 2020; Khalili et al., 2020; Verity et al., 2020). In the following discussion, we present in detail the data sources and model calibration approaches used.

Social contact rates Contact rates across age classes have been estimated in studies by Mossong et al. (2008, 2017) and Béraud et al. (2015). A more recent study of Klepac et al. (2020) reports detailed population contact patterns for the United Kingdom based on self-reported contact data from over 36,000 volunteers that participated in the project. We find contact patterns across various locations and age groups to be consistent between Mossong et al. (2008) and Klepac et al.

(2020). Consequently, we use in our study the estimates of social contact rates provided by Mossong et al. (2008) for the 16 age groups defined in Table 4.1. Using the PyRoss methodology (Adhikari et al., 2020), we further decompose the contact matrix, as in (4.4), into four components representing contacts at home (σ^H), work (σ^W), school (σ^S) and other locations (σ^O). Estimation methods and parameter values for these matrices are outlined in Appendix B.2.

Contact rates may vary across different regions due to the heterogeneity in socio-economic composition structure and specific regional characteristics, such as population density, level of urbanisation and the level of use of public transport. To account for this heterogeneity, we parameterise the (pre-lockdown) contact matrix in region r as $\sigma^r(0) = d_r\sigma$, where the regional adjustment factors $\{d_r\}_{r=1}^{133}$ are estimated to reproduce the regional growth rate of reported cases before the lockdown period.[3] The results are displayed in Figure 4.2. Table 4.2 provides a summary of selected characteristics of five regions with the highest values of the regional adjustment factors d_r.

Region	d_r	**Density**	**Inward mobility**	**Outward mobility**	**Population**
UKC12	1.80	925.9 (#59)	17.6% (#82)	19.1% (#105)	276988 (#102)
UKI62	1.68	4518.4 (#67)	16.6% (#87)	43.2% (#12)	389473 (#59)
UKG32	1.64	1205.5 (#59)	14.5% (#96)	46.2% (#6)	215055 (#120)
UKI53	1.62	6161.9 (#11)	23.4% (#50)	41.6% (#16)	587575 (#25)
UKC23	1.52	2026.9 (#49)	13.7% (#103)	24.0% (#78)	277733 (#99)

Table 4.2: Summary of regions with the highest regional multiplier d_r for social contact matrix. The number in the brackets signify the respective rank of the measured quantity.

As seen in Figure 4.2, our findings imply heterogeneity of social contact rates across regions. As we will observe below, these differences have a considerable impact on regional epidemic dynamics.

Incubation rate Following the study of Ferguson et al. (2020), we use an incubation rate $\beta = 0.2$, which corresponds to an incubation period of approximately 5 days. This is further supported by several empirical studies on diagnosed cases

[3]See Appendix B.2 for more details.

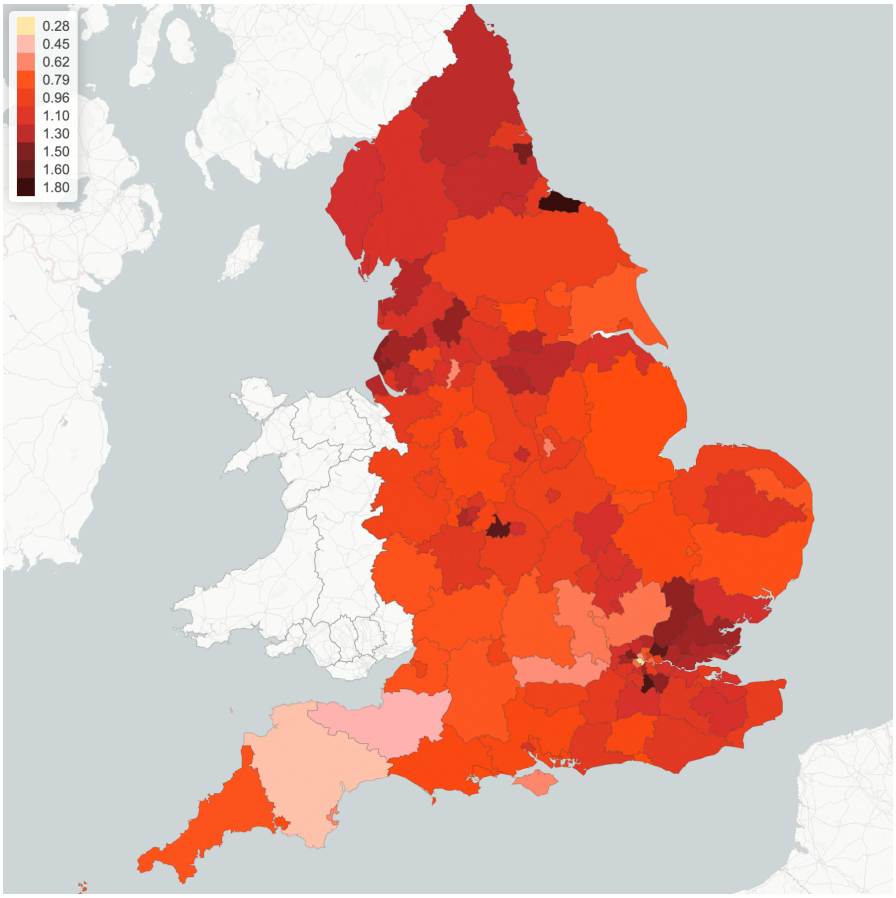


Figure 4.2: Regional multiplier d_r for social contact matrix, implied by epidemic dynamics pre-lockdown (before March 23, 2020).

in China outside the Hubei province. An early study of Backer et al. (2020) based on 88 confirmed cases, which uses data on known travel to and from Wuhan, China to estimate the exposure interval, indicates a mean incubation period of 6.4 days with a 95% confidence interval (CI) of 5.6–7.7 days. Linton et al. (2020), based on 158 confirmed cases in China reported through January 31, 2020, estimate a median incubation period of 5.0 days with 95% CI of 4.4–5.6 days, and an incubation period with a mean of around 5 days and a 95% CI of 4.2–6.0 days. Lauer et al. (2020) estimate a median of incubation period to be 5.1 days with 95% CI of 4.5–5.8 days, based on 181 cases from 50 regions and countries outside Wuhan over the period of January 4 to February 24, 2020.

Epidemiological parameters, such as the incubation rate, are crucial for under-

standing the transmission dynamics of the virus. However, these often are hard to estimate at an early point in a pandemic as the data is limited. Furthermore, existing estimates may vary depending on the number of study participants, the data collection time period, geographical location of the study, and the methodology. Consequently, quality of the estimates may be increased significantly by combining findings of existing studies (Alene et al., 2021). In particular, a more recent work of Alene et al. (2021) uses 23 observational studies to estimate a mean incubation period of 6.5 days with a 95% CI of 5.9–7.1 days. A majority of the included studies in this work are conducted in China (14 out of 23), and most are based on the data within the January to early March time period.

Proportion of symptomatic and asymptomatic infections The probability p that an infected individual develops symptoms is an important parameter for epidemic dynamics, yet it is subject to a high degree of uncertainty: studies on various data sets (Buitrago-Garcia et al., 2020; Davies et al., 2020; Mizumoto et al., 2020; Flaxman et al., 2020; Office for National Statistics, 2020a) are based on small samples and yield a wide range of estimates. In particular, early estimates from the Diamond Princess cruise ship (Mizumoto et al., 2020) and Japanese evacuation flights from Wuhan yielded estimates as high as $p \simeq 0.7$-0.8 (Nishiura et al., 2020), while a July 2020 study by the Office for National Statistics (2020a), based on a much larger sample, show that p can be as low as 0.23. However, clinical studies (Davies et al., 2020) indicate that this probability may strongly depend on the age group considered.

In order to capture the sensitivity of our results to these estimates, we use a range of values for the age-dependent probability p_a whose upper bound is consistent with Davies et al. (2020) and whose lower bound is consistent with the estimates provided by the Office for National Statistics (2020a). These values are displayed in Table 4.3. Given the much larger sample size used in the study of Office for National Statistics (2020a), we use the corresponding estimates ('low values', denoted as p_{low} in Table 4.3) as benchmark unless stated otherwise.

Age group	[0,5)	[5,10)	[10,15)	[15,20)	[20, 25)	[25, 30)	[30, 35)	[35, 40)
p_{low}	0.075	0.075	0·05	0·05	0.15	0.15	0.21	0.21
p_{high}	0.15	0.15	0.1	0.1	0.3	0.3	0.42	0.42
Age group	[40,45)	[45,50)	[50,55)	[55,60)	[60, 65)	[65, 70)	[70, 75)	[75, 100)
p_{low}	0.23	0.23	0.28	0.28	0.41	0.41	0.375	0.375
p_{high}	0.45	0.45	0.56	0.56	0.82	0.82	0.75	0.75

Table 4.3: Age-dependent symptomatic ratios, p. Source: Office for National Statistics (2020a) and Davies et al. (2020).

Recovery rate γ In line with Cao et al. (2020); Li et al. (2020) and Rocklöv et al. (2020), we use a recovery rate $\gamma = 0.1$, which corresponds to an average infectious period of 10 days. However, we note the difficulty in estimating the recovery rate at the beginning of the pandemic due to data limitations. For example, the estimate in Li et al. (2020) is based on the first 425 confirmed cases in Wuhan. The estimate of Cao et al. (2020) is set in accordance with the Chinese Center for Disease Control and Prevention at the time.

Infection fatality rates We denote by f_a the (infection) fatality rate for age group a. In practice, these parameters are difficult to estimate during outbreaks and estimates may be subject to various biases (Lipsitch et al., 2015). Note that the *infection* fatality rate (IFR) is different from (and generally much smaller than) the *case* fatality rate.

Fatality rates for COVID-19 have been observed to be highly variable across age groups (Khalili et al., 2020; Salje et al., 2020; Verity et al., 2020). Based on the infection fatality rates provided in Verity et al. (2020) for different age groups and the UK population distribution, we derive the aggregated IFR for the respective 16 age groups of interest as summarised in Table 4.4. The estimates are based on individual-case data from Hubei, China reported on February 8, 2020, as well as on cases from 37 other countries as reported until February 25, 2020. These results are consistent with data obtained from other countries; for example, see the study of Salje et al. (2020) using French hospitalisation and death data as of May 7, 2020.

Age group	[0,5)	[5,10)	[10,15)	[15,20)	[20, 25)	[25, 30)	[30, 35)	[35, 40)
IFR f (%)	0.002	0.002	0.01	0.01	0.05	0.05	0.1	0.1
Age group	[40,45)	[45,50)	[50,55)	[55,60)	[60, 65)	[65, 70)	[70, 75)	[75, 100)
IFR f (%)	0.2	0.2	0.6	0.6	2.00	2.00	4.0	7.5

Table 4.4: Age-dependent infection fatality rates. Source: Verity et al. (2020).

Infection rate	α	0.055 (0.051,0.062)
Source: Donnat and Holmes (2020); Dorigatti et al. (2020)		
Incubation rate	β	0.2
Source: Ferguson et al. (2020); Lauer et al. (2020); Linton et al. (2020)		
Recovery rate	γ	0.1
Source: Cao et al. (2020); Li et al. (2020); Rocklöv et al. (2020)		
Infection fatality rate	f	See Table 4.4
Source: Verity et al. (2020)		
Symptomatic ratios (low estimate)	p_{low}	See Table 4.3
Source: Office for National Statistics (2020a)		
Symptomatic ratios (high estimate)	p_{high}	See Table 4.3
Source: Davies et al. (2020)		
Social contact matrix	σ	See Appendix B.2
Source: Mossong et al. (2008)		
Symptomatic contact adjustment	κ	0.5
Regional adjustment for contact rates (pre-lockdown time period)	d_r	See Figure 4.2
Regional adjustment for contact rates (lockdown time period)	l_r	See Table 4.6
Inter-regional mobility matrix	M	
Source: Office for National Statistics (2020b)		

Table 4.5: Summary of parameters for the COVID-19 model.

4.3.4 Estimation of the Infection Rate

We use a simulation-based indirect inference method (Gourieroux et al., 1993) for estimating the infection rate α. We simulate the stochastic model (4.3) for a range of values $0.03 \leq \alpha \leq 0.15$. The value of α is estimated by matching the logarithmic growth rates of the simulated reported cases with that of reported

cases in England.

In particular, our data consists of the cumulative number of reported cases C_t in England and C_t^r in the NUTS-3 region r, up to date t included. Since infectious disease outbreaks are characterised by a typical epidemic profile exhibiting an exponential growth (Chowell et al., 2016), we use a logarithmic ordinary least square (OLS) regression to estimate the growth rate of C_t and C_t^r between March 9 and April 2, 2020. The estimated slope and intercept for England are $\widehat{b} = 0.165$ and $\widehat{c} = 7.25$ with $R^2 = 0.982$. Figure 4.3 shows the distribution of regional slope estimates $\{\widehat{b}_r\}_{r=1}^{133}$.

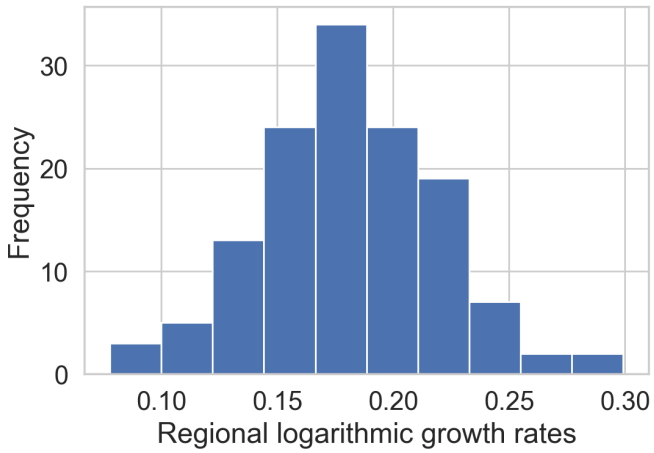


Figure 4.3: Distribution of the estimated logarithmic growth rates $\widehat{b}_r$ in different regions.

To obtain a confidence interval for α we use a simulation-based indirect inference approach of Gourieroux et al. (1993). That is, we perform a parameter sweep, where we simulate the dynamics of our stochastic SEIAR model (4.3) for $\alpha \in [0.03, 0.04, \ldots, 0.15]$. For each α on the grid and every region $r = 1, \ldots, 133$, we consider Z paths simulated independently from our model. For each simulation path $z = 1, \ldots, Z$, we estimate the slope of $\log C_t^r(\alpha)$, denoted by $\hat{b}_r^{(k)}(\alpha)$, by fitting a logarithmic OLS model to the simulated data. We use an analogous procedure to estimate the logarithmic growth rate in England, $\hat{b}^{(k)}(\alpha)$. Consequently, the estimates obtained are used to construct a 95% confidence interval (CI) for the

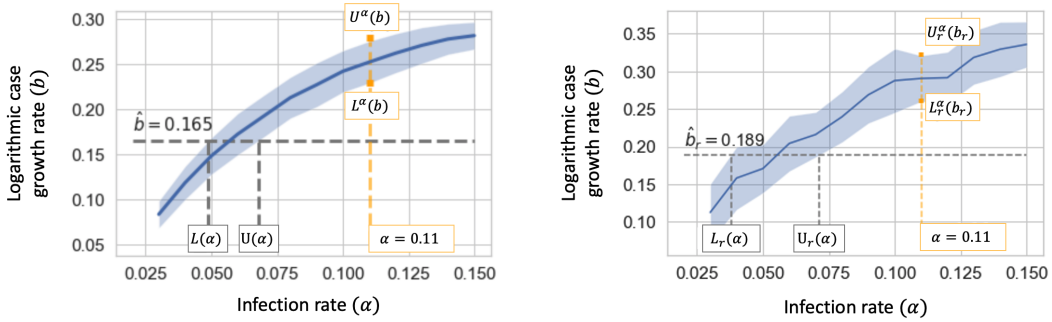
slope b_r in each region r, which we denote by $(L_r^\alpha(b_r), U_r^\alpha(b_r))$; and the 95% CI for England, denoted by $(L^\alpha(b), U^\alpha(b))$.

For the above simulation procedure, we use parameters specified in Table 4.5 and the following initial conditions for $t_0 =$ March 10, 2020:

$$E_{t_0}(r,a) = \frac{N(r,a)}{\sum_{a' \in \mathcal{W}} N(r,a')} \frac{C_{t_0+5}(r)}{p_a\, \pi}\ \mathbb{1}(a \in \mathcal{W}), \tag{4.8}$$

where $\mathcal{W} = \{5,6,7,8,9,10,11,12\}$ corresponds to a set of age groups in the working population, and $A_{t_0}(r,a) = 0$, $D_{t_0}(r,a) = 0$, $I_{t_0}(r,a) = 0$ for all a. These initial conditions ensure that the simulations agree on average with regional case numbers on March 15, 2020, for all values of α.

Recall that $\widehat{b}_r$ is the estimate of the regional logarithmic growth based on reported cases data in region r. By considering all values of α such that $\widehat{b}_r \in (L_r^\alpha(b_r), U_r^\alpha(b_r))$, we can then infer the corresponding CI $(L_r(\alpha), U_r(\alpha))$ for α in each region r. Similarly we construct a CI for α , denoted by $(L(\alpha), U(\alpha))$, based on the slope estimate $\widehat{b}$ of England (see Figure 4.4).



(a) Estimates for England.

(b) Estimates for Leeds (UKE44).

Figure 4.4: Indirect inference method for estimating the infection rate α. By simulating epidemic dynamics for a fixed $\alpha \in [0.03, 0.15]$, we compute a 95% CI for the logarithmic epidemic growth rate b, denoted $(L^\alpha(b), U^\alpha(b))$. Then, we infer the 95% CI for α, denoted $(L(\alpha), U(\alpha))$, by matching growth of reported cases with our simulations. Presented results are based on $K = 50$ simulations.

Our procedure with $Z = 50$ simulation paths yields an estimated value of $\widehat{\alpha} = 0.055$ and a confidence interval $(0.051, 0.062)$ for England. Regional estimates are summarised in Figure 4.5. For the majority of the regions, there is an overlap with

the 95% CI for England. We note that our results are consistent with estimates obtained in Donnat and Holmes (2020) and Dorigatti et al. (2020) using data from other countries.

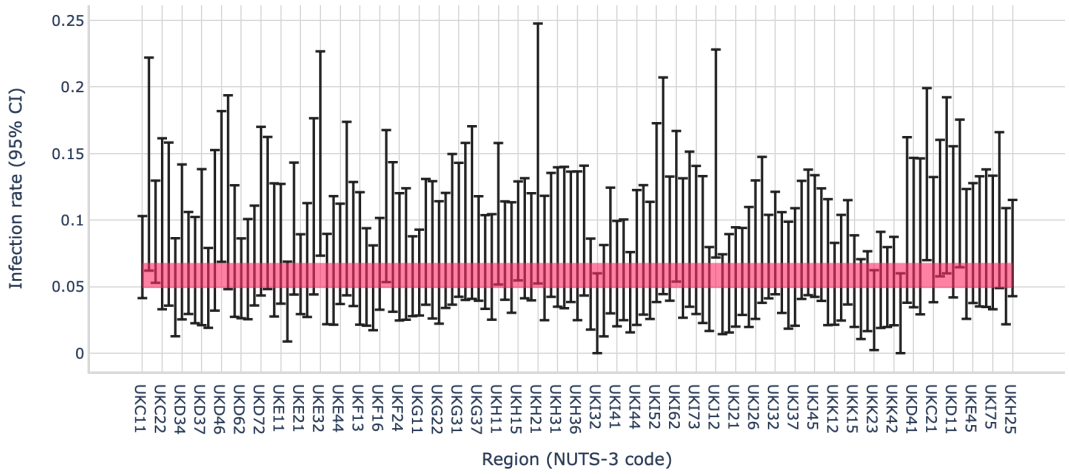


Figure 4.5: Estimated 95% confidence intervals for regional values of the infection rate, α. The red area corresponds to a 95% CI of England.

4.3.5 Inter-Regional Mobility and Social Contact During Confinement

Confinement measures were implemented across the United Kingdom starting March 23, 2020 via the Coronavirus Act.[4] During this lockdown period schools and workplaces were closed and social contact was reduced, as evidenced by mobility data.[5] However mobility data also reveals regional differences in the impact of the lockdown.

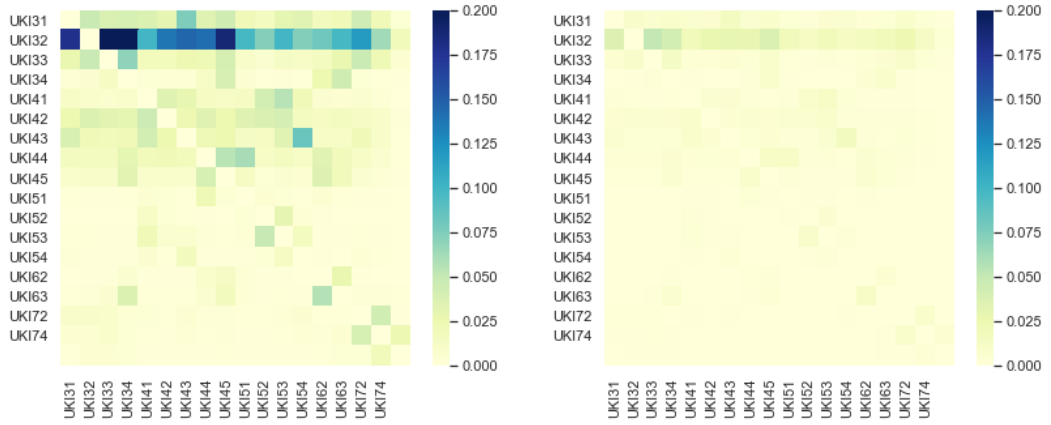
We model the reduction in inter-regional mobility through an adjusted mobility matrix

$$\widehat{M}_{r,r'}(t) = q\ \widehat{M}_{r,r'} + (1-q)\mathbf{I}, \qquad \text{with} \quad 0 < q < 1, \tag{4.9}$$

and where $\widehat{M}_{r,r'}$ is the inter-regional mobility matrix defined in (4.7). According to the Labor Force Survey data from 2018/19 (Farquharson et al., 2020), 7.1 million

[4]See `https://www.legislation.gov.uk/ukpga/2020/7/contents/enacted`.
[5]See `https://www.oxford-covid-19.com/`.

adults across the UK are considered as 'key workers'. We set $q = 20\%$ to take into account the fact that these key workers continued to access their workplace during the lockdown period. This is also consistent with the methodology in Rawson et al. (2020) and empirical studies of Santana et al. (2020) on mobility changes before and after lockdown in the UK. Figure 4.6 shows the submatrix corresponding to daily mobility across London boroughs, and illustrates the observed dramatic drop in commute patterns.



(a) Pre-lockdown: before March 23, 2020. (b) During lockdown: from March 23 to June 10, 2020.

Figure 4.6: Inter-regional mobility across London boroughs.

We model the impact of confinement on the social contact matrix through a regional multiplier l_r,

$$\sigma^r(t) = l_r \times \sigma(0), \tag{4.10}$$

where $l_r \leq d_r$ represents the reduction in social contacts during the lockdown period; $l_r = d_r$ corresponds to the pre-lockdown level of social contact. The value of l_r is estimated from panel data on regional epidemic dynamics during the period from March 23 to June 1, 2020, using a least-squares logarithmic regression on the number of observed regional cases (see Appendix B.2 for more details).

The average value of this reduction factor is found to be

$$\frac{\sum_{r=1}^{133} N(r) l_r}{\sum_{r=1}^{133} N(r)} = 0.12,$$

which is an average reduction of 88% in social contacts, an order of magnitude that is corroborated by mobility data (Santana et al., 2020), showing that the lockdown was very effective in reducing social contacts. Refer to Table 4.6 for a summary of regional adjustments across the regions in England.

NUTS-1 Region	**Pre-lockdown (d_r)**	**Lockdown (l_r)**
South West (UKK)	0.729	0.099
East Midlands (UKF)	0.952	0.134
London (UKI)	1.143	0.100
West Midlands (UKG)	1.020	0.126
Yorkshire and Humber (UKE)	1.069	0.137
South East (UKJ)	0.920	0.116
North East (UKC)	1.260	0.131
North West (UKD)	1.122	0.137
East of England (UKH)	0.994	0.129

Table 4.6: Estimated values for regional adjustments d_r and l_r in NUTS-1 regions.

4.3.6 Goodness-of-Fit

Having estimated the model parameters using data on reported cases between March 10 and May 20, 2020 we assess the goodness-of-fit and out-of-sample performance using reported cases and fatalities between May 21 and June 22, 2020. Figures 4.7 and 4.8 show that the model is able to reproduce the in-sample and out-of-sample evolution of case numbers and fatalities at both national and regional level.

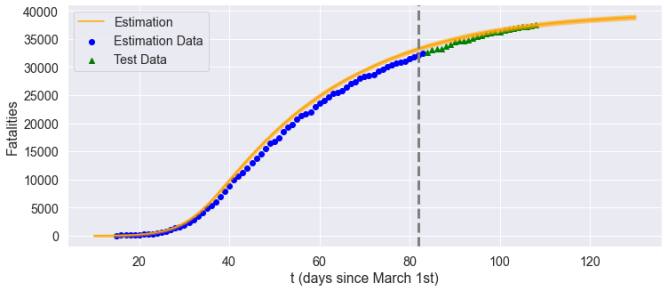
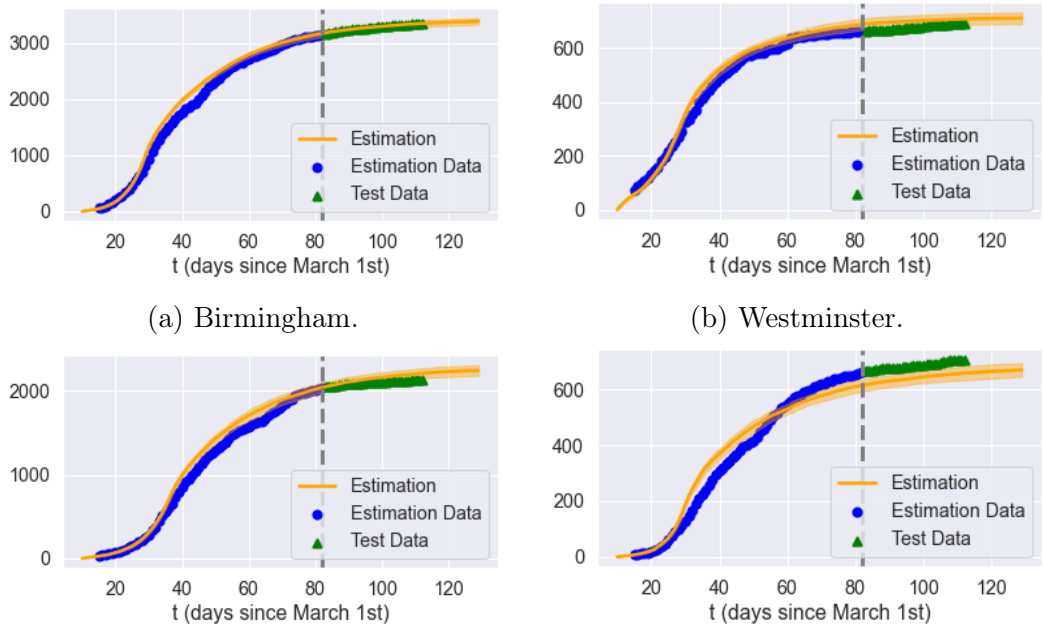


Figure 4.7: Fatalities in England: comparison of model with data. Grey dashed line: separation between estimation sample and test data; orange line: model simulation; blue dot: in-sample data; green triangle: out-of-sample data.



(a) Birmingham.

(b) Westminster.

(c) Oxfordshire.

(d) West Essex.

Figure 4.8: Cumulative reported cases in selected regions. Grey dashed line: separation between estimation sample and test data; orange line: average of 50 simulated scenarios; blue dot: in-sample data; green triangle: out-of-sample data.

4.4 Observable Quantities and Uncertainty

When applying such models to epidemic data, a key point is to realise that the state variables S, E, I, A, R are not directly observed (and certainly not in real time) but need to be inferred from other observable quantities.

In absence of widespread testing, public health authorities are faced with the problem of controlling a system under partial observation. This lack of direct observability has some implications for the estimation and interpretation of the model, which we briefly discuss here.

4.4.1 Observable Quantities

The two main observable quantities in COVID-19 data are

- The cumulative number of *reported* cases; and

- The cumulative number of COVID-19 fatalities D_t.

Of the two, fatalities are generally considered more reliable, as deaths are nearly always reported, while identification of cases requires testing or self-reporting. We thus identify the observed number of fatalities with the state variable D_t.

In absence of widespread testing, only a fraction π of cases is reported. This fraction may change with time due to testing campaigns.[6] We therefore cannot assume the number of infectious cases to be directly observed: rather, we estimate it from the fatality count D_t (see also Jombart et al. (2020)).

Let C_t be the cumulative number of (symptomatic) infectious cases. Assuming that

- The daily number $r(t)$ of reported cases is a fraction $\pi(t)$ of new cases, that is
$$r(t) = \pi(t)\,(C_{t+1} - C_t)\,; \tag{4.11}$$
- Deaths occur on average T days after detection;

we obtain that the daily fatality count is proportional to the lagged number of new cases,
$$D_{t+T+1} - D_{t+T} \simeq \overline{f}\,(C_{t+1} - C_t) = \frac{\overline{f}}{\pi(t)}\, r(t), \tag{4.12}$$
where $\overline{f}$ is the (average) infection fatality rate. We use these relations to obtain an estimate for the cumulative number C_t of symptomatic infections and the reporting ratio $\pi(t)$.

Using (4.12), we estimate the average delay T between case reporting and death by identifying the lag T which maximises the correlation between the $D_{t+T+1} - D_{t+T}$ and $r(t)$. Using an average fatality rate of $\overline{f} = 0.9\%$ for the UK as in Ferguson et al. (2020) (see discussion in Section 4.3.3), we estimate the reporting probability to be
$$\widehat{\pi(t)} = \frac{\overline{f} \quad r(t)}{D_{t+T+1} - D_{t+T}}, \tag{4.13}$$
which implies that the total number of cases in England is more than 20 times the reported number. As shown in Figure 4.9, prior to June 2020 this reporting ratio

[6] See `https://ourworldindata.org/coronavirus-testing`.

was around $\widehat{\pi(t)} = 4.5\%$; with the subsequent increase in testing, the estimated reporting ratio has steadily increased to more than 20% in November 2020.

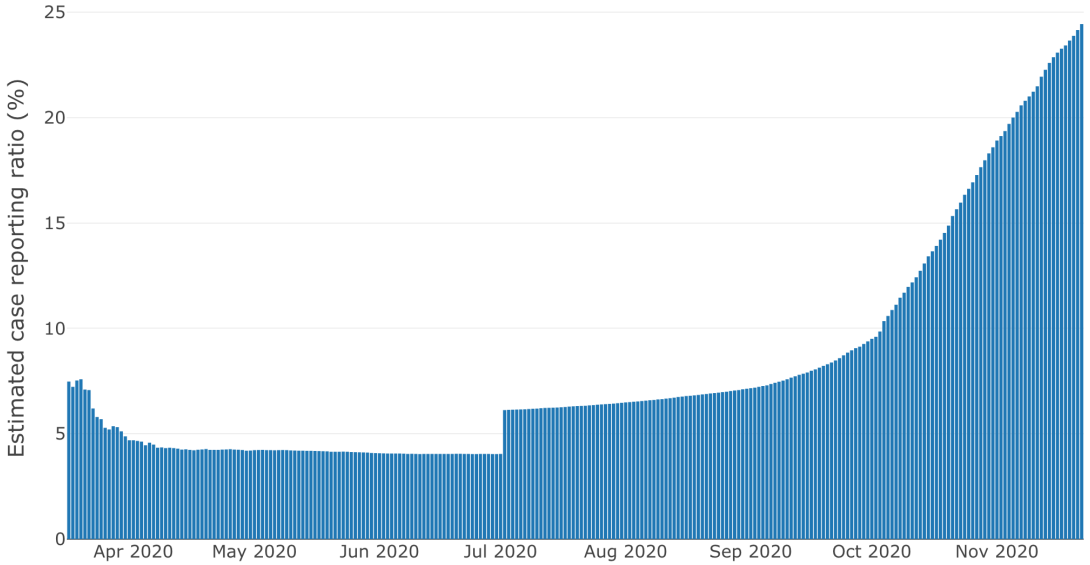


Figure 4.9: Estimate of case reporting ratio $\pi(t)$ based on a comparison of fatalities and reported cases.

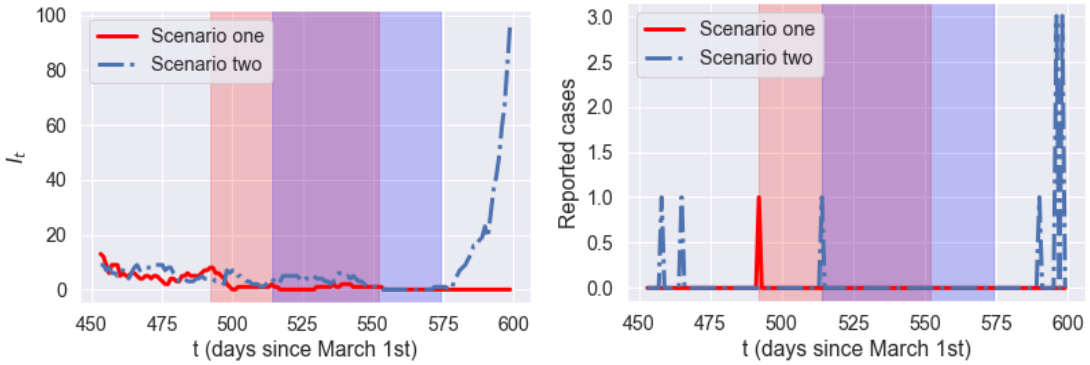
4.4.2 Implications of Partial Observability

A key issue in epidemic control is the availability of reliable indicators for the intensity of an ongoing epidemic. Public health authorities have communicated the daily number of reported cases and fatalities, and these have served as inputs for policy planning.

An important corollary is that, given the combination of random factors affecting dynamics and the considerable uncertainty on the actual number of new infections, it is perfectly possible to observe a run of many consecutive days without new reported cases while in fact the *actual* number of infections is on the rise.

Figure 4.10 shows an example of scenario in our model where for 60 consecutive days, although a small number of (symptomatic and asymptomatic) cases appear, none of them are reported due to the low detection probability ($\pi = 4.5\%$).

Nonetheless, after a run of 60 days without any reported cases (blue shaded area in Figure 4.10), which may prompt public health authorities to lower their guard, the epidemic takes off again. Figure 4.10 displays in fact two sample paths with the same initial conditions, which differ only through the randomness in the dynamics. The fact that the break-out occurs only in one of the two scenarios (in blue) illustrates how random flare-ups may originate from a small group of undetected cases.



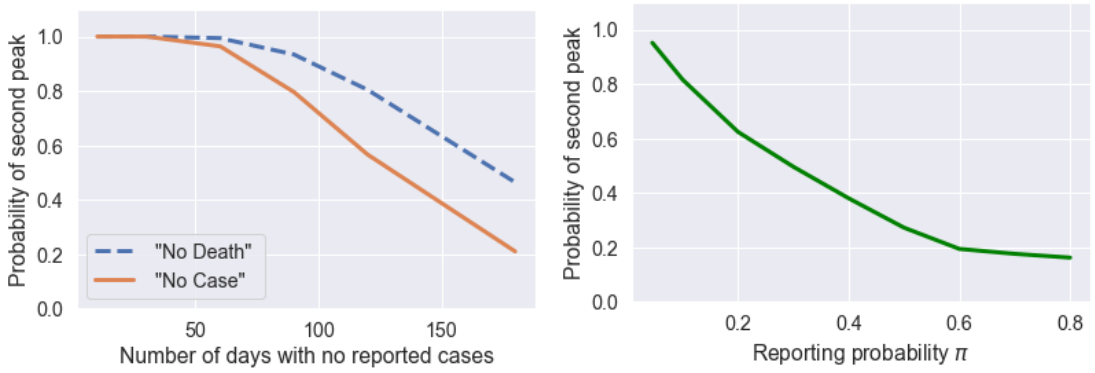
(a) Number of infected I_t.

(b) Reported cases.

Figure 4.10: Example of latent progression of the epidemic with zero reported case for 60 consecutive days (red shaded area for scenario one and blue shaded area for scenario two). Reporting probability $\pi = 4.5\%$.

Figure 4.11a shows the probability of observing a second peak in infections when social distancing measures are lifted after no reported cases for L consecutive days. This probability is estimated using 500 simulated paths from our model given in (4.3). It is striking to observe that, even after 60 days with no reported cases, the probability of observing a resurgence of the epidemic is around 40%. Figure 4.11a (blue dashed line) shows the same probability conditional on observing no fatalities for L consecutive days.

These observations point to the importance of broader testing: as shown in Figure 4.11b, an increase in the probability π of detecting new cases leads to a strong decrease in the probability of misdiagnosing the end of the epidemic, as in the scenario described above.



(a) Probability of having a second peak in infections after no reported cases (solid line) and no fatalities (dashed line) for L consecutive days (low symptomatic ratios).

(b) Probability of having a second peak in infections following 60 consecutive days with no reported cases, as a function of reporting probability π.

Figure 4.11: Probability of observing a second peak after a period with no cases reported.

4.5 Counterfactual Scenario: No Intervention

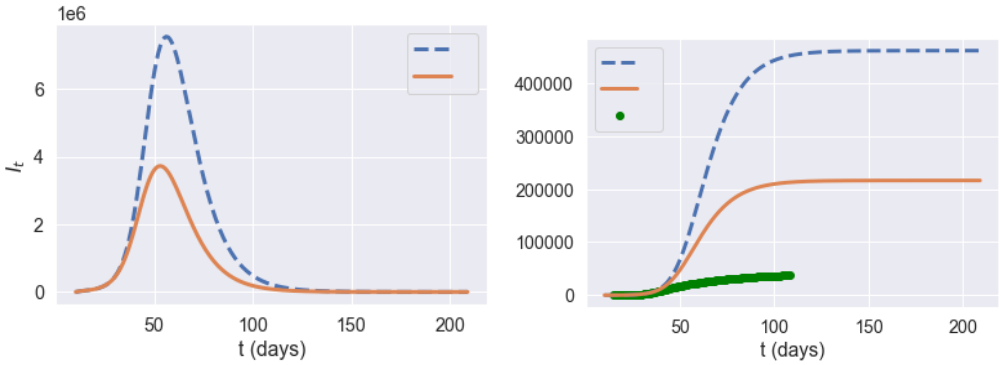
A much debated issue has been whether the lockdown and subsequent social distancing restrictions were necessary or whether health outcomes would have been comparable in absence of any restrictions, eventually leading to herd immunity. To examine this question we consider the counterfactual scenario of no intervention and estimate the fatalities and peak number of infections under such a scenario.

4.5.1 Magnitude and Heterogeneity of Outcomes

Our counterfactual simulations show that, in absence of social distancing and confinement measures, the number of fatalities in England may have exceeded 216,000 by August 1, 2020. This is 174,000 more than the outcome actually observed on this date following the lockdown.

Figure 4.12 displays the evolution of the number of symptomatic infections and fatalities in absence of restrictions, under two different assumptions on symptomatic ratios (see Table 4.3). Under the assumption of low symptomatic ratios, the total fatalities in absence of any mitigation policy are estimated to be 216,000 on average across 100 scenarios, with 12.8 million (more than 20% of England's

population) being infected and symptomatic. A peak number of 3,720,000 symptomatic individuals in England would have been reached on April 30, 2020.

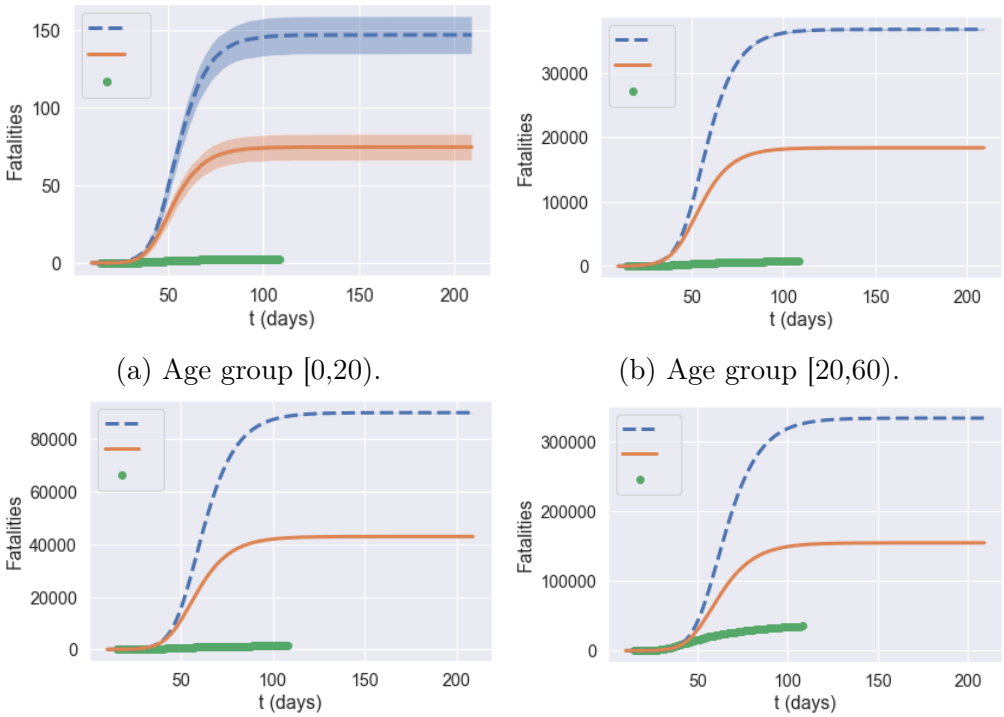


(a) Dynamics of I_t.

(b) Dynamics of D_t. Actual reported fatalities under lockdown policy in England shown as green dots.

Figure 4.12: Comparison of different quantities in England with no intervention: high symptomatic ratios (blue dashed line) against low symptomatic ratios (orange solid line), averaged across 100 simulated scenarios.

Figure 4.13 decomposes these results across different age groups.



(a) Age group [0,20).

(b) Age group [20,60).

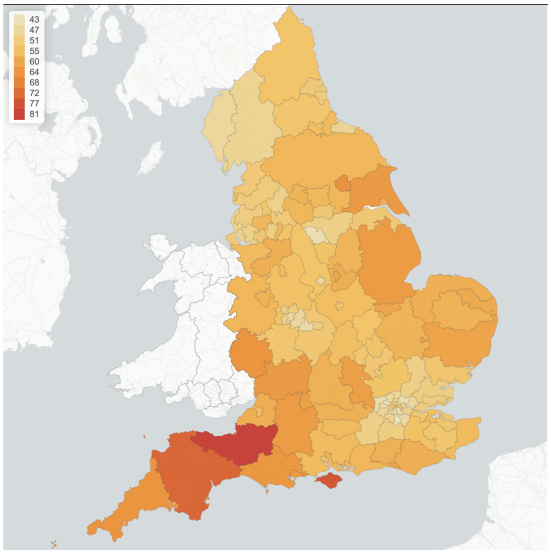
(c) Age group [60, 70).

(d) Age group [70, 100).

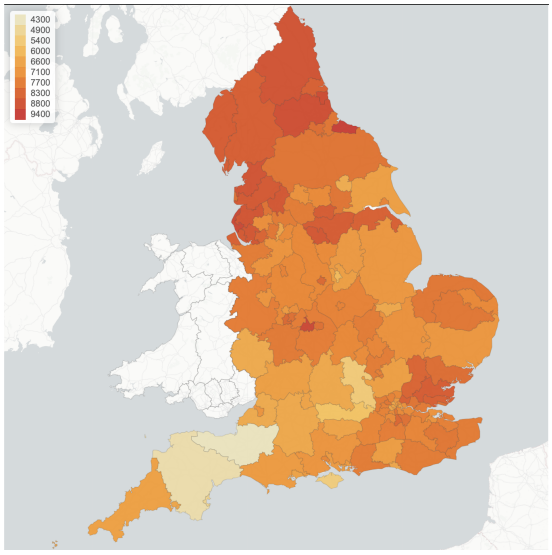
Figure 4.13: Fatalities by age group under no intervention, averaged across 100 scenarios: high symptomatic ratios (blue dashed line) against low symptomatic ratios (orange solid line). Actual reported fatalities under lockdown policy in England shown as green dots.

The low and high estimates for symptomatic ratios lead to very different simulation outcomes in terms of peak I_t values and total number of fatalities, which illustrates the huge impact of parameter uncertainty on model projections.

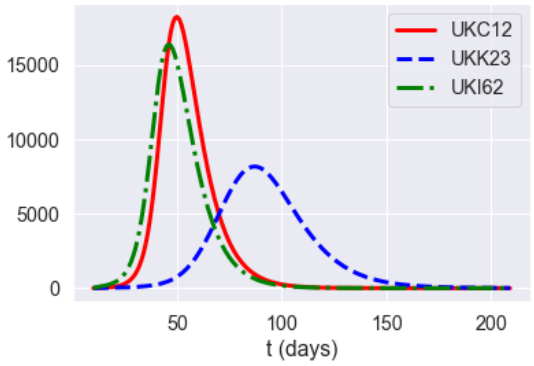
Heterogeneity of regional outcomes As shown in Figure 4.14, regions exhibit heterogeneous outcomes in terms of peak time, peak value, and fatalities (per 100,000 inhabitants).



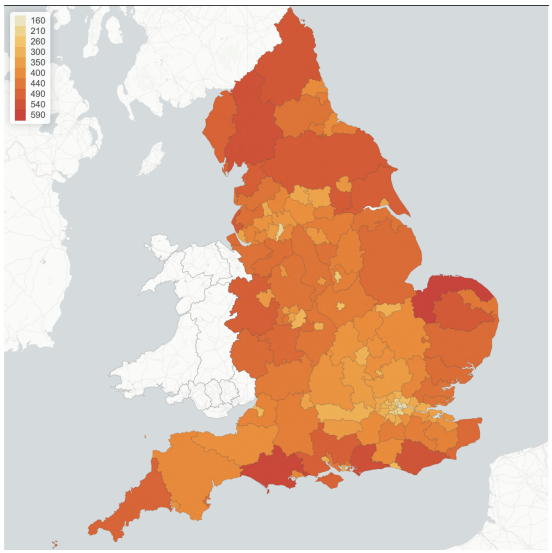
(a) Time of the infection peak (in days) across different regions.



(b) Peak value of infections per 100,000 inhabitants across different regions.



(c) Infections per 100,000 inhabitants for three selected regions: South Teesside (UKC12), Somerset (UKK23), Croydon (UKI62).



(d) Fatalities per 100,000 inhabitants.

Figure 4.14: Heterogeneity of outcomes across different regions in the absence of intervention. Outcomes are averaged over 100 simulated scenarios.

4.5.2 Impact of Demographic and Spatial Heterogeneity

Homogeneous SIR models (Aguilar et al., 2020; Lourenco et al., 2020; Donnat and Holmes, 2020; Kucharski et al., 2020; Pindyck, 2020; Roques et al., 2020b;

Rawson et al., 2020; Rowthorn and Maciejowski, 2020) or age-stratified versions of such models (Acemoglu et al., 2020; Chikina and Pegden, 2020; Davies et al., 2020; Prem et al., 2020; Singh and Adhikari, 2020) have been used in many recent studies on COVID-19 in the UK and other countries.

The heterogeneity of outcomes observed in our simulation suggests that homogeneous epidemic models may fail to capture some important features of COVID-19 dynamics which are relevant for public health policy. We investigate this point further by comparing our simulations with two homogeneous benchmarks: a country-level SEIAR compartmental model and an age-stratified version of it.

Homogeneous SEIAR model The homogeneous SEIAR model (**H**) corresponds to the case where the country is considered as a single region, assuming away geographic heterogeneity:

$$\begin{cases} \dot{S}_t = -\alpha\sigma \frac{A_t+\kappa I_t}{N} S_t, \\ \dot{E}_t = \alpha\sigma \frac{A_t+\kappa I_t}{N} S_t - \beta E_t, \\ \dot{A}_t = (1-p)\beta E_t - \gamma A_t, \qquad \dot{I}_t = p\beta E_t - \gamma I_t \\ \dot{D}_t = f\gamma I_t, \qquad \dot{R}_t = (1-f)\gamma I_t + \gamma A_t, \\ N = S_t + E_t + A_t + I_t + R_t + D_t. \end{cases} \tag{4.14}$$

where σ is the average contact rate in the population, and N is the total population.

We use the values of α, β and γ as specified in Table 4.5 and population-averaged versions of low symptomatic ratios in Table 4.3. We aggregate the age-stratified contact matrix (B.1) from POLYMOD (Mossong et al., 2017), the probability of developing symptoms, and fatality rate using the England population age distribution (Table 4.1). This leads to an aggregate fatality rate of 1.1%, a probability of developing symptoms of 45.4%, and a contact number of 4.157. As in Sections 4.3.3 and 4.3.5, we use an indirect inference method (Gourieroux et al., 1993) to estimate the implied scaling parameters for social contact rates by matching daily case dynamics before and during lockdown, leading to $d = 2.909$ and $l = 0.115$.

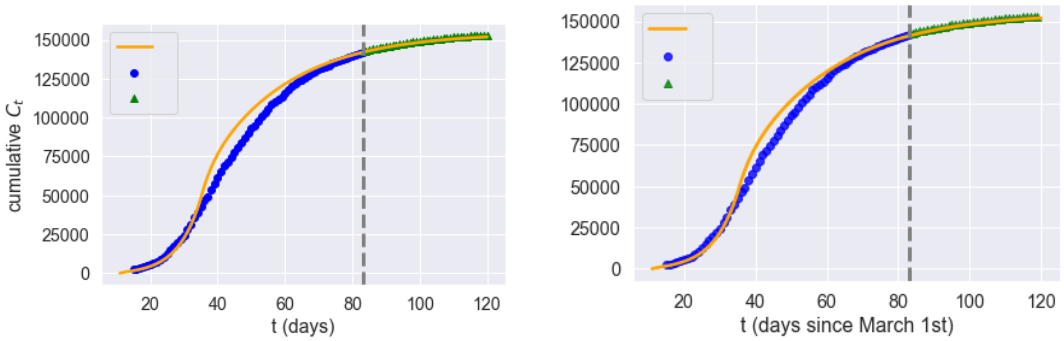
We also consider an age-stratified (**A**) version of this model with 4 age groups:

$[0, 20), [20, 60), [60, 70), [70, 100)$.[7] The rate of exposure for age group a is given by

$$\lambda_t(a) = \alpha \sum_{a'} \sigma(a') \frac{\kappa I_t(a') + A_t(a')}{N(a')}. \tag{4.15}$$

For the age-stratified model, we use α, β and γ as in Table 4.5. We estimate the scaling parameters for social contact rates before and during lockdown as above, yielding $d = 1.07$ and $l = 0.112$.

As shown in Figure 4.15, these country-level models actually yield reasonable fits to aggregate dynamics of cases and fatalities.



(a) Homogeneous model (**H**).

(b) Age-stratified model (**A**) with 4 age groups.

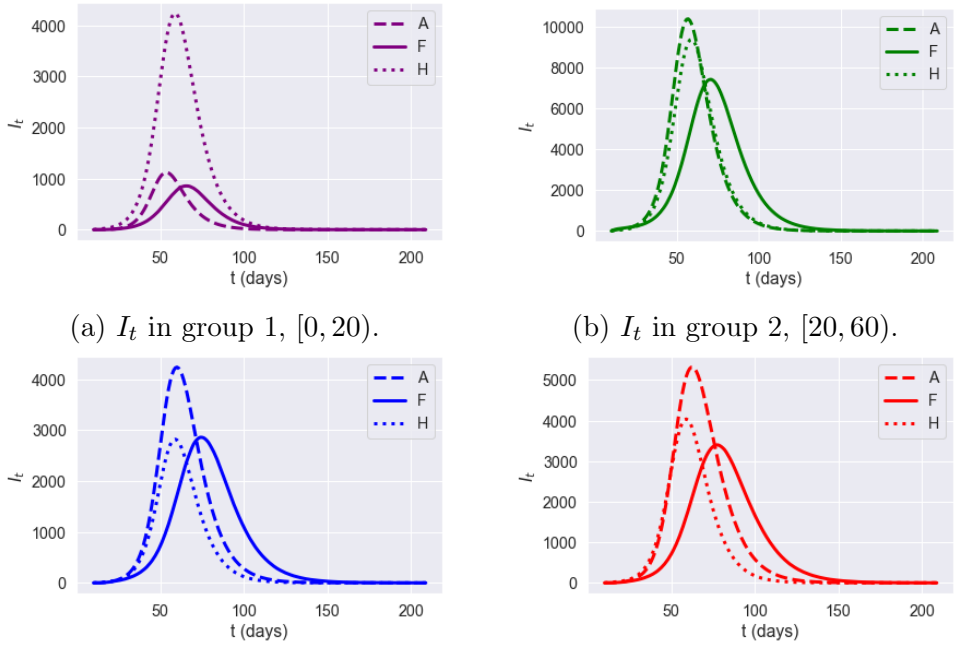
Figure 4.15: Goodness-of-fit for homogeneous and age-stratified country-level models. Grey dashed line: separation of estimation sample from test data; orange line: model; blue dots: data; and green triangles: out-of-sample data.

However, the inability of these country-level models to capture regional heterogeneity and inter-regional exchanges leads to regional outcomes that are very different from our model with spatial heterogeneity (**F**). Figures 4.16–4.19 compare simulation results for the homogeneous model (**H**), the age-stratified model (**A**) and our spatial heterogeneous model (**F**) in two regions: Torbay (UKK42) and Birmingham (UKG31). The homogeneous model is observed to overestimate fatalities in age groups 1 and 2 and underestimate fatalities in age groups 3 and 4. The age-stratified model and the heterogeneous model agree on fatalities across

[7] As shown in Section 4.7.1, the results are robust to changes in model granularity. In this section, we choose to work with a model with four age groups for ease of result presentation.

age groups but, as shown in Figures 4.16–4.19, neither the age-stratified model nor the homogeneous model can capture the regional features of epidemic dynamics, such as the difference in peak time and peak value of infections across regions.

We therefore conclude such country-level models should not be used in the context of policy discussions which target regional measures or regional outcomes.



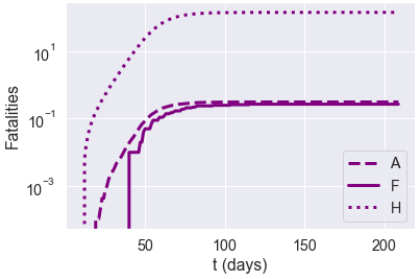
(a) I_t in group 1, $[0, 20)$.

(b) I_t in group 2, $[20, 60)$.

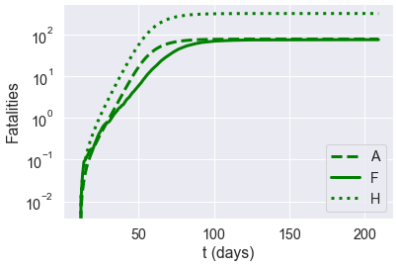
(c) I_t in group 3, $[60, 70)$.

(d) I_t in group 4, $[70, 100)$.

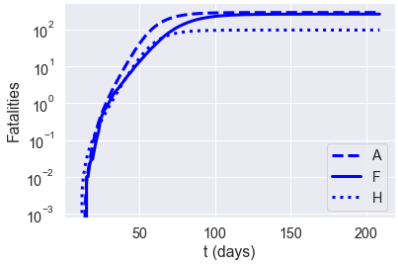
Figure 4.16: Dynamics of infections I_t in Torbay (UKK42): impact of model granularity.



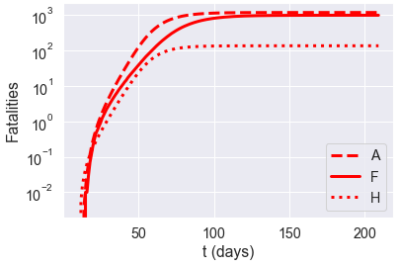
(a) Age group $[0, 20)$.



(b) Age group $[20, 60)$.

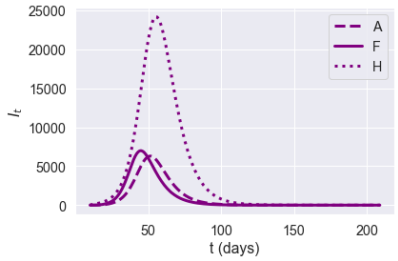


(c) Age group $[60, 70)$.

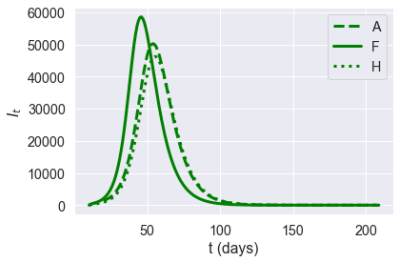


(d) Age group $[70, 100)$.

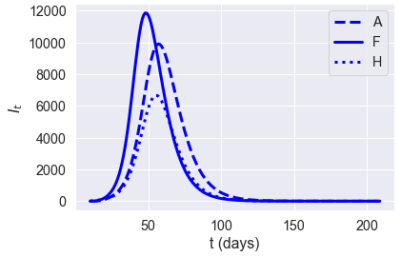
Figure 4.17: Dynamics of fatalities I_t in Torbay (UKK42): impact of model granularity.



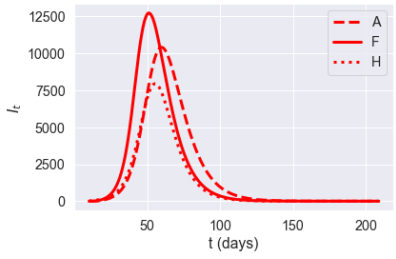
(a) Age group $[0, 20)$.



(b) Age group $[20, 60)$.

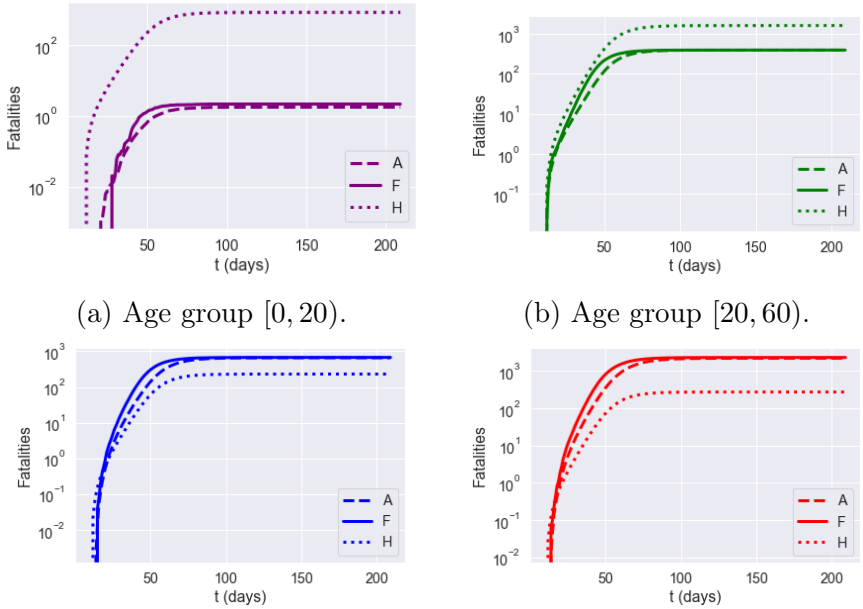


(c) Age group $[60, 70)$.



(d) Age group $[70, 100)$.

Figure 4.18: Dynamics of symptomatic infections (I_t) in Birmingham (UKG31): impact of model granularity.



(a) Age group $[0, 20)$. (b) Age group $[20, 60)$.

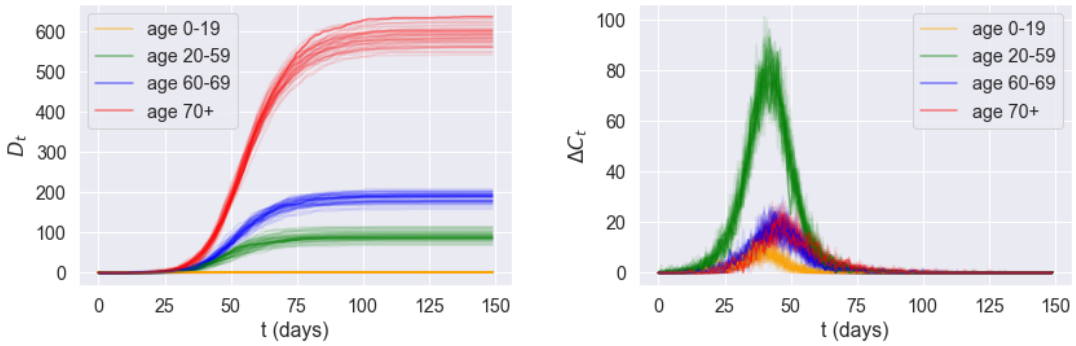
(c) Age group $[60, 70)$. (d) Age group $[70, 100)$.

Figure 4.19: Cumulative fatalities for Birmingham (UKG31): impact of model granularity.

4.5.3 Variability of Outcomes

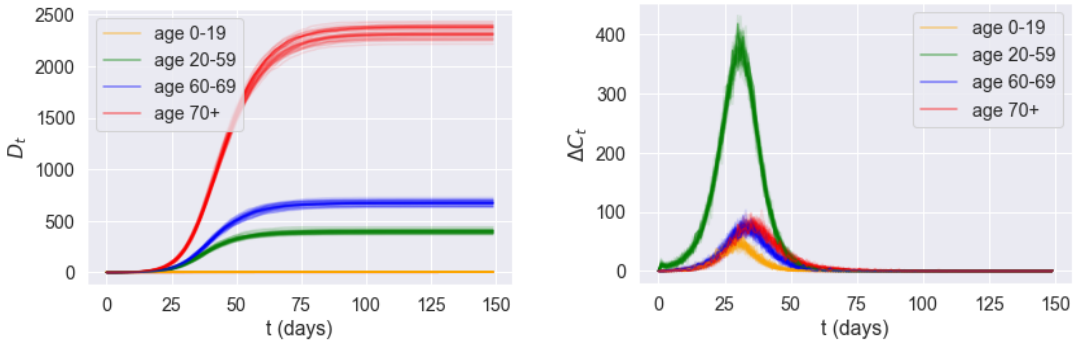
Given a set of initial conditions, the stochastic model (4.3) may lead to a range of outcomes due to the randomness present in the dynamics. Figures 4.20 and 4.21 show an example of variability of outcomes across different scenarios.

As expected from asymptotic analysis of large population limits (Britton et al., 2019), the relative variability across scenarios is of the order of $\sim 1/\sqrt{N(r,a)}$ for a subpopulation of size $N(r,a)$. Although not negligible, especially at the onset of the epidemic, this variability remains small at the regional level given the granularity used in our model. In the sequel we have thus reported the average outcomes across 100 or 1,000 simulated scenarios for each case examined.



(a) Fatalities D_t. (b) Reported daily cases ΔC_t.

Figure 4.20: Effect of randomness: variability of outcomes across 50 sample paths for Kingston upon Hull (UKE11). Results shown for low symptomatic ratios.



(a) Fatalities D_t. (b) Reported daily cases ΔC_t.

Figure 4.21: Effect of randomness: variability of outcomes across 50 sample paths for Birmingham (UKG31). Results shown for low symptomatic ratios.

4.6 Comparative Analysis of Epidemic Control Policies

We outline below an analysis of centrally pre-planned policies, focusing on the comparison of their social cost and the number of fatalities, and hence the identification of *efficiency frontiers*.

4.6.1 Confinement Followed by Social Distancing

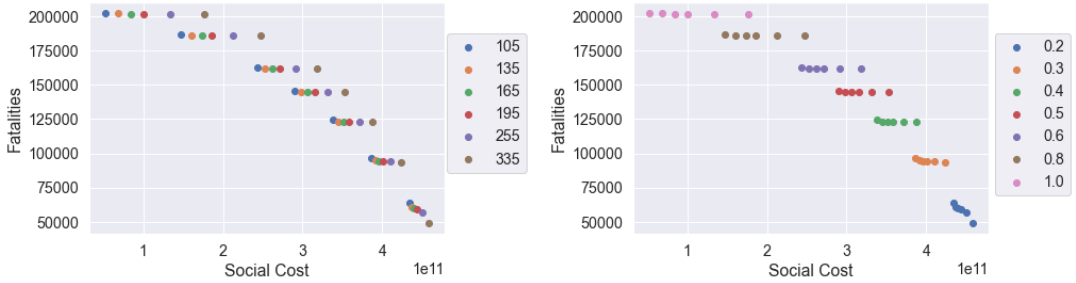
We first consider the impact of a national 'lockdown' followed by social distancing, which reflects the situation in the UK between March 2020 and August 2020. We examine in particular the impact of a lockdown duration T and the level of social distancing after lockdown on the number of fatalities and the associated social cost. To do so, we parameterise the contact matrix as

$$\sigma^r(t) = \begin{cases} l_r\ \sigma & \text{for} \quad t_0 \le t \le t_0 + T \quad \text{(lockdown)}, \\ ((1-m)l_r + m d_r)\ \sigma & \text{for} \quad t > t_0 + T \quad \text{(after lockdown)}, \end{cases} \tag{4.16}$$

where l_r measures the level of social distancing under lockdown, as estimated from observations for the period from March 23 to May 31, and the parameter $m \in [0, 1]$ measures the level of compliance with social distancing measures. A value of m close to zero indicates a level of social contact similar to lockdown, while $m = 1$ corresponds to normal levels of social contact.

The origin date $t = 0$ corresponds to March 1, 2020. All scenario simulations include a lockdown starting at $t_0 =$ March 23, 2020. We consider a range $105 \le T \le 335$ for the lockdown duration and $0.2 \le m \le 1$ for post-lockdown social distancing levels. Note that the actual duration of the first lockdown in England corresponded to $T = 105$.

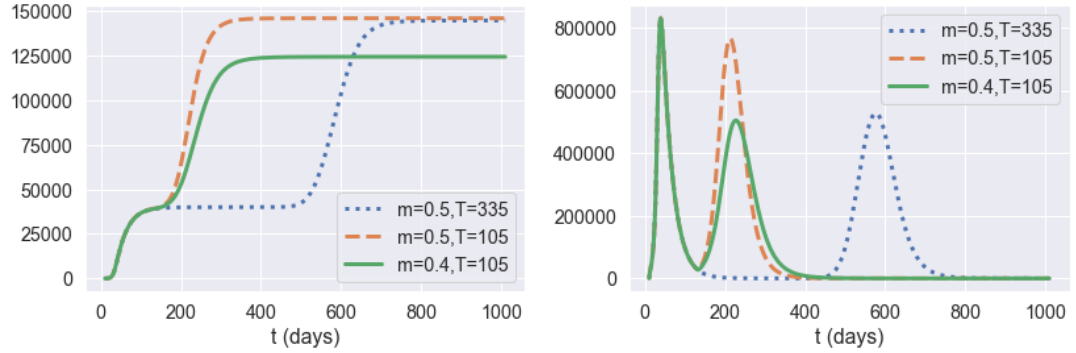
As shown in Figure 4.22a, the level of social distancing *after* the confinement period is observed to be more important (Figure 4.22b) than the length of the confinement period (Figure 4.22a). This is consistent with the findings in Lipton and Lopez de Prado (2020). Smaller values of m, associated with stricter social distancing, lead to lower fatalities but at an increased social cost (Figure 4.22b). On the other hand, the lengthening of the lockdown duration T, while significantly increasing the associated social cost, does not result in a significant reduction in the number of fatalities, especially if social distancing is not respected after lockdown.



(a) Impact of lockdown duration T. (b) Impact of compliance level m.

Figure 4.22: Fatalities against social cost for different T and m values. Results shown for low symptomatic ratios.

Figure 4.22 also shows that some of these policies are inefficient, in the sense that we can reduce fatalities *and* the social cost simultaneously by shortening the lockdown period or by relaxing social distancing constraints, as shown in Figure 4.23.



(a) Cumulative fatalities in England. (b) Dynamics of I_t in England.

Figure 4.23: Comparison of three policies averaged across 50 simulated scenarios. Blue dotted line: $m = 0.5$ and $T = 335$; orange dashed line: $m = 0.5$ and $T = 105$; green solid line: $m = 0.4$ and $T = 105$.

Policy	Blue dotted: $m = 0.5, T = 335$	Orange dash: $m = 0.5, T = 105$	Green solid: $m = 0.4, T = 105$
Social cost ($\times 10^{11}$)	3.5	2.9	3.4
Projected fatalities	144,600	146,000	124,400

Table 4.7: Outcomes for policies represented in Figure 4.23.

By comparing the orange and blue policies in Figure 4.23, which represent the same post-lockdown compliance level ($m = 0.5$), we observe that extending the lockdown duration increases social cost without reducing the total number of fatalities. On the other hand, comparing the orange and green policies, which correspond to the same lockdown duration of $T = 105$ days, shows that moving the compliance level from $m = 0.5$ to $m = 0.4$ reduces the second peak amplitude by 35% and fatalities by 13.9%.

Impact of parameter uncertainty The above results are highly sensitive to the value of the symptomatic ratios which, as noted in Section 4.3, are highly uncertain (see Table 4.3). Figure 4.24 shows the policy outcomes for low against high symptomatic ratios across different compliance levels and lockdown duration. As observed in this figure, while the overall pattern of the efficiency diagram is similar, the projected fatality levels shift considerably depending on the assumption on the symptomatic ratio: from 50,000–200,000 for low symptomatic ratios to 126,000–430,000 for high symptomatic ratios.

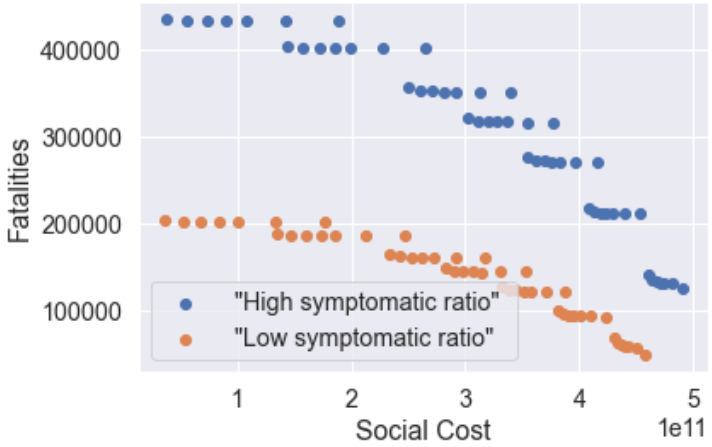


Figure 4.24: Trade-off between fatalities and social cost for a T-day lockdown followed by social distancing ($0.2 \leq m \leq 1$, $105 \leq T \leq 335$): low symptomatic ratio (orange) and high symptomatic ratio (blue).

Regional heterogeneity While the policies discussed here are applied uniformly across all regions, we observe a significant heterogeneity in mortality levels

across regions, as well in terms of the timing and amplitude of a second peak of infections. As shown in Figure 4.25, some regions exhibit mortality levels up to four times higher than others. This huge disparity in mortality rates cannot be explained by demographic differences alone, which are much less pronounced: the differences in social contact patterns seem to be more important, as illustrated in Figure 4.2.

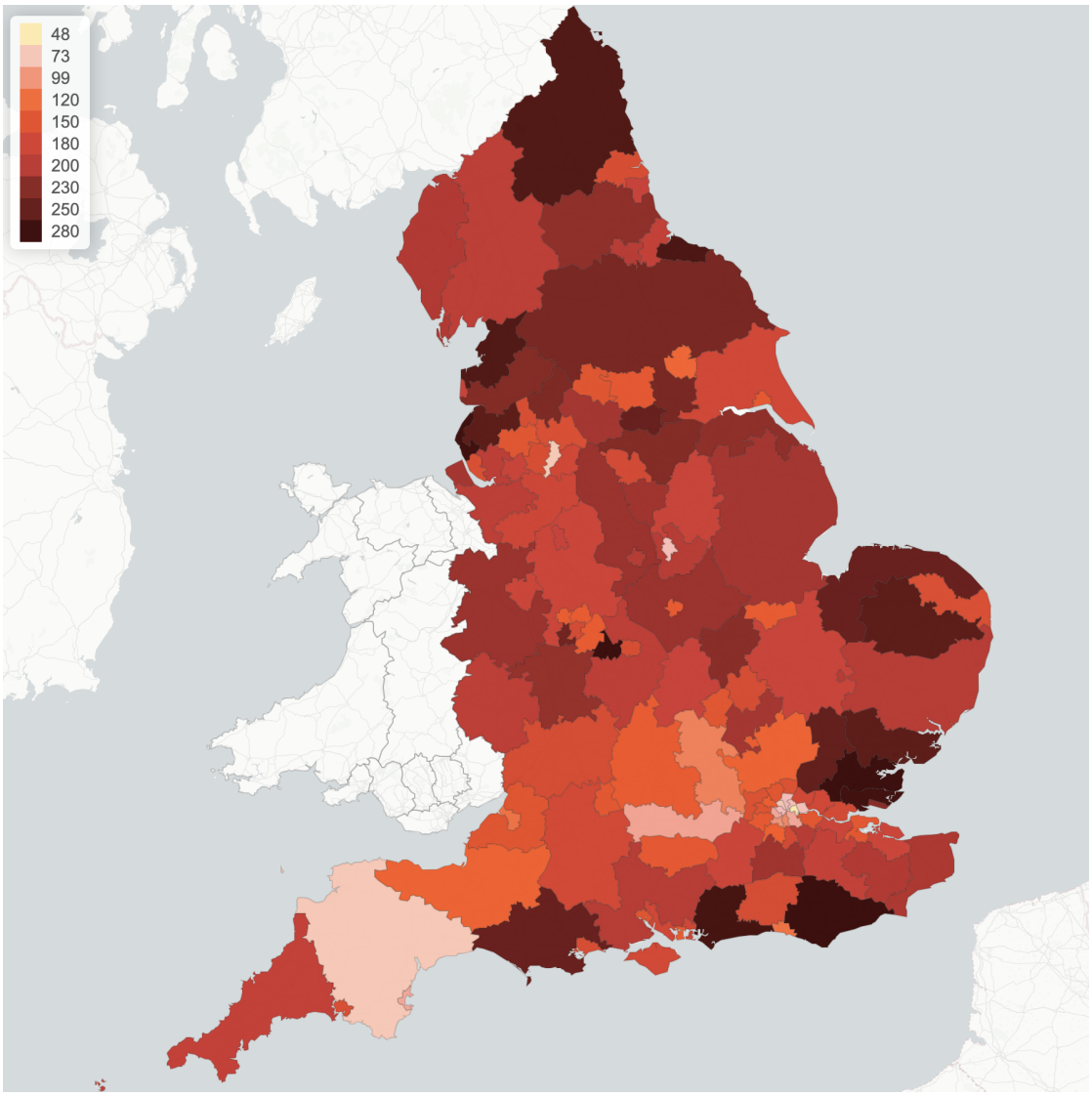
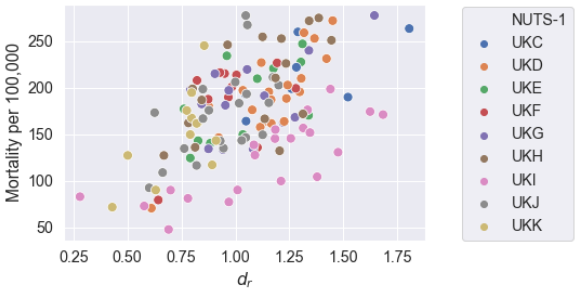
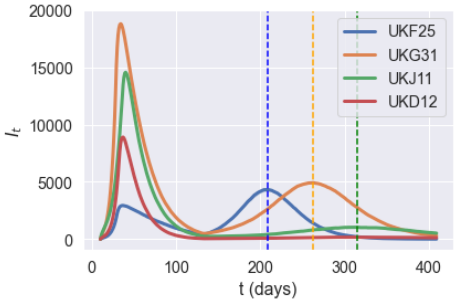


Figure 4.25: Regional mortality per 100,000 inhabitants with a lockdown of 105 days followed by social distancing ($m = 0.3$).

Indeed, as shown in Figure 4.26a, there is a positive correlation (above 40%) between regional COVID-19 mortality and the intensity of social contact as measured by the parameter d_r, defined in Section 4.3.3. Figure 4.26b shows that this heterogeneity is also reflected in the timing and amplitude of second peaks.



(a) Level of social contact (d_r) against COVID-19 mortality (per 100,000 inhabitants).



(b) Regional dynamics of symptomatic infections (I_t): North Northamptonshire (UKF25), Birmingham (UKG31), Berkshire (UKJ11) and East Cumbria (UKD12). Dotted lines denote the second peak times.

Figure 4.26: Regional outcomes for lockdown of 105 days followed by social distancing ($m = 0.3$).

4.6.2 Targeted Policies

We now consider the impact of social distancing measures targeting particular age groups or environments (school, work, etc.) following a lockdown of duration T, by setting

$$\sigma_{ij}^r(t) = \begin{cases} l_r\ \sigma & \text{for} \quad t_0 \le t \le t_0 + T; \\ \sigma_{ij}^{r,H} + u_{ij}^S \sigma_{ij}^{r,S} + u_{ij}^W \sigma_{ij}^{r,W} + u_{ij}^O \sigma_{ij}^{r,O} & \text{for} \quad t > t_0 + T. \end{cases} \tag{4.17}$$

We consider different targeted measures after a lockdown period of $T = 105$ days (the actual duration of the lockdown in England), including school closure, shielding of senior populations, workplace restrictions, restrictions on social gatherings and combinations thereof (see Figure 4.27). Note that there is no control over the social contact at home.

School closures Although most of the infected population in the school age groups are asymptomatic, they may in turn infect the senior population (those 60 or over) who are more likely to develop symptoms. School closure corresponds to $u^S = 0$, school reopening with social distancing correspond to $u^S = 0.5$, and school reopening without social distancing correspond to $u^S = 1$.

Shielding The high infection fatality rates among senior populations (that is, citizens over the age of 60) have naturally lead to the consideration of shielding policies for these populations. We model this as a reduction in social contacts of age groups 13 – 16 to the level observed under lockdown:

$$\sigma^r_{i,j}(t) = l_r \left(\sigma^H_{i,j} + \sigma^S_{i,j} + \sigma^W_{i,j} + \sigma^O_{i,j}\right) \quad \text{if } i \in \{13, \ldots, 16\} \text{ or } j \in \{13, \ldots, 16\}.$$

Workplace restrictions We model the impact of a partial return to work after confinement by assuming different proportions of workforce return after the lockdown period by choosing

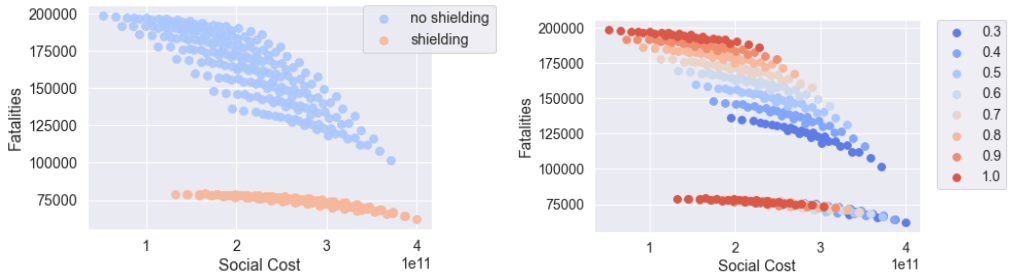
$$0.2 < u^W < 1 \quad \text{for} \quad t > t_0 + T, \tag{4.18}$$

the lower bound $u^W = 0.2$ corresponding to restricting workplace return to 'essential workers', as discussed in Section 4.3.5. Since workplace restrictions have an effect on commuting, such measures also have an impact on the inter-regional mobility matrix

$$M_t = u^W(t) M_0 + (1 - u^W(t))\mathbf{I}, \tag{4.19}$$

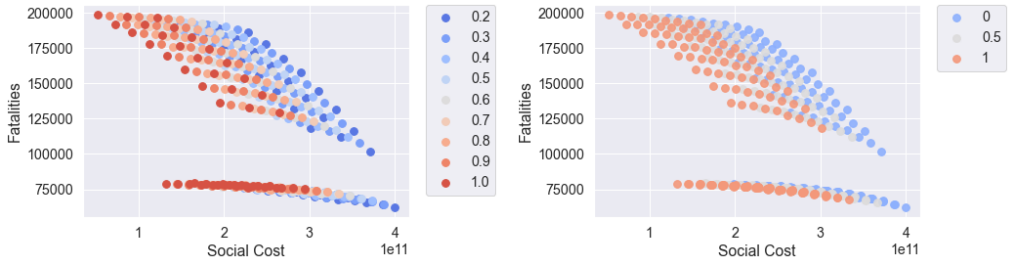
where $M_0(r, r')$ is the baseline mobility matrix defined in (4.7).

Restrictions on social gatherings Although social activities, such as gatherings at pubs or sports events, may aggravate the contagion of COVID-19, keeping certain levels of social activities is important to the economic recovery and the well-being of individuals. The parameter u^O measures the fraction of social gatherings: during the lockdown this fraction was estimated to be as low as 20% (see Section 4.3.5). In what follows, we consider $u^O \in [0.3, 1.0]$ after the period of lockdown.



(a) Impact of the shielding measure for senior citizens.

(b) Social distancing outside work and school: impact of the parameter u^O.



(c) Social distancing at work: impact of the parameter u^W.

(d) Social distancing at school: impact of the parameter u^S.

Figure 4.27: Efficiency plot of social cost against projected fatalities for the shielding measure and various values of u^S, u^W, and u^O ($u^H = 1$ and $T = 105$).

4.6.2.1 Pubs and Schools

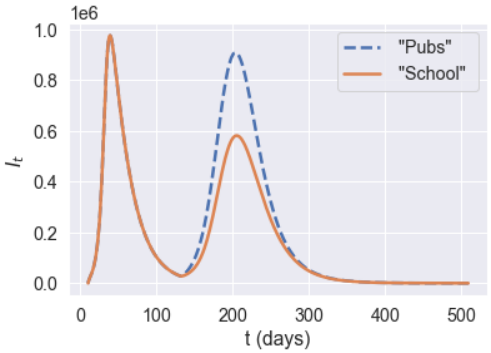
Table 4.8 shows the impact of school closures and social distancing at schools on projected fatalities and social contacts. Reopening schools, while reducing significantly the social cost, does not seem to lead to a significant increase in fatalities.

Policy	School closure: $u^S = 0$	Social distancing at school: $u^S = 0.5$	Normal school regime: $u^S = 1.0$
Social cost ($\times 10^{11}$)	2.2	1.9	1.5
Projected fatalities	153,900	157,000	159,300

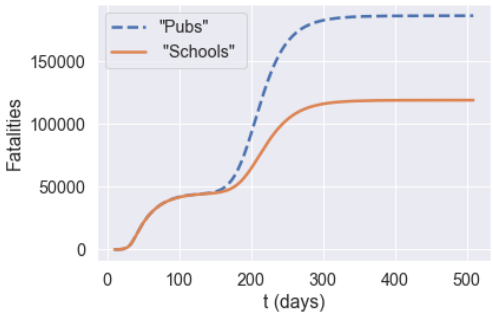
Table 4.8: Impact of school closures and social distancing at schools: outcomes averaged across 50 simulated scenarios, $u^H = u^W = 1$, $u^O = 0.5$.

We compare two post-confinement policies, one (labeled as 'schools') involves

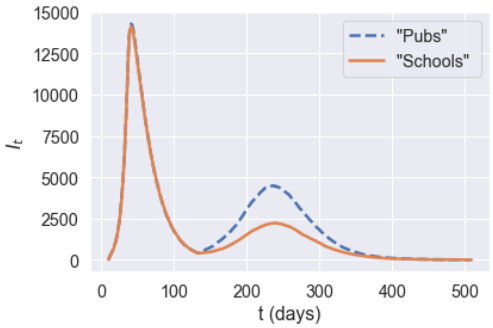
leaving schools open while social gatherings are restricted ($u^S = 1, u^O = 0.2$), and the other (labeled as 'pubs') involves closing schools while not restricting social gatherings ($u^S = 0, u^O = 1$). The social cost for the 'pubs' policy is 2.3×10^{11}, while the cost for the 'schools' policy is 3.0×10^{11}. However, as shown in Figure 4.28, the 'schools' policy leads to 35% fewer fatalities compared to the 'pubs' policy.



(a) Infections in England.



(b) Fatalities in England.



(c) Infections in Oxfordshire (UKJ14).

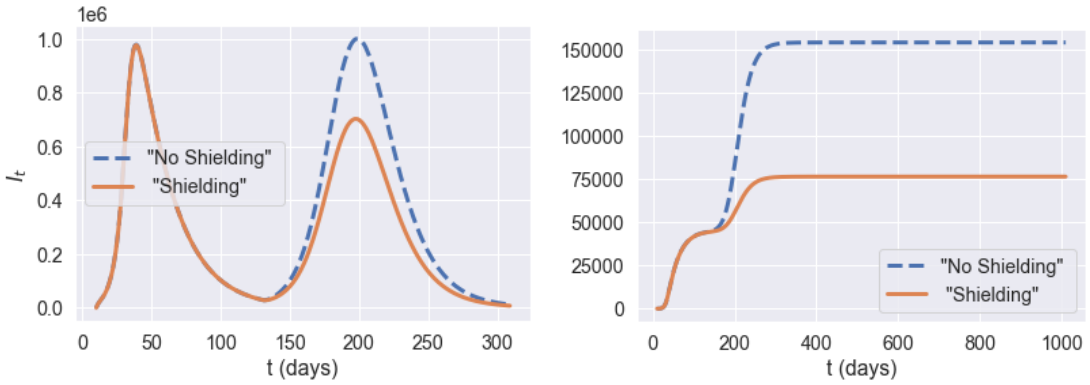
Figure 4.28: Comparison of post-confinement policies. Blue dashed line: restrictions in schools but not social gatherings (policy 'pubs'); orange solid line: restrictions in social gatherings but not in schools (policy 'schools').

4.6.2.2 Shielding of Senior Citizens

We have examined the impact of shielding in isolation and also in combination with other measures such as school closure and social distancing.

As shown in Figure 4.27a, whether applied in isolation or in combination with other measures, shielding of senior populations is by far the most effective measure for reducing the number of fatalities. As clearly shown in Figure 4.27a, regardless

of the trade-off between social cost and health outcome, a policy which neglects the shielding of seniors is not efficient and its outcomes can always be improved through shielding measures.



(a) Symptomatic infections in England.

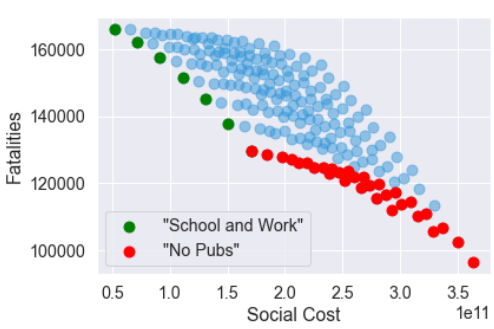
(b) Fatalities in England.

Figure 4.29: Comparison of policies with and without shielding in place, when $u = (1, 0.0, 1.0, 0.5)$. Blue dashed line: no shielding; orange solid line: shielding in place.

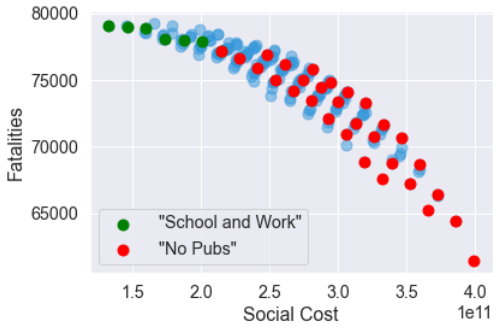
Impact of shielding on policy efficiency For policies without shielding, the level of social gatherings, u^O, is the leading factor to determine the efficiency frontier. In Figures 4.30a, the efficiency frontier contains two classes of policies:

- 'School and Work' policies: do not include any restrictions on school or work ($u^S = 1, u^W = 1$) but have a varying level of restrictions on social gatherings ($0.3 \leq u^O \leq 1$). Within this class of policies, different level of social gatherings lead to very different outcome of fatalities, as illustrated in Figure 4.30a; and
- 'No Pubs' policies: social gatherings outside school and work are restricted ($u^O = 0.3$), with different levels of social distancing $u^S \in \{0, 0.5, 1\}$ and $u^W \in [0.2, 1]$ at school and work.

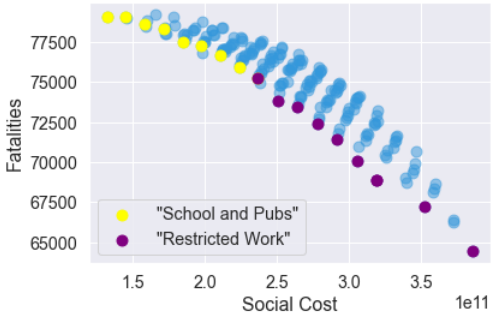
However, as observed in Figure 4.30b, these policies are not efficient when shielding measures are put in place for seniors.



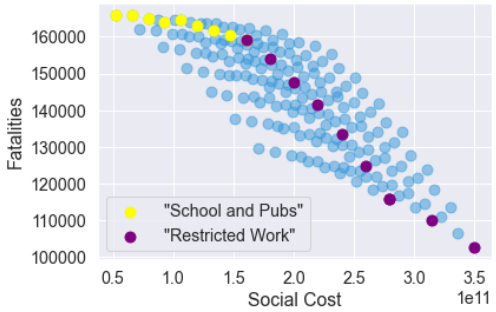
(a) Efficient policies without shielding: 'School and Work' and 'No Pubs'.



(b) Policies 'School and Work' and 'No Pubs' are not efficient when shielding is applied.



(c) Efficient policies with shielding: 'School and Work' and 'No work'.



(d) Policies 'School and Pubs' and 'No work' are not efficient when shielding is removed.

Figure 4.30: Impact of shielding on the efficiency frontier.

Under shielding, the spectrum of efficient policies is parameterised by the fraction u^W of the workforce returning to work. As shown in Figure 4.30c, we can distinguish two classes of efficient policies under shielding:

- 'School and Pubs', consisting of policies without restrictions on schools or social gatherings ($u^S = 1, u^O = 1$) and different levels u^W of restrictions on workplace gatherings.

- 'Restricted Work' policies, under which only 'essential' workers are allowed on-site work ($u^W = 0.2$), with either

  (i) No school restrictions ($u^S = 1$) and different levels of restrictions on social gatherings ($0.2 \leq u^O \leq 1$); or

(ii) Restrictions on social gatherings ($u^O = 0.3$, that is 'no pubs') and different levels of social distancing in school ($0 \leq u^S \leq 1$).

As Figure 4.30d illustrates, 'School and Pubs' and 'Restricted Work' policies are not efficient without shielding.

In absence of shielding, social gatherings seem to be the main vector for contagion. When shielding measures are put in place, the social contacts associated with seniors are reduced to the same level as under lockdown; in this case, contacts at work become the main vector of contagion.

4.7 Adaptive Mitigation Policies

We now consider *adaptive* mitigation policies, in which the daily number of (national or regional) reported cases is used as a trigger for social distancing measures. Such policies have been recently implemented, in the UK and elsewhere, at a local or national level with various degrees of success. We distinguish *centralised* policies, based on monitoring of national case numbers, from *decentralised* policies where monitoring and implementation of measures are done at the level of NUTS-3 regions.

4.7.1 Country-Wide Restrictions

We first consider centralised policies that monitor the number of daily reported cases at country level. Whenever the number of daily reported cases (per 100,000 inhabitants) exceeds a threshold B_{on}, confinement measures are imposed for a minimum of L days, until the number of daily reported cases falls below the threshold $B_{\text{off}} < B_{\text{on}}$. Outside these lockdown periods, we assume social distancing is in place with a compliance level m; we use a default value of $m = 0.5$.

This policy is implemented after the initial lockdown (that is, after July 4, 2020). In terms of the social contact matrix, we have, for $t > t_0 + T$,

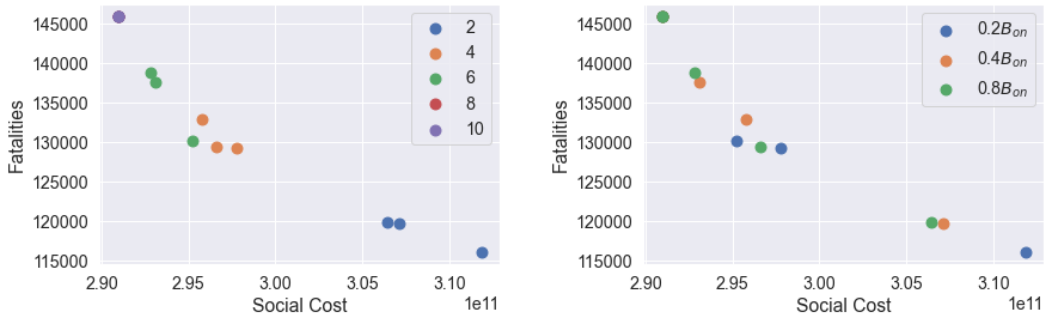
$$\begin{cases} \sigma^r(t) = ((1-m)l_r + md_r)\,\sigma;\ i_s = 0, & \text{if } (C_t \leq \frac{NB_{\text{off}}}{100{,}000} \text{ and } \Pi_{s=t-L}^{t-1} i_s = 1); \\ & \text{or } (C_t \leq \frac{NB_{\text{on}}}{100{,}000} \text{ and } i_{t-1} = 0); \\ \sigma^r(t) = l_r \times \sigma;\ i_t = 1, & \text{if } C_t > \frac{NB_{\text{on}}}{100{,}000} \text{ or } \Pi_{s=t-L}^{t-1} i_s = 1. \end{cases} \tag{4.20}$$

Here $T = 105$, i_t is the indicator of whether lockdown is applied on day t and C_t is the daily reported cases in England on day t. $\Pi_{s=t-L}^{t-1} i_s = 1$ if lockdown has been applied for L consecutive days during the period $[t - L, t - 1]$.

We simulate the dynamics with various choices of B_{off} and B_{on}:

- $B_{\text{on}} \in \{2, 4, 6, 8, 10\}$ (daily reported cases per 100,000 inhabitants); and
- $B_{\text{off}} = 0.2\ B_{\text{on}}$, $B_{\text{off}} = 0.4\ B_{\text{on}}$ or $B_{\text{off}} = 0.8\ B_{\text{on}}$.

We assume that once a lockdown is triggered it lasts a minimum of $L = 7$ days and that, once lockdown is removed, individuals continue to observe social distancing as measured by the parameter $m \in [0, 1]$. Data on real-time mobility monitoring in the UK indicate mobility to be at 50% of normal level during the post-lockdown period.[8] We therefore use $m = 0.5$ as a default value.



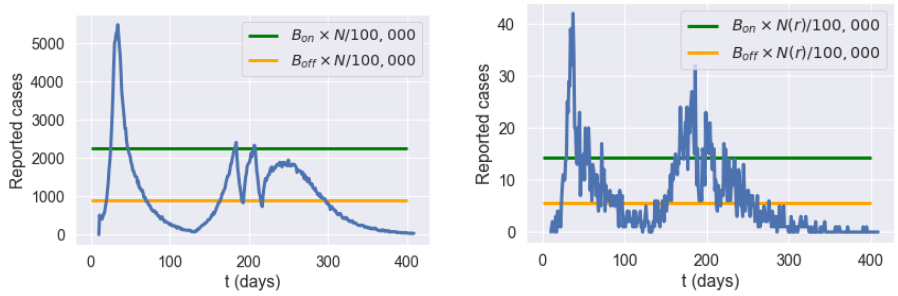
(a) Influence of the threshold B_{on} to resume lockdown.

(b) Influence of the threshold B_{off} to lift lockdown.

Figure 4.31: Social cost against fatalities when $m = 0.5$.

Example Figure 4.32 shows an example of such an adaptive policy, where lockdown is triggered when daily cases exceeds 2,240 nationally, and maintained until the count of new daily cases drops to 896. In the scenario shown in Figure 4.32a, this results in two short lockdowns, totaling 19 days. Consequently, this brings under control the national progression of the epidemic and avoids a second peak at the national level. However, as shown in Figure 4.32b, this policy is less successful at the regional level, resulting in a regional outbreak in Leicester.

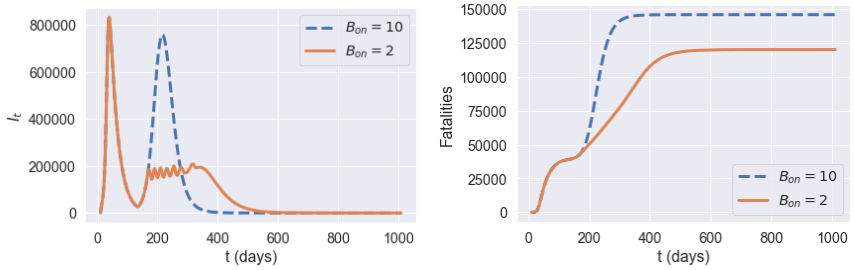
[8] See `https://www.oxford-covid-19.com/`.



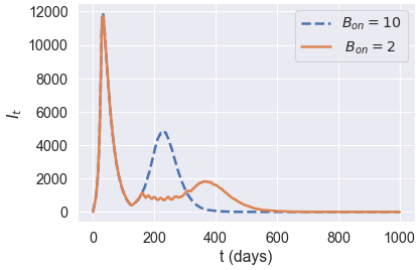
(a) Daily reported cases in England. (b) Daily reported cases in Leicester.

Figure 4.32: Simulation of reported cases in England and Leicester under a centralised triggering policy with $B_{\text{on}} = 4$, $B_{\text{off}} = 0.4 \times B_{on}$, $m = 0.5$ and no shielding.

Impact of the triggering threshold B_{on} The trigger threshold B_{on} has a significant impact on the efficiency of the policy. Smaller B_{on} values correspond to more frequent lockdowns, leading to a larger social cost and fewer fatalities. Here, we compare the impact of the triggering threshold B_{on} when $m = 0.5$ and $B_{\text{off}} = 0.4 \times B_{\text{on}}$ (see Figure 4.33.



(a) Dynamics of I_t in England. (b) Cumulative fatalities in England.



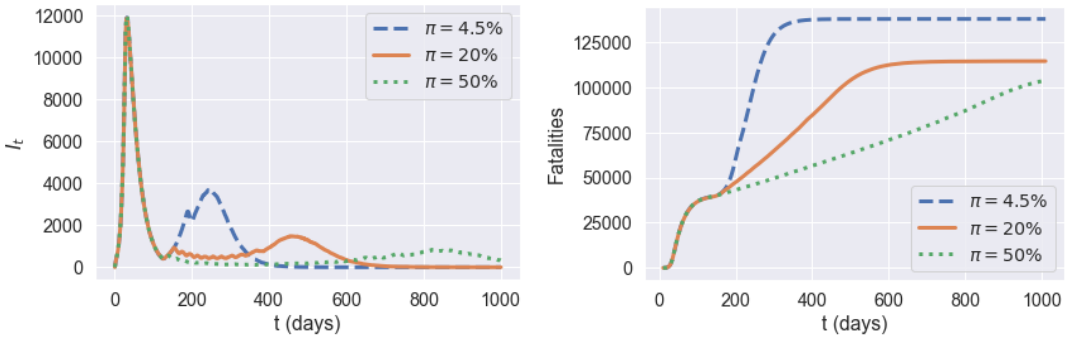
(c) I_t in Oxfordshire (UKJ14).

Figure 4.33: Comparison between triggering thresholds $B_{\text{on}} = 10$ and $B_{\text{on}} = 2$.

We observe in our simulations a second peak in I_t for England when $B_{\text{on}} = 10$, while we observe no second peak when $B_{\text{on}} = 2$. When $B_{\text{on}} = 2$, I_t remains at level 2×10^5 with frequent interventions for 200 days and then decreases to zero. The social cost for policy $B_{\text{on}} = 10$ and policy $B_{\text{on}} = 2$ are 2.9×10^{11} and 3.1×10^{11}, respectively. Policy $B_{\text{on}} = 2$ has 18% fewer fatalities compared to policy $B_{\text{on}} = 10$. Oxfordshire exhibits the same profile as England when $B_{\text{on}} = 10$. However, the shape of I_t is different for $B_{\text{on}} = 2$ where Oxfordshire experiences a small outbreak around day 350.

In summary, smaller B_{on} values correspond to more frequent lockdowns and result in damping or elimination of the second peak.

Increasing testing capacity To study the effect of an increased testing capacity, we assume wider testing is adopted such that the reporting probability is increased from 4.5% to a significantly higher level (20%, 50%) on July 4, 2020. Results are summarised in Figure 4.34 and Table 4.9.



(a) Dynamics of I_t in England.

(b) Cumulative fatalities in England.

Figure 4.34: Reporting probabilities $\pi = 4.5\%$ (blue line) versus $\pi = 20\%$ (orange line) and $\pi = 50\%$ (green line). Policy: $B_{\text{on}} = 6$, $B_{\text{off}} = 0.2B_{\text{on}}$, and $m = 0.5$.

Reporting probability π	4.5%	20%	50%
Social Cost ($\times 10^{11}$)	2.9	3.2	3.6
Fatalities	137,600	114,300	103,700

Table 4.9: Average social cost and fatalities for a given policy with different testing capacities (50 paths). Policy: $B_{\text{on}} = 6$, $B_{\text{off}} = 0.4B_{\text{on}}$, and $m = 0.5$.

By increasing the testing capacity, the observable quantity of daily reported cases becomes more consistent with the underlying dynamics of I_t. Compared to the policy with a reporting probability $\pi = 4.5\%$ throughout the reference period, we see that the dynamics of I_t when $\pi = 50\%$ decrease to a small value rapidly. Increasing the testing capacity also implies a more efficient control and as a result leads to fewer fatalities.

Impact of demographic granularity Several studies on the impact of public health policies on COVID-19 dynamics have used less granular models with fewer age groups (Acemoglu et al., 2020). To assess whether such coarse-graining may result in a loss of accuracy for the model projections, we have compared our present model, which has 16 age groups, with coarse-grained versions of the model in which all individuals in the 20–59 age range are grouped into 2 age groups (leading to a total of 5 age groups) or a single group (leading to 4 age groups). Parameters for the coarse-grained models are obtained as population-weighted averages of parameters from the granular model.

Comparison of model projections, shown in Figure 4.35, indicate that the results are robust to changes in model granularity. Some quantitative differences may emerge when assessing the impact of targeted policies, but the overall dynamics of infections, cases and fatalities are rather insensitive to the demographic granularity.

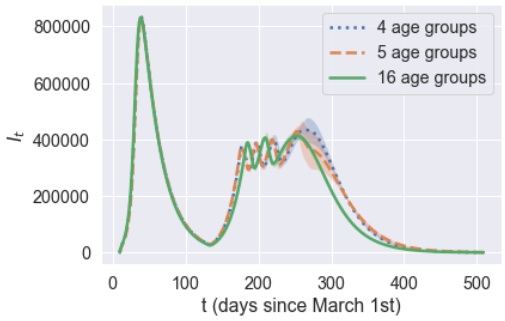
4.7.2 Decentralised Policies

We now consider a decentralised version of the above policies, based on the regional monitoring of cases as triggers for regional confinement measures.

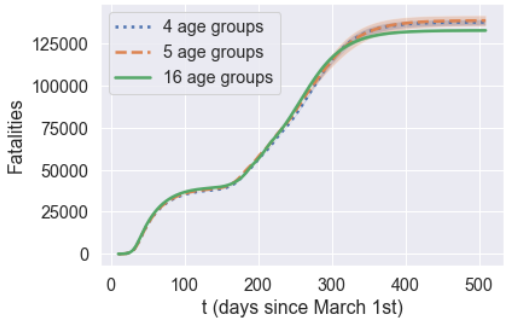
In terms of the social contact matrices, we have, for $t > t_0 + T$,

$$\begin{cases} \sigma^r(t) = ((1-m)l_r + md_r)\,\sigma;\ i_t^r = 0, & \text{if } (C_t(r) \le \frac{N(r)\,B_{\text{off}}}{100000} \text{ and } \Pi_{s=t-L}^{t-1} i_s^r = 1) \\ & \text{or } (C_t(r) \le \frac{N(r)\,B_{\text{on}}}{100000} \text{ and } i_{t-1}^r = 0) \\ \sigma^r(t) = l_r \times \sigma;\ i_t^r = 1, & \text{if } C_t(r) > \frac{N(r)\,B_{\text{on}}}{100000} \text{ or } \Pi_{s=t-L}^{t-1} i_s^r = 1. \end{cases} \quad (4.21)$$

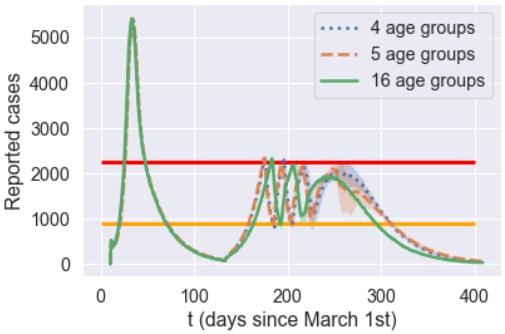
Here, i_t^r is the indicator of whether lockdown is applied in region r on day t and $C_t(r)$ is the daily number of cases reported in region r on day t. The term $\Pi_{s=t-L}^{t-1} i_s^r$ is used to track if lockdown has been applied in region r for L consecutive days



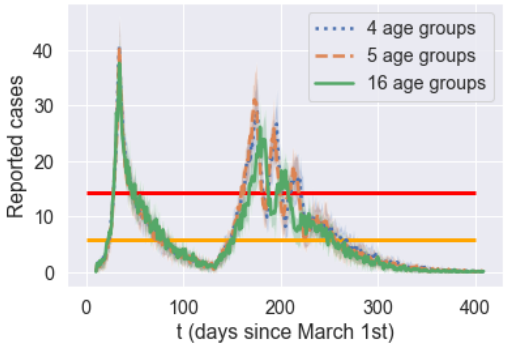
(a) Symptomatic infections (I_t): England.



(b) Projected fatalities: England.



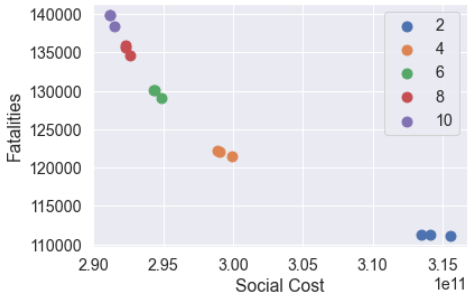
(c) Reported cases projection: England.



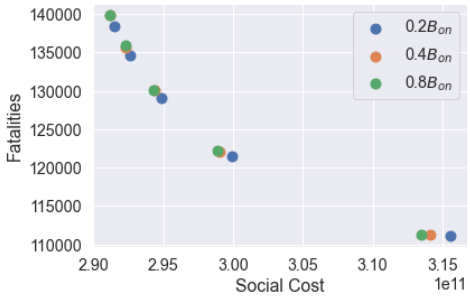
(d) Reported cases projection: Leicester.

Figure 4.35: Impact of model granularity: projections for an adaptive policy with $R_{\text{on}} = 4$, $R_{\text{off}} = 0.4 \times R_{on}$, $m = 0.5$ and no shielding.

during $[t - L, t - 1]$. We use the same values of B_{on} and B_{off} as in Section 4.7.1. Results are shown in Figure 4.36.



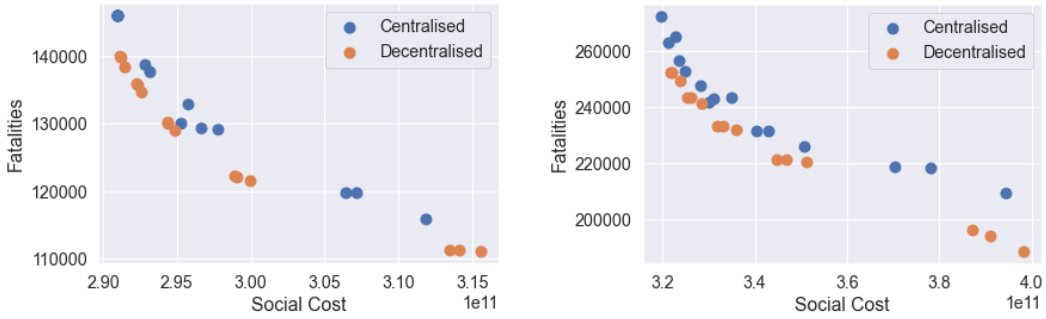
(a) Influence of the threshold B_{on} for triggering lockdown.



(b) Influence of the threshold B_{off} for lifting lockdown.

Figure 4.36: Decentralised confinement triggered by regional daily case numbers: social cost against fatalities ($m = 0.5$).

Figure 4.37 compares the outcomes of centralised and decentralised triggering policies. Decentralised policies are observed to always improve over centralised policies. For example, when $B_{\text{on}} = 4$ and $B_{\text{off}} = 0.4B_{\text{on}}$ there are 133,000 fatalities in England under the centralised policy, compared with 122,000 fatalities under the decentralised policy – that is an 8% reduction.

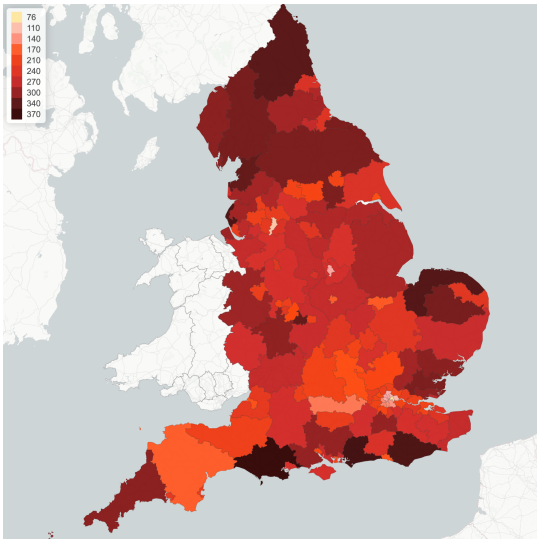


(a) Low symptomatic ratios.

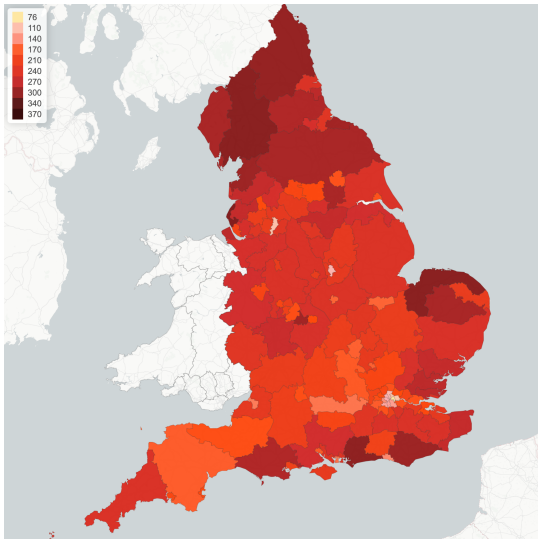
(b) High symptomatic ratios.

Figure 4.37: Efficiency analysis for centralised (blue) and decentralised (orange) adaptive mitigation policies. Outcomes averaged across 100 simulated scenarios.

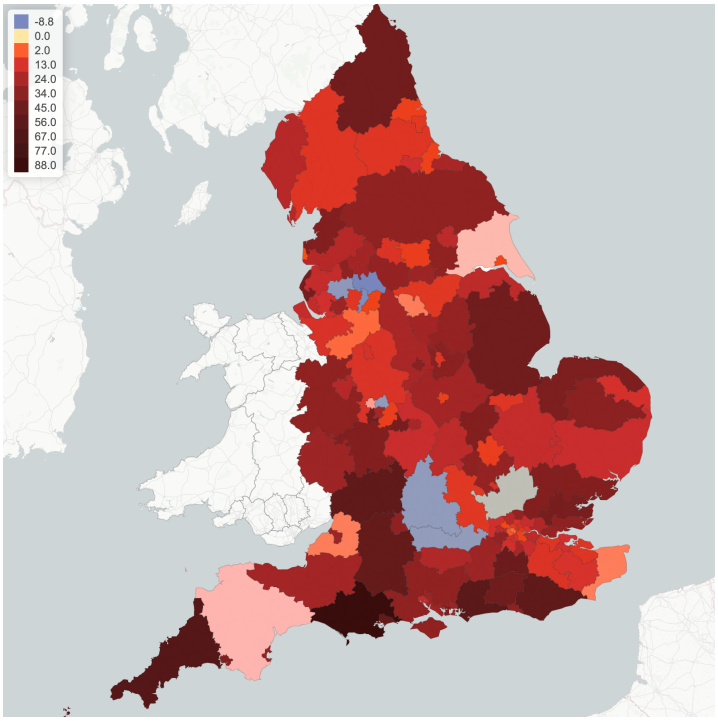
Figure 4.38 compares regional fatalities per 100,000 habitants for these policies. For more than 90% of the regions, decentralised measures lead to fewer fatalities. The most effective reductions are in Dorset, South West England (UKK22) with 23% fewer fatalities, and in Cornwall and Isles of Scilly (UKK30) with 21% fewer fatalities. However, there are a few exceptions (see regions in light blue in Figure 4.38c). These regions are already under control before adaptive policies are applied. Therefore, the improvement of moving from centralised policy to decentralised policy is limited.



(a) Centralised (country-level) adaptive policy.



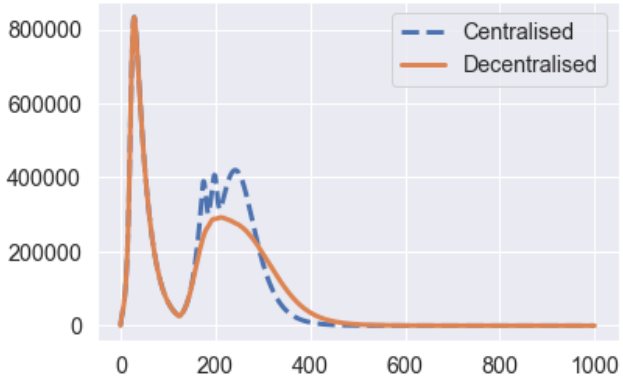
(b) Decentralised (regional) adaptive policy.



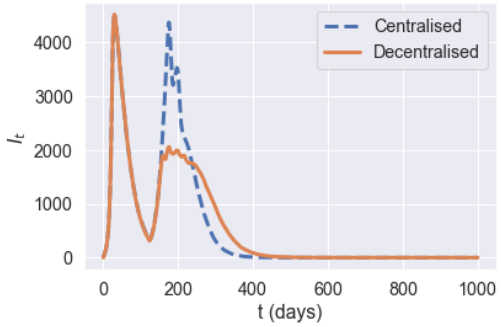
(c) Increase in fatalities (per 100,000 inhabitants) when moving from regional to centralised policy.

Figure 4.38: Fatalities per 100,000 inhabitants for centralised (left) and regional (right) adaptive mitigation policies. Same triggering thresholds are used in both cases: $B_{\text{on}} = 4$ and $B_{\text{off}} = 0.4B_{\text{on}}$.

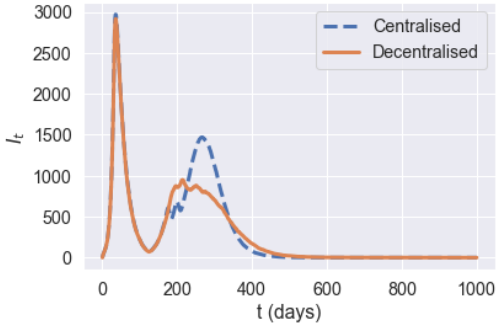
Figure 4.39a compares the dynamics of symptomatic infections (I_t) for the same example. There is a reduction of 100,000 symptomatic individuals in the amplitude of the second peak value when moving from a centralised policy to a decentralised one. The decentralised policy also dampens the second peak values in most regions. Similar effects are observed for York (Figure 4.39c) and Leicester (Figure 4.39b).



(a) Number of symptomatic individuals (I_t) in England under centralised and decentralised policies.



(b) Number of symptomatic individuals (I_t) in Leicester (UKF21).



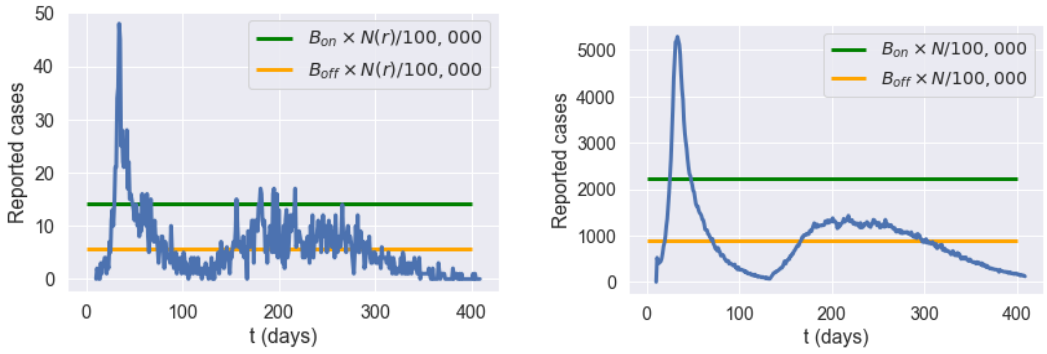
(c) Number of symptomatic individuals (I_t) in York (UKE21).

Figure 4.39: Number of infected individuals under centralised (blue dashed line) and decentralised (orange solid line) policies. Same triggering thresholds are used in both cases: $B_{\text{on}} = 4$ and $B_{\text{off}} = 0.4B_{\text{on}}$.

On June 29, 2020, Leicester became the first city in Britain to be placed in a local lockdown, after public health officials voiced concern at the city's alarming rise in COVID-19 cases. Earlier in June, the Government announced that parts

of the city would be released from lockdown, while a 'targeted' approach will see pockets remain under tighter restrictions. Our simulations indicate a 60% reduction of the second-peak value in Leicester when a decentralised policy is implemented (Figure 4.39b).

Example Figure 4.40 shows an example of such a decentralised triggering policy, with the same triggering thresholds as in the centralised example in Figure 4.32. At the regional level, we see in Figure 4.40a that this policy is more successful than the centralised policy in taming the local outbreaks in Leicester, substantially reducing the second peak through four week long regional lockdowns. At the national level this results in a strong damping of second wave infections, as shown in Figure 4.40b (compare with Figure 4.32a).



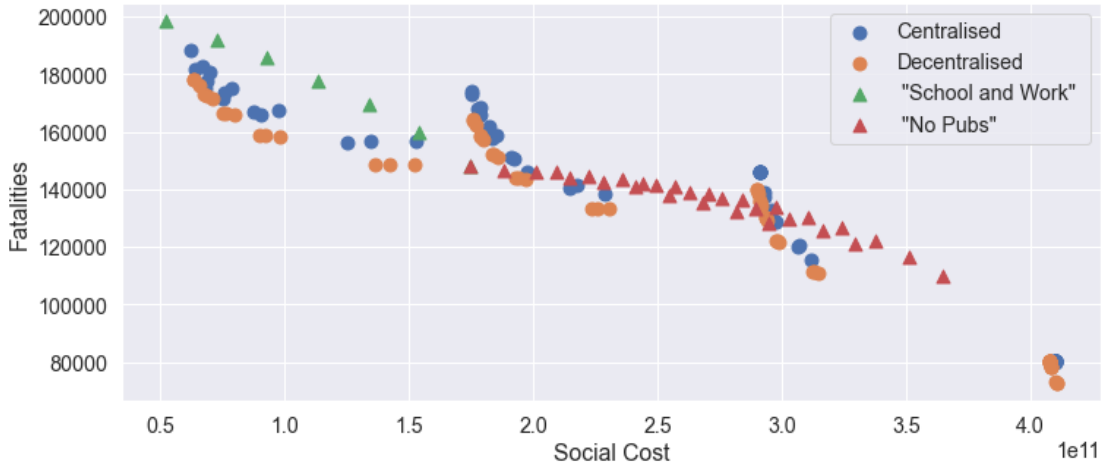
(a) Daily reported cases in Leicester.

(b) Daily reported cases in England.

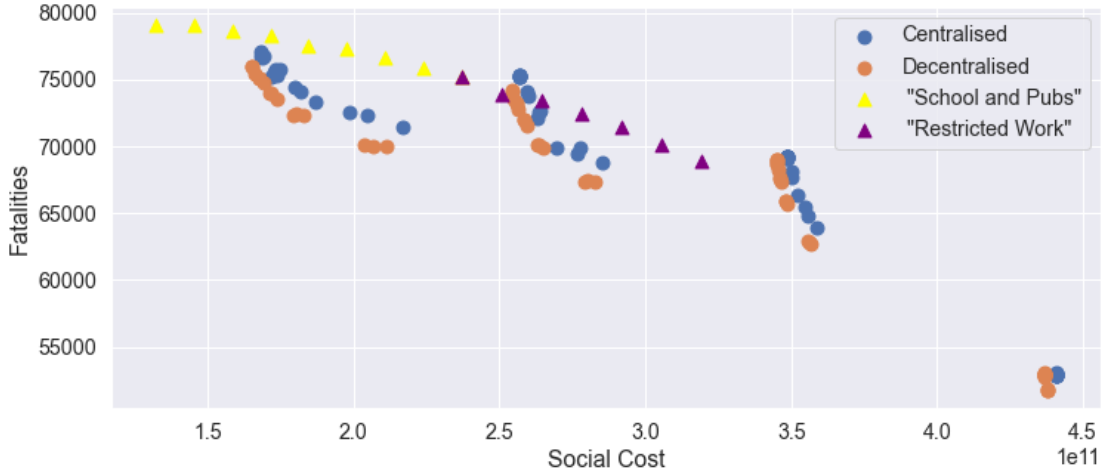
Figure 4.40: Reported cases in England and Leicester under a decentralised triggering policy: an average of 50 simulated scenarios with $B_{\text{on}} = 4$, $B_{\text{off}} = 0.4 \times R_{on}$, $m = 0.5$, no shielding.

4.7.3 Adaptive and Pre-Planned Policies

Figure 4.41 compares the health outcome and social cost of the efficient policies considered in Sections 4.6.2, 4.7.1 and 4.7.2. The efficient frontier of pre-planned policies are among policies with $u^S \in \{0, 0.5, 1\}$, $0.2 \leq u^W \leq 1.0$ and $0.3 \leq u^O \leq 1.0$. For centralised and decentralised policies, $m = 0.25, 0.5, 0.75, 1$; $B_{\text{on}} = 2, 4, 6, 8, 10$; and $B_{\text{off}} = p \times B_{\text{on}}$ with $p = 0.2, 0.4, 0.8$.



(a) Results for policies with no shielding of senior citizens.



(b) Results for policies with shielding of senior citizens enacted.

Figure 4.41: Efficiency plot: pre-planned versus adaptive mitigation policies.

We observe that:

- Adaptive policies, in which measures are triggered when the number of daily new cases exceeds a threshold, are more efficient than pre-planned policies; and
- As shown in Figures 4.41a and 4.41b, a decentralised policy is more efficient than both centralised and pre-planned policies.

In Table 4.10, we provide a summary of outcomes for five different types of policies, defined as follows. We note that the outcomes are averaged across 50 scenarios, starting from the same initial conditions on July 4 (end of the UK lockdown).

1. Confinement of $T = 105$ days followed by social distancing ($m = 0.3$ or $m = 0.5$), no shielding.

2. Pre-planned policy: social distancing at work and school ($u^H = 1$, $u^S = 0.5$, $u^W = 0.5$), restrictions on social gatherings ($u^O = 0.3$) and no shielding.

3. Centralised and decentralised triggering policies (Sections 4.7.1 and 4.7.2) with $m = 0.5$, $B_{\text{on}} = 4$, $B_{\text{off}} = 0.4B_{\text{on}}$ and no shielding.

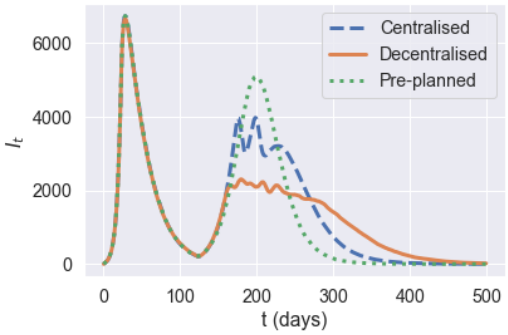
4. Decentralised triggering combined with shielding of senior populations: $m = 0.5$, $B_{\text{on}} = 4$, $B_{\text{off}} = 0.4B_{\text{on}}$.

5. 'Protect Lives' policy: in the range of efficient policies, the one which results in the fewest fatalities is a decentralised triggering policy with $B_{\text{on}} = 2$, $B_{\text{off}} = 0.2B_{\text{on}}$ (so more frequent triggering of confinement measures than the above), high degree of social distancing ($m = 0.25$) and shielding of senior populations. This policy corresponds to the point in the lower right corner of Figure 4.41b. The social cost is 4.3×10^{11}, which is much higher than for the other considered policies.

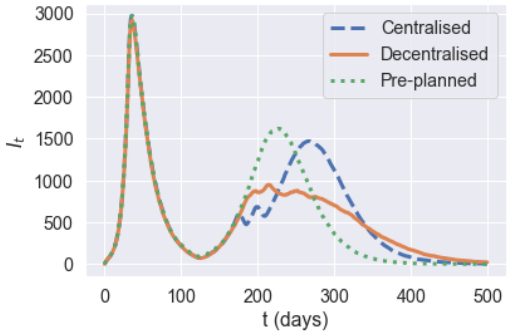
Policy	I_t (Aug 1)	A_t (Aug 1)	Fatalities (Aug 1)	max I_t (2^{nd} peak)	Social cost (10^{11})	Projected fatalities (1,000 days)
Confinement followed by strict social distancing ($m = 0.3$)	47,400	188,700	39,400	255,700	3.8	96,600
Confinement followed by moderate social distancing ($m = 0.5$)	98,400	392,400	40,700	766,800	2.9	146,100
Pre-planned	84,700	360,100	45,500	613,300	2.9	122,900
Centralised triggering	80,300	321,200	40,500	423,200	3.0	133,500
Decentralised triggering	80,100	320,200	40,400	292,200	3.0	122,100
Decentralised triggering and shielding	55,000	266,100	39,700	267,700	3.4	65,900
'Protect Lives'	25,900	118,400	39,600	63,900	4.3	51,700

Table 4.10: Summary of outcomes for different policies, starting from the same initial conditions on July 4, 2020.

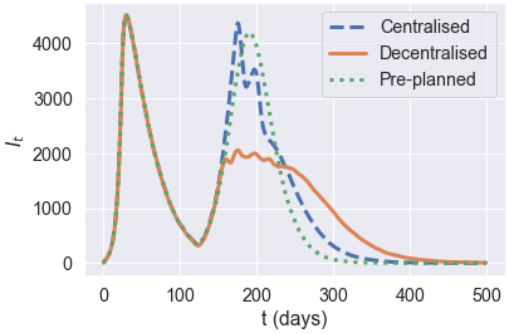
Regional outcomes Comparison of the regional outcomes of the centralised, decentralised and pre-planned policies displayed in Table 4.10 shows that the decentralised triggering policies are able in many cases to considerably dampen the second wave of infections. Figure 4.42 illustrates this in the case of Mid Lancashire, York, Leicester and Birmingham: the decentralised triggering policy reduces the second peak amplitude by around one half as compared to the pre-planned policy.



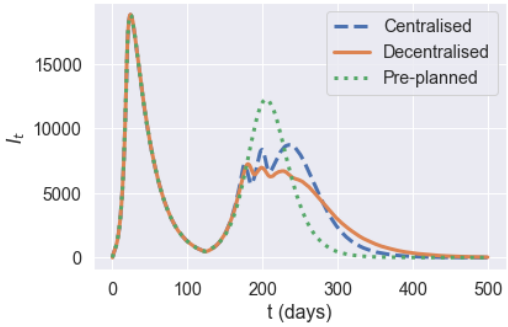
(a) I_t in Mid Lancashire.



(b) I_t in York.



(c) I_t in Leicester.



(d) I_t in Birmingham.

Figure 4.42: Regional comparison of pre-planned and adaptive mitigation policies.

Appendix A

Supplementary Material for the UK Reinsurance Market Study

A.1 Treaty and Facultative Statistics

Table A.1 presents a quantitative summary of premiums ceded by contract type and by line of business in the treaty and facultative data set.

Type of contracts	**Frequency** (percent of total)	**Value in £m** (percent of total)
Life, external	1905 (4.5%)	62181 (57.7%)
Life, internal	81 (0.2%)	7557 (7%)
Non-life, external	37048 (88.5%)	25286 (23.5%)
Non-life, internal	1042 (2.5%)	11833 (11%)
Life, proportional	1721 (4.1%)	69160 (64.2%)
Life, non-proportional	258 (0.6%)	115 (0.1%)
Non-life, proportional	6177 (14.7%)	27621 (25.6%)
Non-life, non-proportional	33700 (80.5%)	10306 (9.6%)
Fire and other damage to property	15941 (38.1%)	12717 (11.8%)
Marine, Aviation, Transport	12253 (29.3%)	3288 (3.1%)
General liability	5556 (13.3%)	4973 (4.6%)
Other non-life	1689 (4%)	1566 (1.5%)
Other life	1428 (3.4%)	52003 (48.3%)
Motor insurance	1315 (3.1%)	2642 (2.5%)
Credit and suretyship insurance	1311 (3.1%)	685 (0.6%)
Health reinsurance	1162 (2.8%)	4316 (4%)
Unit-linked or index-linked	318 (0.8%)	16681 (15.5%)
Multiline	317 (0.8%)	7326 (6.8%)
Medical expense insurance	294 (0.7%)	335 (0.3%)
Life reinsurance	201 (0.5%)	916 (0.9%)
With profit participation	56 (0.1%)	139 (0.1%)
Non-life annuities	42 (0.1%)	110 (0.1%)

Table A.1: Summary of premiums ceded in the treaty and facultative data set for the 2016 year-end.

A.2 Overview of Original Data and Adjustments

Table A.2 presents a detailed overview of original data and performed adjustments for the following networks and layers:

- (TF-A) Treaty and Facultative, all contracts.
- (TF-L) Treaty and Facultative, only life contracts.
- (TF-NL) Treaty and Facultative, only non-life contracts.
- (TF-G) Treaty and Facultative, group contracts.
- (R-A) Recoverables (net of collateral), all contracts.
- (R-L) Recoverables (net of collateral), only life contracts.
- (R-NL) Recoverables (net of collateral), only non-life contracts.
- (R-G) Recoverables (net of collateral), group contracts.

Network		**TR-A**	**TR-L**	**TR-NL**	**TR-G**	**R-A**	**R-L**	**R-NL**	**R-G**
Nodes	Total	893	162	825	385	4378	427	3682	2954
	Incoming links	799	125	743	345	1494	156	1354	920
	Outgoing links	225	49	202	117	3940	351	3246	2640
	Both	131	12	120	77	1056	80	918	606
Edges	All contracts	41883	2003	39880	38190	25891	834	21346	25432
	Links	8564	305	8338	3545	22713	822	19125	11649
Adjustments	Initial contracts	50771	3913	46858	46281	26269	834	21374	26269
	Weight threshold above £0	1359	276	1083	1292	0	0	0	0
	Self-links	14	0	14	768	34	0	27	0
	Blanks	7515	1634	5881	6031	344	337	0	340
	Reverse negatives	0	0	0	0	5349	159	5076	5349
	Final number of contracts	41883	2003	39880	38190	25891	834	21346	25432

Table A.2: Overview of data adjustments and quantitative statistics. Networks based on the treaty and facultative data set (TR) and the recoverables data set (R). Information shown for all contracts (A), life contracts (L), non-life contracts (NL), and contracts between distinct insurance groups (G).

Appendix B

Supplementary Material for the COVID-19 Model

B.1 Demographic Regions

Table B.1 details the used mapping between Upper Tier Local Authority (UTLA) region codes and the Nomenclature of Territorial Units for Statistics at level 3 (NUTS-3) codes. If more than one UTLA region falls within the boundary of a single NUTS-3 region, the data is then aggregated. On the other hand, if a single UTLA region lies within more than one NUTS-3 region, the data is distributed among NUTS-3 regions in proportion to the total number of people living in each region.

UTLA Code	**UTLA Region Name**	**NUTS-3 Code Mapping**
E06000001	Hartlepool	UKC11
E06000002	Middlesbrough	UKC12
E06000003	Redcar and Cleveland	UKC12
E06000004	Stockton-on-Tees	UKC11
E06000005	Darlington	UKC13
E06000006	Halton	UKD71
E06000007	Warrington	UKD61
E06000008	Blackburn with Darwen	UKD41
E06000009	Blackpool	UKD42
E06000010	Kingston upon Hull, City of	UKE11
E06000011	East Riding of Yorkshire	UKE12
E06000012	North East Lincolnshire	UKE13
E06000013	North Lincolnshire	UKE13
E06000014	York	UKE21
E06000015	Derby	UKF11
E06000016	Leicester	UKF21
E06000017	Rutland	UKF22
E06000018	Nottingham	UKF14
E06000019	Herefordshire, County of	UKG11

E06000020	Telford and Wrekin	UKG21
E06000021	Stoke-on-Trent	UKG23
E06000022	Bath and North East Somerset	UKK12
E06000023	Bristol, City of	UKK11
E06000024	North Somerset	UKK12
E06000025	South Gloucestershire	UKK12
E06000026	Plymouth	UKK41
E06000027	Torbay	UKK42
E06000030	Swindon	UKK14
E06000031	Peterborough	UKH11
E06000032	Luton	UKH21
E06000033	Southend-on-Sea	UKH31
E06000034	Thurrock	UKH32
E06000035	Medway	UKJ41
E06000036	Bracknell Forest	UKJ11
E06000037	West Berkshire	UKJ11
E06000038	Reading	UKJ11
E06000039	Slough	UKJ11
E06000040	Windsor and Maidenhead	UKJ11
E06000041	Wokingham	UKJ11
E06000042	Milton Keynes	UKJ12
E06000043	Brighton and Hove	UKJ21
E06000044	Portsmouth	UKJ31
E06000045	Southampton	UKJ32
E06000046	Isle of Wight	UKJ34
E06000047	County Durham	UKC14
E06000049	Cheshire East	UKD62
E06000050	Cheshire West and Chester	UKD63
E06000051	Shropshire	UKG22
E06000052	Cornwall and Isles of Scilly	UKK30
E06000054	Wiltshire	UKK15
E06000055	Bedford	UKH24
E06000056	Central Bedfordshire	UKH25
E06000057	Northumberland	UKC21
E06000058	Bournemouth and Poole	UKK21
E06000059	Dorset	UKK22
E08000001	Bolton	UKD36
E08000002	Bury	UKD37
E08000003	Manchester	UKD33
E08000004	Oldham	UKD37
E08000005	Rochdale	UKD37
E08000006	Salford	UKD34
E08000007	Stockport	UKD35
E08000008	Tameside	UKD35
E08000009	Trafford	UKD34
E08000010	Wigan	UKD36
E08000011	Knowsley	UKD71
E08000012	Liverpool	UKD72
E08000013	St. Helens	UKD71
E08000014	Sefton	UKD73
E08000015	Wirral	UKD74
E08000016	Barnsley	UKE31
E08000017	Doncaster	UKE31
E08000018	Rotherham	UKE31
E08000019	Sheffield	UKE32
E08000021	Newcastle upon Tyne	UKC22
E08000022	North Tyneside	UKC22
E08000023	South Tyneside	UKC22

E08000024	Sunderland	UKC23
E08000025	Birmingham	UKG31
E08000026	Coventry	UKG33
E08000027	Dudley	UKG36
E08000028	Sandwell	UKG37
E08000029	Solihull	UKG32
E08000030	Walsall	UKG38
E08000031	Wolverhampton	UKG39
E08000032	Bradford	UKE41
E08000033	Calderdale	UKE44
E08000034	Kirklees	UKE44
E08000035	Leeds	UKE42
E08000036	Wakefield	UKE45
E08000037	Gateshead	UKC22
E09000001	City of London	UKI31
E09000002	Barking and Dagenham	UKI52
E09000003	Barnet	UKI71
E09000004	Bexley	UKI51
E09000005	Brent	UKI72
E09000006	Bromley	UKI61
E09000007	Camden	UKI31
E09000008	Croydon	UKI62
E09000009	Ealing	UKI73
E09000010	Enfield	UKI54
E09000011	Greenwich	UKI51
E09000012	Hackney	UKI41
E09000013	Hammersmith and Fulham	UKI33
E09000014	Haringey	UKI43
E09000015	Harrow	UKI74
E09000016	Havering	UKI52
E09000017	Hillingdon	UKI74
E09000018	Hounslow	UKI75
E09000019	Islington	UKI43
E09000020	Kensington and Chelsea	UKI33
E09000021	Kingston upon Thames	UKI63
E09000022	Lambeth	UKI45
E09000023	Lewisham	UKI44
E09000024	Merton	UKI63
E09000025	Newham	UKI41
E09000026	Redbridge	UKI53
E09000027	Richmond upon Thames	UKI75
E09000028	Southwark	UKI44
E09000029	Sutton	UKI63
E09000030	Tower Hamlets	UKI42
E09000031	Waltham Forest	UKI53
E09000032	Wandsworth	UKI34
E09000033	Westminster	UKI32
E10000002	Buckinghamshire	UKJ13
E10000003	Cambridgeshire	UKH12
E10000006	Cumbria	UKD11, UKD12
E10000007	Derbyshire	UKF13, UKF12
E10000008	Devon	UKK43
E10000011	East Sussex	UKJ22
E10000012	Essex	UKH37, UKH34, UKH35, UKH36
E10000013	Gloucestershire	UKK13, UKK12
E10000014	Hampshire	UKJ36, UKJ37, UKJ35
E10000015	Hertfordshire	UKH23
E10000016	Kent	UKJ43, UKJ44, UKJ45, UKJ46

E10000017	Lancashire	UKD45, UKD46, UKD47, UKD44
E10000018	Leicestershire	UKF22
E10000019	Lincolnshire	UKE13, UKF30
E10000020	Norfolk	UKH15, UKH17, UKH16
E10000021	Northamptonshire	UKF24, UKF25
E10000023	North Yorkshire	UKE22
E10000024	Nottinghamshire	UKF15, UKF16
E10000025	Oxfordshire	UKJ14
E10000027	Somerset	UKK12, UKK23
E10000028	Staffordshire	UKG24
E10000029	Suffolk	UKH14
E10000030	Surrey	UKJ25, UKJ26
E10000031	Warwickshire	UKG13
E10000032	West Sussex	UKJ28, UKJ27
E10000034	Worcestershire	UKG12

Table B.1: Mapping between the Upper Tier Local Authority (UTLA) regions and the Nomenclature of Territorial Units for Statistics at level 3 codes (NUTS-3).

B.2 Baseline Parameters for Social Contact Rates

This appendix outlines the sources used for the baseline social contact rate parameters. In particular, two sources have been used: the POLYMOD study (Mossong et al., 2017), processed using the methodology of PyRoss (Adhikari et al., 2020), and the BBC Pandemic study (Klepac et al., 2020), which is used as a robustness check. We find our estimates of contact rates to be consistent with (Klepac et al., 2020) across different ages and locations.

We use the following estimates for social contact rates across the 16 age groups (detailed in Table 4.1) given in Mossong et al. (2017):

$$\sigma = \begin{bmatrix} 1.92 & 0.81 & 0.47 & 0.30 & 0.49 & 0.79 & 0.89 & 1.07 & 0.44 & 0.27 & 0.35 & 0.27 & 0.22 & 0.15 & 0.10 & 0.02 \\ 0.78 & 6.64 & 1.24 & 0.58 & 0.49 & 0.72 & 1.09 & 1.40 & 1.10 & 0.36 & 0.35 & 0.23 & 0.35 & 0.24 & 0.07 & 0.23 \\ 0.42 & 1.16 & 6.85 & 1.30 & 0.25 & 0.37 & 0.57 & 1.10 & 1.18 & 0.64 & 0.35 & 0.35 & 0.2 & 0.2 & 0.17 & 0.14 \\ 0.26 & 0.52 & 1.26 & 6.71 & 1.24 & 0.72 & 0.47 & 0.87 & 0.97 & 0.97 & 0.52 & 0.31 & 0.2 & 0.26 & 0.24 & 0.28 \\ 0.43 & 0.44 & 0.24 & 1.26 & 2.59 & 1.36 & 0.84 & 0.76 & 0.83 & 0.93 & 0.63 & 0.5 & 0.31 & 0.22 & 0.16 & 0.17 \\ 0.73 & 0.68 & 0.38 & 0.76 & 1.42 & 1.83 & 1.13 & 0.92 & 0.9 & 0.92 & 0.85 & 0.72 & 0.45 & 0.38 & 0.18 & 0.12 \\ 0.73 & 0.93 & 0.53 & 0.44 & 0.79 & 1.02 & 1.67 & 1.27 & 0.98 & 0.72 & 0.7 & 0.63 & 0.48 & 0.27 & 0.09 & 0.27 \\ 0.79 & 1.06 & 0.89 & 0.74 & 0.63 & 0.74 & 1.12 & 1.5 & 1.27 & 0.86 & 0.63 & 0.55 & 0.53 & 0.43 & 0.14 & 0.31 \\ 0.32 & 0.83 & 0.97 & 0.83 & 0.69 & 0.73 & 0.87 & 1.28 & 1.35 & 1.21 & 0.7 & 0.55 & 0.55 & 0.35 & 0.33 & 0.43 \\ 0.24 & 0.32 & 0.62 & 0.96 & 0.91 & 0.87 & 0.75 & 1.02 & 1.41 & 1.87 & 0.75 & 0.64 & 0.51 & 0.26 & 0.32 & 0.33 \\ 0.34 & 0.34 & 0.37 & 0.57 & 0.68 & 0.88 & 0.81 & 0.82 & 0.90 & 0.82 & 0.74 & 0.98 & 0.46 & 0.31 & 0.28 & 0.76 \\ 0.24 & 0.20 & 0.35 & 0.32 & 0.50 & 0.69 & 0.67 & 0.66 & 0.66 & 0.66 & 0.91 & 1.17 & 0.73 & 0.43 & 0.20 & 0.46 \\ 0.24 & 0.40 & 0.25 & 0.25 & 0.38 & 0.54 & 0.65 & 0.80 & 0.82 & 0.65 & 0.53 & 0.91 & 0.65 & 0.55 & 0.30 & 0.66 \\ 0.19 & 0.31 & 0.29 & 0.39 & 0.32 & 0.52 & 0.42 & 0.74 & 0.60 & 0.39 & 0.42 & 0.63 & 0.64 & 0.70 & 0.52 & 0.20 \\ 0.15 & 0.11 & 0.28 & 0.40 & 0.26 & 0.29 & 0.17 & 0.28 & 0.66 & 0.55 & 0.44 & 0.34 & 0.40 & 0.60 & 0.59 & 0.57 \\ 0.01 & 0.18 & 0.12 & 0.24 & 0.14 & 0.10 & 0.24 & 0.32 & 0.43 & 0.28 & 0.60 & 0.39 & 0.45 & 0.12 & 0.29 & 0.86 \end{bmatrix} \tag{B.1}$$

Entry (i, j) of the matrix (B.1) corresponds to the average daily number of contacts between a person in age group i with people in age group j. By convention, we let $i = 1$ correspond to age group $[0, 5)$, $i = 2$ to $[5, 10)$, ..., and $i = 16$ to $[75, 100)$.

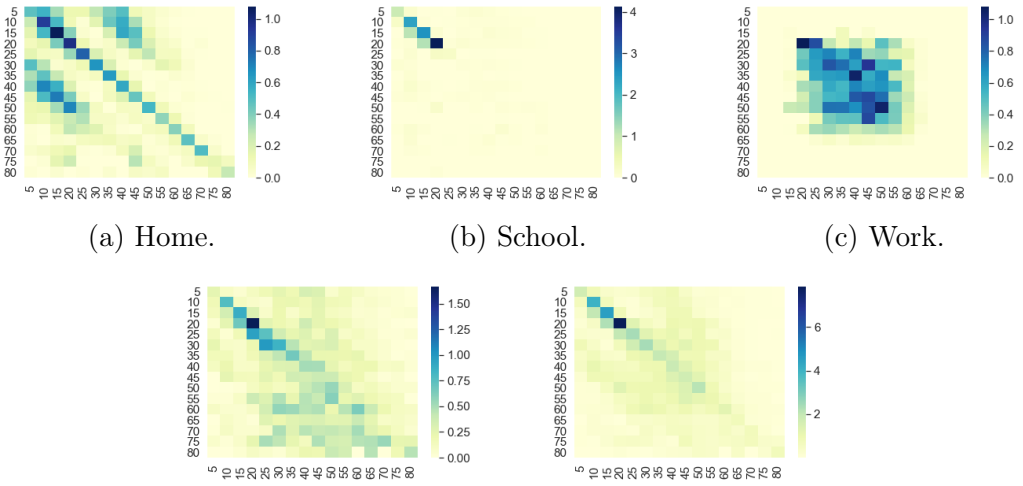
We use the Bayesian hierarchical framework provided by the PyRoss library (Adhikari et al., 2020) to decompose contact rates into 'home', 'school', 'work', and 'other' (Prem et al., 2017). The results are stated below, and also visualised in Figure B.1.

$$\sigma^H = \begin{bmatrix}
0.48 & 0.55 & 0.33 & 0.13 & 0.14 & 0.28 & 0.41 & 0.49 & 0.11 & 0.07 & 0.04 & 0.02 & 0.01 & 0.00 & 0.00 & 0.00 \\
0.26 & 0.92 & 0.52 & 0.12 & 0.03 & 0.17 & 0.45 & 0.58 & 0.32 & 0.07 & 0.03 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.17 & 0.54 & 1.08 & 0.39 & 0.04 & 0.01 & 0.22 & 0.59 & 0.49 & 0.13 & 0.04 & 0.02 & 0.00 & 0.01 & 0.00 & 0.00 \\
0.09 & 0.15 & 0.42 & 0.98 & 0.13 & 0.03 & 0.06 & 0.25 & 0.42 & 0.21 & 0.07 & 0.05 & 0.01 & 0.02 & 0.00 & 0.00 \\
0.17 & 0.08 & 0.07 & 0.36 & 0.80 & 0.21 & 0.07 & 0.04 & 0.16 & 0.28 & 0.08 & 0.09 & 0.02 & 0.01 & 0.00 & 0.00 \\
0.49 & 0.30 & 0.04 & 0.07 & 0.13 & 0.66 & 0.21 & 0.03 & 0.01 & 0.05 & 0.17 & 0.11 & 0.03 & 0.00 & 0.00 & 0.00 \\
0.32 & 0.47 & 0.27 & 0.08 & 0.05 & 0.09 & 0.64 & 0.15 & 0.03 & 0.00 & 0.01 & 0.04 & 0.02 & 0.00 & 0.00 & 0.00 \\
0.38 & 0.70 & 0.56 & 0.20 & 0.02 & 0.01 & 0.09 & 0.59 & 0.15 & 0.00 & 0.02 & 0.01 & 0.01 & 0.01 & 0.00 & 0.00 \\
0.17 & 0.52 & 0.73 & 0.42 & 0.07 & 0.00 & 0.09 & 0.19 & 0.46 & 0.10 & 0.02 & 0.00 & 0.03 & 0.01 & 0.01 & 0.00 \\
0.13 & 0.15 & 0.33 & 0.71 & 0.35 & 0.08 & 0.01 & 0.08 & 0.07 & 0.53 & 0.11 & 0.03 & 0.01 & 0.00 & 0.00 & 0.01 \\
0.12 & 0.10 & 0.18 & 0.26 & 0.31 & 0.07 & 0.07 & 0.04 & 0.08 & 0.06 & 0.37 & 0.13 & 0.01 & 0.00 & 0.00 & 0.00 \\
0.02 & 0.01 & 0.11 & 0.28 & 0.23 & 0.18 & 0.06 & 0.02 & 0.01 & 0.08 & 0.09 & 0.39 & 0.10 & 0.00 & 0.00 & 0.00 \\
0.02 & 0.00 & 0.05 & 0.03 & 0.09 & 0.10 & 0.11 & 0.07 & 0.08 & 0.02 & 0.04 & 0.11 & 0.48 & 0.06 & 0.01 & 0.00 \\
0.05 & 0.07 & 0.11 & 0.15 & 0.02 & 0.01 & 0.02 & 0.10 & 0.22 & 0.03 & 0.02 & 0.05 & 0.08 & 0.51 & 0.05 & 0.00 \\
0.00 & 0.02 & 0.16 & 0.25 & 0.01 & 0.00 & 0.00 & 0.00 & 0.30 & 0.08 & 0.03 & 0.00 & 0.06 & 0.14 & 0.16 & 0.09 \\
0.02 & 0.00 & 0.04 & 0.00 & 0.02 & 0.00 & 0.04 & 0.00 & 0.07 & 0.05 & 0.08 & 0.00 & 0.00 & 0.00 & 0.11 & 0.27
\end{bmatrix} \tag{B.2}$$

$$\sigma^S = \begin{bmatrix}
0.97 & 0.15 & 0.01 & 0.03 & 0.01 & 0.09 & 0.13 & 0.09 & 0.04 & 0.03 & 0.00 & 0.00 & 0.01 & 0.00 & 0.00 & 0.00 \\
0.24 & 2.35 & 0.06 & 0.00 & 0.02 & 0.04 & 0.06 & 0.06 & 0.06 & 0.05 & 0.04 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 1.12 & 2.56 & 0.12 & 0.01 & 0.05 & 0.04 & 0.12 & 0.09 & 0.07 & 0.03 & 0.02 & 0.01 & 0.00 & 0.00 & 0.00 \\
0.04 & 0.08 & 1.17 & 4.14 & 0.06 & 0.12 & 0.08 & 0.08 & 0.07 & 0.05 & 0.04 & 0.03 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.13 & 0.00 & 0.27 & 0.23 & 0.01 & 0.02 & 0.03 & 0.01 & 0.01 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.11 & 0.07 & 0.01 & 0.00 & 0.00 & 0.06 & 0.04 & 0.06 & 0.04 & 0.03 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.10 & 0.01 & 0.01 & 0.01 & 0.00 & 0.04 & 0.00 & 0.02 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.04 & 0.17 & 0.07 & 0.03 & 0.00 & 0.02 & 0.05 & 0.07 & 0.07 & 0.05 & 0.00 & 0.01 & 0.00 & 0.01 & 0.00 & 0.00 \\
0.07 & 0.10 & 0.03 & 0.00 & 0.04 & 0.06 & 0.03 & 0.06 & 0.02 & 0.01 & 0.01 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.02 & 0.00 & 0.02 & 0.21 & 0.00 & 0.00 & 0.05 & 0.01 & 0.07 & 0.05 & 0.06 & 0.04 & 0.01 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.02 & 0.00 & 0.00 & 0.04 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.05 & 0.14 & 0.05 & 0.00 & 0.00 & 0.00 & 0.03 & 0.00 & 0.15 & 0.03 & 0.00 & 0.08 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.13 & 0.00 & 0.00 & 0.02 & 0.01 & 0.05 & 0.11 & 0.02 & 0.03 & 0.01 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.06 & 0.04 & 0.00 & 0.00 & 0.00 & 0.04 & 0.03 & 0.00 & 0.00 & 0.00 & 0.03 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.06 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00
\end{bmatrix} \tag{B.3}$$

$$\sigma^W = \begin{bmatrix}
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.02 & 0.00 & 0.00 & 0.00 & 0.10 & 0.00 & 0.04 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.01 & 1.08 & 0.87 & 0.28 & 0.11 & 0.28 & 0.15 & 0.48 & 0.32 & 0.06 & 0.02 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.02 & 0.44 & 0.64 & 0.71 & 0.56 & 0.71 & 0.35 & 0.37 & 0.17 & 0.16 & 0.02 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.01 & 0.49 & 0.61 & 0.84 & 0.77 & 0.62 & 0.95 & 0.54 & 0.55 & 0.09 & 0.02 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.05 & 0.10 & 0.59 & 0.63 & 0.65 & 1.05 & 0.58 & 0.69 & 0.27 & 0.21 & 0.02 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.02 & 0.33 & 0.37 & 0.59 & 0.54 & 0.70 & 0.62 & 0.66 & 0.52 & 0.19 & 0.01 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.01 & 0.23 & 0.39 & 0.53 & 0.55 & 0.83 & 0.88 & 0.77 & 0.35 & 0.25 & 0.02 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.26 & 0.28 & 0.39 & 0.76 & 0.76 & 0.68 & 0.85 & 1.04 & 0.42 & 0.19 & 0.03 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.12 & 0.43 & 0.47 & 0.48 & 0.88 & 0.52 & 0.42 & 0.16 & 0.05 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.01 & 0.09 & 0.33 & 0.36 & 0.31 & 0.25 & 0.29 & 0.33 & 0.30 & 0.14 & 0.01 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.03 & 0.07 & 0.03 & 0.09 & 0.07 & 0.05 & 0.06 & 0.04 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00
\end{bmatrix} \tag{B.4}$$

$$\sigma^O = \begin{bmatrix}
0.26 & 0.10 & 0.05 & 0.13 & 0.19 & 0.26 & 0.19 & 0.34 & 0.31 & 0.07 & 0.15 & 0.11 & 0.06 & 0.04 & 0.04 & 0.01 \\
0.18 & 0.77 & 0.13 & 0.09 & 0.06 & 0.22 & 0.23 & 0.18 & 0.26 & 0.17 & 0.16 & 0.06 & 0.09 & 0.06 & 0.04 & 0.01 \\
0.15 & 0.35 & 0.88 & 0.31 & 0.09 & 0.21 & 0.09 & 0.17 & 0.26 & 0.13 & 0.11 & 0.05 & 0.06 & 0.02 & 0.04 & 0.05 \\
0.04 & 0.26 & 0.68 & 1.67 & 0.27 & 0.20 & 0.16 & 0.37 & 0.25 & 0.21 & 0.16 & 0.08 & 0.06 & 0.04 & 0.01 & 0.00 \\
0.15 & 0.04 & 0.18 & 0.96 & 0.75 & 0.37 & 0.24 & 0.30 & 0.16 & 0.32 & 0.09 & 0.14 & 0.08 & 0.03 & 0.02 & 0.05 \\
0.25 & 0.09 & 0.08 & 0.20 & 0.99 & 0.85 & 0.46 & 0.27 & 0.29 & 0.32 & 0.20 & 0.10 & 0.09 & 0.05 & 0.02 & 0.00 \\
0.16 & 0.17 & 0.09 & 0.18 & 0.32 & 0.44 & 0.67 & 0.44 & 0.23 & 0.23 & 0.22 & 0.25 & 0.10 & 0.06 & 0.01 & 0.02 \\
0.13 & 0.22 & 0.10 & 0.10 & 0.24 & 0.30 & 0.31 & 0.52 & 0.51 & 0.25 & 0.15 & 0.20 & 0.21 & 0.10 & 0.11 & 0.04 \\
0.00 & 0.25 & 0.15 & 0.15 & 0.27 & 0.18 & 0.36 & 0.35 & 0.39 & 0.39 & 0.24 & 0.20 & 0.10 & 0.02 & 0.05 & 0.00 \\
0.00 & 0.01 & 0.03 & 0.13 & 0.19 & 0.13 & 0.17 & 0.39 & 0.46 & 0.56 & 0.32 & 0.12 & 0.16 & 0.10 & 0.06 & 0.06 \\
0.03 & 0.03 & 0.10 & 0.39 & 0.30 & 0.48 & 0.24 & 0.46 & 0.41 & 0.60 & 0.23 & 0.29 & 0.20 & 0.14 & 0.12 & 0.01 \\
0.11 & 0.06 & 0.06 & 0.11 & 0.36 & 0.57 & 0.55 & 0.46 & 0.49 & 0.22 & 0.46 & 0.64 & 0.40 & 0.20 & 0.12 & 0.11 \\
0.00 & 0.09 & 0.07 & 0.03 & 0.29 & 0.22 & 0.28 & 0.31 & 0.32 & 0.27 & 0.28 & 0.40 & 0.26 & 0.09 & 0.11 & 0.08 \\
0.02 & 0.11 & 0.06 & 0.06 & 0.11 & 0.49 & 0.29 & 0.46 & 0.42 & 0.43 & 0.39 & 0.52 & 0.29 & 0.22 & 0.14 & 0.06 \\
0.07 & 0.03 & 0.09 & 0.24 & 0.51 & 0.44 & 0.24 & 0.14 & 0.49 & 0.36 & 0.34 & 0.47 & 0.48 & 0.55 & 0.24 & 0.25 \\
0.00 & 0.00 & 0.04 & 0.11 & 0.07 & 0.04 & 0.13 & 0.22 & 0.20 & 0.41 & 0.35 & 0.00 & 0.47 & 0.00 & 0.19 & 0.47
\end{bmatrix} \tag{B.5}$$



(a) Home. (b) School. (c) Work.

(d) Other. (e) Total.

Figure B.1: Baseline social contact matrices with 16 age groups.

Adjusting for the regional contact heterogeneity It should be noted that the above parameters are used as baseline contact rates, and a further detailed calibration is carried out region by region to account for heterogeneity of social contact patterns across UK regions. In particular, we use a simulation-based indirect inference approach of Gourieroux et al. (1993) to estimate these regional adjustments in order to match the heterogeneous pandemic dynamics across 133 regions in England.[1]

[1]Our approach is analogous to the one used in estimating the infection rate α, detailed in Section 4.3.4.

Denote by b_r^1 and b_r^2 the respective slopes of log reported cumulative cases $\{\log C_t^r\}_{t=1}^T$ in region r before and after a given date t_r, for each region $r = 1, 2, \ldots, K$. We introduce t_r to be the effective day of the lockdown in region r (that is, March 23 + lag) to accommodate for the observed regional heterogeneity in data. In particular, the effective date may be region-dependent since different regions have different hospital capacities and limitations in case testing. We denote C_t^r and C_t to be the respective number of reported cases in region r and in England. Correspondingly, we let $\widetilde{C}_t^r$ and $\widetilde{C}^r$ be the estimated respective number of reported cases in region r and in England obtained from the simulation in our model, where we use values averaged over 100 simulated paths.

The procedure for the estimation of the adjustments factors for the regional contacts $\{d_r\}_{r=1}^{133}$ and $\{l_r\}_{r=1}^{133}$ is based on an indirect inference approach of Gourieroux et al. (1993). Algorithm 1 summarises our estimation procedure. Empirically, we find that 1,000 iteration steps and setting $\eta_1 = \eta_2 = 0.2$ results in a good rate of convergence for the parameters of interest (d_r and l_r). We also note that $t_r = 32$ corresponds to March 23.

Algorithm 1 Estimation of Contact Adjustment Factors

Input: Sampling horizon T; model parameters α, β, γ, p; slopes estimated from data $\{b_r^1\}_{r=1}^{133}$ and $\{b_r^2\}_{r=1}^{133}$; learning rates η_1 and η_2; effective lockdown start times $\{t_r\}_{r=1}^{133}$ with $t_r = 32$ for region r.
Initialisation: set regional contact adjustment factors $d_r = 1$ and $l_r = 1$, for every $r = 1, \ldots, 133$.
for $i = 1, \ldots, 1000$ **do**
    Compute regional cases $(\tilde{C}_t^1, \cdots, \tilde{C}_t^{133})_{t=1}^T$ as an average of 100 simulations paths from the SEAIR$(\{t_r\}_{r=1}^{133}, \alpha, \beta, \gamma, p, f, \{d_r\}_{r=1}^{133}, \{l_r\}_{r=1}^{133})$ model.
    for $r = 1, 2, \ldots, 133$ **do**
        Fit ordinary least squares model to estimate the slope of $\log(\widetilde{C}_t^r)$ before lockdown, $\widehat{b}_1^r$, using $\{\tilde{C}_t^r\}_{t=1}^{t_r-1}$; and the slope of $\log \widetilde{C}_t^r$ after lockdown, $\widehat{b}_2^r$, using $\{\widetilde{C}_t^r\}_{t=t_r}^T$.
    for $r = 1, 2, \ldots, 133$ **do**
        Update $d_r \leftarrow d_r + \eta_1(b_1^r - \hat{b}_1^r)$;
        Update $l_r \leftarrow l_r + \eta_2(b_2^r - \hat{b}_2^r)$.
        Increment t_r by a day ($\Delta t_r \in \{-1, +1\}$) to better match the observed peak in the simulation to data, where $t_r \geq 32$.
Return the adjustment factors $\{d_r\}_{r=1}^{133}$, $\{l_r\}_{r=1}^{133}$, and dates $\{t_r\}_{r=1}^{133}$.

Bibliography

Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608, February 2015.

Daron Acemoglu, Victor Chernozhukov, Iván Werning, and Michael D Whinston. A multi-risk SIR model with optimally targeted lockdown. Technical report, National Bureau of Economic Research, 2020.

Viral Acharya and Alberto Bisin. Counterparty risk externality: Centralized versus over-the-counter markets. *Journal of Economic Theory*, 149(C):153–182, 2014.

Viral V. Acharya, Thomas Philippon, and Matthew Richardson. Measuring systemic risk for insurance companies. In Felix Hufeld, Ralph S. J. Koijen, and Christian Thimann, editors, *The Economics, Regulation, and Systemic Risk of Insurance Markets*. Oxford Scholarship Online, 2016.

Ronojoy Adhikari, Austen Bolitho, Fernando Caballero, Michael E. Cates, Jakub Dolezal, Timothy Ekeh, Jules Guioth, Robert L. Jack, Julian Kappler, Lukas Kikuchi, et al. Inference, prediction and optimization of non-pharmaceutical interventions using compartment models: the PyRoss library. *arXiv*, 2020.

Jacob Aguilar, Jeremy Samuel Faust, Jeremy Samuel Faust, Lauren Westafer, and Juan Gutierrez. A model describing COVID-19 community transmission taking into account asymptomatic carriers and risk mitigation. *medRxiv*, 2020.

Muluneh Alene, Leltework Yismaw, Moges Agazhe Assemie, Daniel Bekele Ketema, Wodaje Gietaneh, and Tilahun Yemanu Birhan. Serial interval and

incubation period of COVID-19: a systematic review and meta-analysis. *BMC Infectious Diseases*, 21(257), 2021.

Franklin Allen and Ana Babus. Networks in finance. Working Paper No. 08-07, Wharton Financial Institutions Center, 2008.

Franklin Allen and Douglas Gale. Optimal financial crises. *The Journal of Finance*, 53(4):1245–1284, 1998.

Franklin Allen and Douglas Gale. Financial contagion. *Journal of Political Economy*, 108(1):1–33, 2000.

Linda J. S. Allen. An introduction to stochastic epidemic models. In Fred Brauer, Pauline van den Driessche, and Jianhong Wu, editors, *Mathematical Epidemiology*, pages 81–130. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

Linda J. S. Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2(2):128–142, 2017.

Hamed Amini, Rama Cont, and Andreea Minca. Stress Testing the Resilience of Financial Networks. *International Journal of Theoretical and Applied Finance (IJTAF)*, 15(1):1–20, 2012.

Roy M. Anderson and Robert M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, United Kingdom, 1992.

Jantien A Backer, Don Klinkenberg, and Jacco Wallinga. Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from Wuhan, China. *Eurosurveillance*, 25(5), 2020.

Andrew Bain. Insurance spirals and the Lloyd's market. *The Geneva Papers on risk and insurance*, 24(2):228–242, 1999.

Shweta Bansal, Bryan T Grenfell, and Lauren Ancel Meyers. When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society, Interface*, 4(16):879–891, 2007.

Linlin Bao, Wei Deng, Hong Gao, Chong Xiao, Jiayi Liu, Jing Xue, Qi Lv, Jiangning Liu, Pin Yu, Yanfeng Xu, Feifei Qi, Yajin Qu, Fengdi Li, Zhiguang Xiang, Haisheng Yu, Shuran Gong, Mingya Liu, Guanpeng Wang, Shunyi Wang, Zhiqi Song, Wenjie Zhao, Yunlin Han, Linna Zhao, Xing Liu, Qiang Wei, and Chuan Qin. Reinfection could not occur in SARS-CoV-2 infected rhesus macaques. *bioRxiv*, 2020.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016.

Paolo Barucca and Fabrizio Lillo. Disentangling bipartite and core-periphery structure in financial networks. *Chaos, Solitons & Fractals*, 88:244–253, 2016.

Basel Committee on Banking Supervision. Fundamental review of the trading book. Consultative document, Bank for International Settlements, 2012.

Basel Committee on Banking Supervision. Making supervisory stress tests more macroprudential: Considering liquidity and solvency interactions and systemic risk. Working Paper 29, Bank for International Settlements, November 2015.

Stefano Battiston, Domenico Delli Gatti, Mauro Gallegati, Bruce Greenwald, and Joseph Stiglitz. Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of Economic Dynamics and Control*, 36(8):1121–1141, 2012a.

Stefano Battiston, Michelangelo Puliga, Rahul Kaushik, Paolo Tasca, and Guido Caldarelli. DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk. *Scientific Reports*, 2(541), 2012b.

Patrizia Baudino, Roland Goetschmann, Jérôme Henry, Ken Taniguchi, and Weisha Zhu. Stress-testing banks – a comparative analysis. Financial Stability Institute Insights on policy implementation No 12, Bank for International Settlements, November 2018.

Guillaume Béraud, Sabine Kazmercziak, Philippe Beutels, Daniel Levy-Bruhl, Xavier Lenne, Nathalie Mielcarek, Yazdan Yazdanpanah, Pierre-Yves Boëlle, Niel Hens, and Benoit Dervaux. The French connection: the first large population-based contact survey in france relevant for the spread of infectious diseases. *PLoS one*, 10(7), 2015.

Ben Bernanke. *The Federal Reserve and the financial crisis*. Princeton University Press, 2013.

David Bernoulli. Essai d'une nouvelle analyse de la mortalite causee par la petite verole. *Mem. Math. Phys. Acad. Roy. Sci., Paris*, 1766.

Monica Billio, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104:535–559, 2011.

Ulrich Bindseil, Marco Corsi, Benjamin Sahel, and Ad Visser. The Eurosystem collateral framework explained. European Central Bank Occasional Paper, 2017.

John R Birge, Ozan Candogan, and Yiding Feng. Controlling epidemic spread: Reducing economic losses with targeted closures. Working Paper 2020-57, University of Chicago, 2020.

Kristian Blickle, Markus K Brunnermeier, and Stephan Luck. Micro-Evidence From a System-Wide Financial Meltdown: The German Crisis of 1931. *SSRN Electronic Journal*, 2019.

Frederic Boissay and Phurichai Rungcharoenkitkul. Macroeconomic effects of Covid-19: an early review. Bulletin No. 7, Bank for International Settlements, 2020.

Eric Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99:7280–7287, 2002.

Michael Boss, Helmut Elsinger, Martin Summer, and Stefan Thurner. Network topology of the interbank market. *Quantitative Finance*, 4(6):677–684, 2004.

Fred Brauer. Compartmental Models in Epidemiology. In Fred Brauer, Pauline van den Driessche, and Jianhong Wu, editors, *Mathematical Epidemiology*, pages 19–80. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

Fred Brauer and Carlos Castillo-Chavez. *Mathematical Models for Communicable Diseases*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2012.

Damiano Brigo, Andrea Pallavicini, and Roberto Torresetti. *Credit Models and the Crisis: A Journey into CDOs, Copulas, Correlations and Dynamic Models*. The Wiley Finance Series. Wiley, 2010.

British Academy. The COVID decade: Understanding the long-term societal impacts of COVID-19. Policy report, The British Academy, London, 2021.

Tom Britton, Etienne Pardoux, Franck Ball, Catherine Laredo, David Sirl, and Viet Chí Tran. *Stochastic epidemic models with inference*. Springer, 2019.

Tom Britton, Frank Ball, and Trapman Pieter. A Mathematical Model Reveals the Influence of Population Heterogeneity on Herd Immunity to SARS-CoV-2. *Science*, 369(6505):846–849, 2020.

Christian T. Brownlees and Robert F. Engle. SRISK: A Conditional Capital Shortfall Measure of Systemic Risk. *The Review of Financial Studies*, 30(1):48–79, 2017.

Diana Buitrago-Garcia, Michel Counotte, Dianne Egli-Gany, Stefanie Hossmann, Hira Imeri, and Nicola Low. The role of asymptomatic SARS-CoV-2 infections: A rapid systematic review, April 2020.

Nicole Bäuerle and Alexander Glauner. Optimal risk allocation in reinsurance networks. *Insurance: Mathematics and Economics*, 82:37–47, 2018.

Fabio Caccioli, Thomas A. Catanach, and J. Doyne Farmer. Heterogeneity, correlations and financial contagion. *Advances in Complex Systems*, 15(2):1250058, 2012.

Fabio Caccioli, J. Doyne Farmer, Nick Foti, and Daniel Rockmore. Overlapping portfolios, contagion, and financial stability. *Journal of Economic Dynamics and Control*, 51:50–63, 2015.

Fabio Caccioli, Paolo Barucca, and Teruyoshi Kobayashi. Network models of financial systemic risk: a review. *Journal of Computational Social Science*, 1: 81–114, 2018.

Zhidong Cao, Qingpeng Zhang, Xin Lu, Dirk Pfeiffer, Zhongwei Jia, Hongbing Song, and Daniel Dajun Zeng. Estimating the effective reproduction number of the 2019-nCoV in China. *medRxiv*, 2020.

Hans Carlsson and Eric van Damme. Global Games and Equilibrium Selection. *Econometrica*, 61(5):989–1018, 1993.

Fabio Castiglionesi and Noemi Navarro. Optimal Fragile Financial Networks. Discussion Paper Vol. 2007-100, CentER, 2007.

Stephen Cecchetti and Anil Kashyap. What Binds? Interactions between Bank Capital and Liquidity Regulations. *The Changing Fortunes of Central Banking*, pages 192–202, 2018.

Hua Chen, J. David Cummins, Tao Sun, and Mary A. Weiss. The Reinsurance Network Among U.S. Property–Casualty Insurers: Microstructure, Insolvency Risk, and Contagion. *Journal of Risk and Insurance*, 87(2):253–284, 2020.

Maria Chikina and Wesley Pegden. Modeling strict age-targeted mitigation strategies for COVID-19. *PLoS ONE*, 15(7), 2020.

Gerardo Chowell, Lisa Sattenspiel, Shweta Bansal, and Cécile Viboud. Mathematical models to characterize early epidemic growth: A Review. *Physics of Life Reviews*, 18:66–97, 2016.

Aaron Clauset, Cosma R. Shalizi, and Mark E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

David Clayton and Michael Hills. *Statistical Models in Epidemiology.* Oxford University Press, Incorporated, 2013.

Laurent Clerc, Alberto Giovannini, Sam Langfield, Tuomas Peltonen, Richard Portes, and Martin Scheicher. Indirect contagion: the policy problem. Occasional Paper Series No. 9, European Systemic Risk Board, 2016.

Rama Cont. Central clearing and risk transformation. *Financial Stability Review*, pages 127–140, April 2017.

Rama Cont and Eric Schaanning. Fire sales, indirect contagion and systemic stress-testing. Working paper, Norges Bank, 2017.

Rama Cont and Emily Tanimura. Small-world graphs: Characterization and alternative constructions. *Advances in Applied Probability*, 40(4):939–965, December 2008.

Rama Cont and Lakshithe Wagalath. Fire Sales Forensics: Measuring endogenous risk. *Mathematical Finance*, 26:835–866, 2016.

Rama Cont, Amal Moussa, and Edson B Santos. Network structure and systemic risk in banking systems. In Jean-Pierre Fouque & Joseph A. Langsam, editor, *Handbook of Systemic Risk*, pages 327–368. Cambridge University Press, July 2013.

Sander Conzen. *Liquidity at Risk (LaR) und LiquidityValue at Risk (LVaR): Zwei neue Ansätze für das Liquiditätsmanagement.* Deplomica Verlag, GmbH, 2009.

Marcia Millon Cornett, Jamie John McNutt, Philip E Strahan, and Hassan Tehranian. Liquidity risk management and credit supply in the financial crisis. *Journal of Financial Economics*, 101(2):297–312, 2011.

Ben Craig and Goetz von Peter. Interbank tiering and money center banks. *Journal of Financial Intermediation*, 23:322–347, 2014.

J. David Cummins and Mary A. Weiss. Systemic Risk and the U.S. Insurance Sector. *Journal of Risk and Insurance*, 81(3):489–528, 2014.

David M. Cutler and Lawrence H. Summers. The COVID-19 Pandemic and the $16 Trillion Virus. *JAMA*, 324(15):1495–1496, 2020.

Leon Danon, Ellen Brooks-Pollock, Mick Bailey, and Matt J Keeling. A spatial model of CoVID-19 transmission in England and Wales: early spread and peak timing. *medRxiv*, 2020.

Nicholas G. Davies, Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, Rosalind M. Eggo, and CMMID COVID-19 working group. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med*, 2020.

Matt Davison, Darrell Leadbetter, Bin Lu, and Jane Voll. Are Counterparty Arrangements in Reinsurance a Threat to Financial Stability? Staff Working Papers 39, Bank of Canada, 2016.

Oliver De Brandt and Phillip Hartmann. Systemic risk: a survey. Working Paper No. 35, European Central Bank, 2000.

Hans Degryse and Gregory Nguyen. Interbank Exposures: An Empirical Examination of Systemic Risk in the Belgian Banking System. Working Paper No. 43, National Bank of Belgium, 2004.

Kieran Dent, Ben Westwood, and Miguel Segoviano. Stress testing of banks: an introduction. Quarterly Bulletin Q3, Bank of England, 2016.

Douglas W Diamond and Philip H Dybvig. Bank runs, deposit insurance, and liquidity. *The journal of political economy*, pages 401–419, 1983.

Douglas W. Diamond and Raghuram G. Rajan. Liquidity Shortages and Banking Crises. *The Journal of Finance*, 60(2):615–647, 2005.

Odo Diekmann, J.A.P. Heesterbeek, and J.A.J. Metz. On the definition and the computation of the basic reproduction ratio $\mathcal{R}_0$ in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28:365–382, 1950.

Reinhard Diestel. *Graph Theory*. Springer-Verlag Berlin Heidelberg, Germany, 2017.

Claire Donnat and Susan Holmes. Modeling the heterogeneity in COVID-19 reproductive number and its impact on predictive scenarios. *arXiv:2004.05272*, 2020.

Ilaria Dorigatti, Lucy Okell, Anne Cori, Natsuko Imai, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Z Cucunubá, G Cuomo-Dannenburg, R FitzJohn, et al. Severity of 2019-novel coronavirus (ncov). Technical report, Imperial College London, 2020.

Wenxin Du, Salil Gadgil, Michael B Gordy, and Clara Vega. Counterparty risk and counterparty choice in the credit default swap market. Working paper, Federal Reserve Board, 2019.

Darrell Duffie. *How Big Banks Fail and What to Do about It*. Princeton University Press, 2010.

EIOPA. 2016 EIOPA Insurance Stress Test Report. Report EIOPA 16/302, European Insurance and Occupational Pensions Authority, 2016.

EIOPA. Insurance Stress Test 2018. Technical specifications EIOPA-BoS-18-189, European Insurance and Occupational Pensions Authority, 2018.

EIOPA. Second Discussion Paper on Methodological Principles of Insurance Stress Testing. Discussion Paper EIOPA-BoS-20/341, European Insurance and Occupational Pensions Authority, 2020.

Larry Eisenberg and Thomas Noe. Systemic Risk in Financial Systems. *Management Science*, 47(2):236–249, 2001.

Abdulrahman M. El-Sayed, Peter Scarborough, Lars Seemann, and Sandro Galea. Social network analysis and agent-based modeling in social epidemiology. *Epidemiologic Perspectives & Innovations*, 9(1):1–9, 2012.

Matthew Elliot, Benjamin Golub, and Matthew O. Jackson. Financial networks and contagion. *American Economic Review*, 104(10):3115–3153, 2014.

Andrew Ellul, Chotibhak Jotikasthira, Christian T. Lundblad, and Yihui Wang. Is Historical Cost Accounting a Panacea? Market Stress, Incentive Distortions, and Gains Trading. *Journal of Finance*, 70(6):2489–2538, 2015.

Helmut Elsinger, Alfred Lehar, and Summer Martin. Risk Assessment for Banking Systems. *Management Science*, 52(9):1301–1314, 2006.

European Banking Authority. Guidelines on the revised common procedures and methodologies for the supervisory review and evaluation process (SREP) and supervisory stress testing. Guidelines EBA/GL/2014/13, European Banking Authority, 2018.

European Central Bank. Financial stability review. European Central Bank, December, 2009a.

European Central Bank. The concept of systemic risk. Financial stability review, European Central Bank, 2009b.

European Central Bank. ECB Sensitivity analysis of Liquidity Risk - Stress Test 2019. Methodological note, Banking Supervision, February 2019.

European Court of Auditors. EU-wide stress tests for banks: unparalleled amount of information on banks provided but greater coordination and focus on risks needed. Special Report No. 10, European Court of Auditors, 2019.

Eurostat. Population on 1 January by age group, sex and NUTS 3 region [computer file]. European Commission. Downloaded from: `https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_r_pjangrp3&lang=en`, January 2020a. Table demo_r_pjangrp3.

Eurostat. NUTS - Nomenclature Of Territorial Units For Statistics. European Commission. Available at: `https://ec.europa.eu/eurostat/web/nuts/background`, January 2020b.

Rudiger Fahlenbrach, Robert Prilmeier, and Rene M. Stulz. This Time Is the Same: Using Bank Performance in 1998 to Explain Bank Performance during the Recent Financial Crisis. *Journal of Finance*, 67(6):2139–2185, 2012.

Marc Farag, Damian Harland, and Dan Nixon. Bank capital and liquidity. *Bank of England Quarterly Bulletin*, September 2013.

Christine Farquharson, Imran Rasul, and Luke Sibieta. Key workers: key facts and questions. IFS Observation, Institute for Fiscal Studies, March 2020.

Federal Reserve. August 2019 Senior Financial Officer Survey. Board of Governors of the Federal Reserve System, August 2019.

Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunuba-Perez, Gina Cuomo-Dannenburg, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Report 9, Imperial College COVID-19 Response Team, 2020.

Neil M. Ferguson, Derek A. T. Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sopon Iamsirithaworn, and Donald S. Burke. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437(7056):209–214, 2005.

Jose Fique. *The MacroFinancial Risk Assessment Framework (MFRAF), Version 2.0.* Technical Report No. 111, Bank of Canada, 2017.

Seth Flaxman, Swapnil Mishra, Axel Gandy, et al. Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. Technical Report 30, Imperial College COVID-19 Response Team, 2020.

Andrea French, Mathieu Vital, and Dean Minot. Insurance and Financial Stability. *Bank of England Quarterly Bulletin*, Q3 2015.

Daniel Fricke and Thomas Lux. Core–Periphery Structure in the Overnight Money Market: Evidence from the e-MID Trading Platform. *Computational Economics*, 45(3):359–395, 2015.

Craig H. Furfine. Interbank exposures: Quantifying the risk of contagion. *Journal of Money, Credit and Banking*, 35:111–128, 2003.

Prasanna Gai. The Robust-Yet-Fragile Nature of Financial Systems. In Prasanna Gai, editor, *Systemic Risk: The Dynamics of Modern Financial Systems*, pages 9–28. Oxford University Press, Oxford, 2013.

Prasanna Gai and Sujit Kapadia. Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466 (2120):2401–2423, 2010.

Douglas Gale. Regulation and sausages. *The Manchester School*, 83(S2):1–26, 2015.

Timothy F. Geithner. *Stress Test: Reflections on Financial Crises*. Random House Business Books, London, United Kingdom, 2014.

Geneva Association. Systemic Risk in Insurance: An Analysis of Insurance and Financial Stability. Special Report of The Geneva Association Systemic Risk Working Group, The Geneva Association, March 2010.

Itay Goldstein and Ady Pauzner. Demand–Deposit Contracts and the Probability of Bank Runs. *The Journal of Finance*, 60(3):1293–1327, 2005.

Gary Gorton. Bank suspension of convertibility. *Journal of monetary Economics*, 15(2):177–193, 1985.

Gary Gorton. Banking panics and business cycles. *Oxford economic papers*, 40(4): 751–781, 1988.

Gary Gorton. *Misunderstanding financial crises*. Oxford University Press, 2012.

Gary Gorton and Andrew Metrick. Securitized banking and the run on repo. *Journal of Financial economics*, 104(3):425–451, 2012.

Christian Gourieroux, Alain Monfort, and Eric Renault. Indirect inference. *Journal of applied econometrics*, 8(1):85–118, 1993.

Andrew G. Haldane and Robert M. May. Systemic risk in banking ecosystems. *Nature*, 469:351–355, 2011.

Laurent Hebert-Dufresne, Benjamin M. Althouse, Samuel V. Scarpino, and Antoine Allard. Beyond R_0: Heterogeneity in Secondary Infections and Probabilistic Epidemic Forecasting. *Journal of the Royal Society Interface*, 17, 2020.

Martin F. Hellwig. Systemic risk in the financial sector: An analysis of the subprime mortgage financial crisis. *De Economist*, 157(2), 2009.

IAIS. Insurance and Financial Stability. Technical report, International Association of Insurance Supervisors, 2011.

IAIS. Reinsurance and Financial Stability. Technical report, International Association of Insurance Supervisors, 2012.

IAIS. Systemic risk from insurance product features. Technical report, International Association of Insurance Supervisors, 2016.

Carmela Iazzetta and Michele Manna. The topology of the interbank market: developments in Italy since 1990. Temi di Discussione (Working Paper) No. 711, Bank of Italy, Economic Research and International Relations Area, May 2009.

Bjorn Imbierowicz and Christian Rauch. The relationship between liquidity risk and credit risk in banks. *Journal of Banking & Finance*, 40:242 – 256, 2014.

International Monetary Fund. Chapter 3. The Insurance Sector – Trends and Systemic Risk Implications. In *Global Financial Stability Report: Potent Policies for a Successful Normalization*. International Monetary Fund, April 2016.

International Monetary Fund. Switzerland: Financial Sector Assessment Program; Technical Note-Stress Testing the Banking Sector. Country Report No. 19/189, International Monetary Fund, 2019.

Valerie Isham. Assessing the variability of stochastic epidemics. *Mathematical biosciences*, 107(2):209–224, 1991.

Matthew O. Jackson. *Social and Economic Networks*. Princeton University Press, Princeton, NJ, 2008.

Andreas A. Jobst, Nobuyasu Sugimoto, and Timo Broszeit. Macroprudential Solvency Stress Testing of the Insurance Sector. Working Paper WP/14/133, International Monetary Fund, 2014.

Thibaut Jombart, Kevin Van Zandvoort, Timothy W Russell, Christopher I Jarvis, Amy Gimma, Sam Abbott, Sam Clifford, Sebastian Funk, Hamish Gibbs, Yang Liu, et al. Inferring the number of COVID-19 cases from recently reported deaths. *Wellcome Open Research*, 5, 2020.

Masayasu Kanno. The network structure and systemic risk in the global non-life insurance market. *Insurance: Mathematics and Economics*, 67(C):38–53, 2016.

Sujit Kapadia, Matthias Drehmann, John Elliott, and Gabriel Sterne. Liquidity risk, cash flow constraints, and systemic feedbacks. In *Quantifying Systemic Risk*, pages 29–61. University of Chicago Press, 2012.

Matt J Keeling and Ken T.D Eames. Networks and epidemic models. *Journal of The Royal Society Interface*, 2(4):295–307, 2005.

William O. Kermack and Anderson G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A*, 115:700–721, 1927.

William O. Kermack and Anderson G. McKendrick. Contributions to the mathematical theory of epidemics. II. – The problem of endemicity. *Proceedings of the Royal Society of London A*, 138:55–83, 1932.

William O. Kermack and Anderson G. McKendrick. Contributions to the mathematical theory of epidemics. III. – Further studies of the problem of endemicity. *Proceedings of the Royal Society of London A*, 141:94–122, 1933.

Malahat Khalili, Mohammad Karamouzian, Naser Nasiri, Sara Javadi, Ali Mirzazadeh, and Hamid Sharifi. Epidemiological characteristics of COVID-19: A systemic review and meta-analysis. *medRxiv*, 2020.

Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3): 203–271, September 2014.

Ariah Klages-Mundt and Andreea Minca. Cascading Losses in Reinsurance Networks. *Management Science*, 66(9):4246–4268, September 2020.

Petra Klepac, Adam J Kucharski, Andrew JK Conlan, Stephen Kissler, Maria Tang, Hannah Fry, and Julia R Gog. Contacts in context: large-scale setting-specific social mixing matrices from the BBC Pandemic project. *medRxiv*, 2020.

Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo, Fiona Sun, Mark Jit, James D Munday, Nicholas Davies, Amy Gimma, Kevin van Zandvoort, Hamish Gibbs, Joel Hellewell, Christopher I Jarvis, Sam Clifford, Billy J Quilty, Nikos I Bosse, Sam Abbott, Petra Klepac, and Stefan Flasche. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(5):553–558, 2020.

Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172 (9):577–582, 2020.

Iman Lelyveld, Franka Liedorp, and Manuel Kampman. An empirical assessment of reinsurance risk. *Journal of Financial Stability*, 7(4):191–203, 2011.

Jeffrey Levine and Asani Sarkar. How Do Large Banks Manage Their Cash? Liberty Street Economics, Federal Reserve Bank of New York, 2019.

Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382:1199–1207, 2020.

Gechun Liang, Eva Lütkebohmert, and Yajun Xiao. A multiperiod bank run model for liquidity risk. *Review of Finance*, pages 1–40, 2013.

Natalie M Linton, Tetsuro Kobayashi, Yichi Yang, Katsuma Hayashi, Andrei R Akhmetzhanov, Sung-mok Jung, Baoyin Yuan, Ryo Kinoshita, and Hiroshi Nishiura. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of clinical medicine*, 9(2):538, 2020.

Marc Lipsitch, Christl A. Donnelly, Christophe Fraser, Isobel M. Blake, Anne Cori, Ilaria Dorigatti, Neil M. Ferguson, Tini Garske, Harriet L. Mills, Steven Riley, Maria D. Van Kerkhove, and Miguel A. Hernan. Potential Biases in Estimating Absolute and Relative Case-Fatality Risks during Outbreaks. *PLOS Neglected Tropical Diseases*, 9(7):1–16, July 2015.

Alex Lipton and Marcos Lopez de Prado. Mitigation strategies for COVID-19: Lessons from the K-SEIR model. *SSRN Electronic Journal*, 2020.

Alun L. Lloyd and Vincent A. A. Jansen. Spatiotemporal dynamics of epidemics: synchrony in metapopulation models. *Mathematical Biosciences*, 188(1):1–16, 2004.

Alun L. Lloyd and Robert M. May. Spatial heterogeneity in epidemic models. *Journal of Theoretical Biology*, 179(1):1–11, 1996.

Jose Lourenco, Robert Paton, Mahan Ghafari, Moritz Kraemer, Craig Thompson, Peter Simmonds, Paul Klenerman, and Sunetra Gupta. Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. *MedRxiv*, 2020.

Sheheryar Malik and TengTeng Xu. Interconnectedness of Global Systemically-Important Banks and Insurers. Working Paper No. 17/210, International Monetary Fund, September 2017.

Robert McDonald and Anna Paulson. AIG in Hindsight. *Journal of Economic Perspectives*, 29(2):81–106, 2015.

Robert C Merton. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, 29(2):449–70, May 1974.

Kenji Mizumoto, Katsushi Kagaya, Alexander Zarebski, and Gerardo Chowell. Estimating the asymptomatic proportion of Coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance*, 25(10):2000180, 2020.

Stephen Morris and Hyun Song Shin. Unique equilibrium in a model of self-fulfilling currency attacks. *American Economic Review*, pages 587–597, 1998.

Stephen Morris and Hyun Song Shin. *Global Games: Theory and Applications*, volume 1 of *Econometric Society Monographs*, pages 56–114. Cambridge University Press, Cambridge, United Kingdom, 2003.

Stephen Morris and Hyun Song Shin. Financial regulation in a system context. *Brookings papers on economic activity*, 2008(2):229–274, 2008.

Stephen Morris and Hyun Song Shin. Illiquidity component of credit risk. *International Economic Review*, 57(4):1135–1148, 2016.

Joel Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, 5(3):1–1, March 2008.

Joel Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds. POLYMOD social contact data, 2017. Version 1.1.

Munich RE. Reinsurance: A Basic Guide to Facultative and Treaty Reinsurance, 2010.

Mark Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167–256, January 2003.

Mark Newman. *Networks: An Introduction*. Oxford University Press, Oxford, United Kingdom, 2010.

Erlend Nier, Jing Yang, Tanju Yorulmazer, and Amadeo Alentorn. Network models and financial stability. *Journal of Economic Dynamics and Control*, 31(6):2033–2060, 2007.

Hiroshi Nishiura, Tetsuro Kobayashi, Ayako Suzuki, Sung-Mok Jung, Katsuma Hayashi, Ryo Kinoshita, Yichi Yang, Baoyin Yuan, Andrei R. Akhmetzhanov, Natalie M. Linton, and Takeshi Miyama. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *International Journal of Infectious Diseases*, 2020.

Office for National Statistics. Coronavirus (COVID-19) infections in the community in England: July 2020. Downloaded from: `https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsinthecommunityinengland/july2020`, January 2020a.

Office for National Statistics. 2011 Census: Special Workplace Statistics (United Kingdom) [computer file]. UK Data Service Census Support. Downloaded from: `https://wicid.ukdataservice.ac.uk`, January 2020b. Table WU01EW.

Office for National Statistics. LAU2 to LAU1 to NUTS3 to NUTS2 to NUTS1 (January 2018) Lookup in United Kingdom v3 [computer file]. Downloaded from: `https://geoportal.statistics.gov.uk/datasets/c893dfece45f465f857ac34641041863_0`, January 2020c.

Sandra Y. Okita. Social interactions and learning. In Norbert M. Seel, editor, *Encyclopedia of the Sciences of Learning*, pages 3104–3107. Springer US, Boston, MA, 2012.

Sojung Carol Park and Xiaoying Xie. Reinsurance and systemic risk: The impact of reinsurer downgrading on property-casualty insurers. *Journal of Risk and Insurance*, 81(3):587–622, 2014.

Diane Pierret. Systemic Risk and the Solvency-Liquidity Nexus of Banks. *International Journal of Central Banking*, 11(3):193–227, 2015.

Robert S Pindyck. Covid-19 and the welfare effects of reducing contagion. Technical report, National Bureau of Economic Research, 2020.

Michael Pohl. Basel III liquidity risk monitoring tools. Occasional paper no. 14, Bank for International Settlements, October 2017.

Natacha Postel-Vinay. What Caused Chicago Bank Failures in the Great Depression? A Look at the 1920s. *The Journal of Economic History*, 76(2):478–519, 2016.

Kiesha Prem, Alex R. Cook, and Mark Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS computational biology*, 13(9):e1005697, 2017.

Kiesha Prem, Yang Liu, Timothy W. Russell, Adam J. Kucharski, Rosalind M Eggo, Nicholas Davies, Stefan Flasche, Samuel Clifford, Carl A. B. Pearson, James D. Munday, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*, 5(5):261–270, May 2020.

Public Health England and NHSX. GOV.UK Coronavirus (COVID-19) in the UK. Downloaded from: `https://coronavirus.data.gov.uk`, June 2020.

Thomas Rawson, Tom Brewer, Dessislava Veltcheva, Chris Huntingford, and Michael B. Bonsall. How and When to End the COVID-19 Lockdown: An Optimization Approach. *Frontiers in Public Health*, 8:262, 2020.

Jean-Charles Rochet and Xavier Vives. Coordination failures and the lender of last resort: was Bagehot right after all? *Journal of the European Economic Association*, 2(6):1116–1147, 2004.

Joacim Rocklöv, Henrik Sjödin, and Annelies Wilder-Smith. COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures. *Journal of travel medicine*, 2020.

Dani Rodrik and Andres Velasco. Short-term capital flows. Working Paper 7364, National Bureau of Economic Research, 1999.

L. C. G. Rogers and Luitgard A. M. Veraart. Failure and Rescue in an Interbank Network. *Management Science*, 59(4):882–898, April 2013.

Lionel Roques, Olivier Bonnefon, Virgile Baudrot, Samuel Soubeyrand, and Henri Berestycki. A parsimonious model for spatial transmission and heterogeneity in the COVID-19 propagation. *Royal Society Open Science*, 7(12), 2020a.

Lionel Roques, Etienne K. Klein, Julien Papaix, Antoine Sar, and Samuel Soubeyrand. Impact of lockdown on the epidemic dynamics of COVID-19 in France. *Frontiers in Medicine*, 7:274, 2020b.

Tarik Roukny, Hugues Bersini, Hugues Pirotte, Guido Caldarelli, and Stefano Battiston. Default Cascades in Complex Networks: Topology and Systemic Risk. *Scientific Reports*, 3(2759), 2013.

Robert Rowthorn and Jan Maciejowski. A cost-benefit analysis of the Covid-19 disease. *Oxford Review of Economic Policy*, 2020.

Henrik Salje, Cécile Tran Kiem, Noémie Lefrancq, Noémie Courtejoie, Paolo Bosetti, Juliette Paireau, Alessio Andronico, Nathanaël Hozé, Jehanne Richet, Claire-Lise Dubost, Yann Le Strat, Justin Lessler, Daniel Levy-Bruhl, Arnaud Fontanet, Lulla Opatowski, Pierre-Yves Boelle, and Simon Cauchemez. Estimating the burden of SARS-CoV-2 in France. *Science*, 369(6500):208–211, 2020.

Gillian M. Sandstrom and Elizabeth W. Dunn. Social Interactions and Well-Being: The Surprising Power of Weak Ties. *Personality and Social Psychology Bulletin*, 40:910–922, 2014.

Clodomir Santana, Federico Botta, Hugo Barbosa, Filippo Privitera, Ronaldo Menezes, and Riccardo Di Clemente. Analysis of human mobility in the UK during the COVID-19 pandemic, 2020.

Vicky Saporta. Macroprudential Policy for Insurers. Speech at the association of british insurers annual conference, Bank of England, 2016.

Christian Schmieder, Heiko Hesse, Benjamin Neudorfer, Claus Puhr, and Stefan W. Schmitz. Next Generation System-Wide Liquidity Stress Testing. Working Paper 12, International Monetary Fund, 2012.

Stefan W. Schmitz, Michael Sigmund, and Laura Valderrama. The interaction between bank solvency and funding costs: A crucial effect in stress tests. *Economic Notes*, (in press), 2019.

Til Schuermann. Stress testing banks. *International Journal of Forecasting*, 30(3): 717 – 728, 2014.

Hyun Song Shin. Reflections on Northern Rock: The Bank Run that Heralded the Global Financial Crisis. *Journal of Economic Perspective*, 23(1):101–119, 2009.

Andrei Shleifer and Robert Vishny. Liquidation values and debt capacity: A market equilibrium approach. *The Journal of Finance*, 47(4):1343–1366, 1992.

Andrei Shleifer and Robert Vishny. Fire Sales in Finance and Macroeconomics. *Journal of Economic Perspectives*, 25(1):29–48, 2010.

Constantinos I. Siettos and Lucia Russo. Mathematical modeling of infectious disease dynamics. *Virulence*, 4(4):295–306, 2013.

Rajesh Singh and Ronojoy Adhikari. Age-structured impact of social distancing on the COVID-19 epidemic in India. *arXiv:2003.12055*, 2020.

Robin Swain and David Swallow. The prudential regulation of insurers under Solvency II. Quarterly Bulletin Q2, Bank of England, 2015.

Loring J. Thomas, Peng Huang, Fan Yin, Xiaoshuang Iris Luo, Zack W. Almquist, John R. Hipp, and Carter T. Butts. Spatial Heterogeneity Can Lead to Substantial Local Variations in COVID-19 Timing and Severity. *Proceedings of the National Academy of Sciences of the United States of America*, 117(39):24180–24187, 2020.

Yannick Timmer. Cyclical investment behavior across financial institutions. *Journal of Financial Economics*, 129(2):268–286, 2018.

Christian Upper. Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability*, 7(3):111–125, 2011.

Robert Verity, Lucy C. Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick Walker, Han Fu, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 20(6):669–677, 2020.

Patrick G. T. Walker, Charles Whittaker, Oliver Watson, Marc Baguelin, Kylie E C Ainslie, Sangeeta Bhatia, Samir Bhatt, Adhiratha Boonyasiri, Olivia Boyd, Lorenzo Cattarino, et al. The Global Impact of COVID-19 and Strategies for Mitigation and Suppression. Report, Imperial College COVID-19 Response Team, 2020.

Duncan J. Watts. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press, Princeton, 1999.

Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

Timothy L. Wiemken and Robert R. Kelley. Machine learning in epidemiology and health outcomes research. *Annual Review of Public Health*, 41(1):21–36, 2020.