Multi-agent, non-equilibrium, and network models of labour economics and financial contagion



R. Maria del Rio-Chanona

Balliol College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2021

Abstract

The financial crisis of 2009 showed that the economy is not in equilibrium. That is, the economy is in a state where agents have incentives to modify their chosen actions. The recent COVID-19 pandemic has shown that the economy changes rapidly and that reaching an equilibrium can take years, if one is reached at all. However, assuming equilibrium has long formed the basis of many economic models. Although a non-equilibrium approach may seem better suited for answering economic policy questions during times of rapid change, previously, the lack of data and computing power made it difficult to build these type of models. Today, these factors are less constraining.

This thesis contributes to the development of non-equilibrium economic models that can provide accurate answers to policy-relevant questions efficiently. The models we develop have the potential to enable fast economic analysis and can be run in near real-time.

In the first part of this thesis, we focus on the labour market. We develop a model to analyse how workers move through an empirically derived occupational mobility network in response to labour demand shocks. We show that we can approximate the behavior of the agent-based model with a discrete dynamical system. At the macro level, our model reproduces the Beveridge curve's dynamics, i.e. the negative relationship between vacancy and unemployment rate. At the occupation level, we study the potential impact of automation and the COVID-19 pandemic on employment. We find that the network structure plays an important role in determining unemployment levels, with occupations in particular areas of the network having few job transition opportunities.

In the second part, we use multiplex networks to study global financial contagion. This work focuses on identifying systemically important countries, i.e. countries whose financial distress can propagate to other countries. We build a global financial multiplex network and develop a multiplex threshold model of financial contagion. We find that the number of systemically important countries can be twice as large when we consider a multiplex network structure compared to a monoplex structure.

These two models have the potential to be used for policy design. Arguably, support and skill development programs should be directed towards workers who are likely to face longer periods of unemployment. Our model can help policymakers identify who these workers are. Our multiplex model of financial contagion represents a first step towards developing multiplex network models which policy institutions could use to prevent the spread of financial contagion.

Acknowledgements

The work presented in this thesis was possible thanks to the guidance, encouragement, and support from several brilliant people. Among them, I would like to first thank my supervisors:

Doyne Farmer, for sharing with me your passion for answering questions that *matter*; fostering, within the INET complexity group, a friendly and fun environment of academic freedom and excellence; and showing me a human side to research. Thank you for your trust and life teachings. Your support and encouragement in crucial times made all the difference.

François Lafond, for your support and guidance from start to end; and for your patience during countless discussions. I learned academic rigor and responsibility from you. Most of all, thank you for believing a young physicist could do economics and then for teaching me economics.

Mariano Beguerisse-Díaz, thank you for your friendly supervision and for your advice in academic as well as professional-life choices.

I would also like to thank my co-authors:

Penny Mealy, it's been an absolute joy to do research with you. Your passion, brilliance, and eloquence for discussing research were a constant source of inspiration and learning. Thank you for all the fun, for your encouragement during the good, the stressful, and exciting times, and, of course, for your appreciation of Mexican tequila.

Anton Pichler, for the reminders that "we should not work too hard" and company when doing so. Your dedicated and diligent work, as well as your kindness and friendship, made work so much easier and enjoyable. Collaborating with the INET team during a pandemic has been a thrill!

Yevgeniya Korniyenko and Manasa Patnam, for venturing into networks research and their supervision during my internship at the International Monetary Fund.

Mason A. Porter, for your mentorship; it was an amazing experience to learn from your mathematical rigor.

Special thanks to Marco Pangallo for insightful discussions, help and advice, and friendly tutoring, as well as to Eloise Hamilton, Blas Kolic, Bernardo Perez, and my parents for proof-reading this thesis.

I am deeply thankful to the INET and Balliol community: fellow students, post-docs, and staff, for their valuable feedback and for their help and support during the pandemic and non-pandemic times.

I would also like to acknowledge the National Council for Science and Technology in Mexico (CONACYT), the Mexican Energy Department (SENER), and Baillie Gifford for their generous funding.

. . . . . .

Pursuing a DPhil has been a fantastic experience and I have had the pleasure to share it with wonderful people. I am immensely grateful to all of them.

To my college friends.

Cat lovers: Eloise Hamilton, James Farmer, Helen Hailes, Miles Kempton, Mirri Kummerlin, Janina Nahler, Tobias Schillings, and Daan Wesselink, for the fun trips, bops, walks, jogs, pull-up sessions, and for making college life ever so wholesome.

Fellow Mexicans: Bernardo Perez, Marcela Mendoza, Leila Walker, and Sofia Castello, for the dinners, coffees, heated discussions, reggaeton nights, and for making Oxford more homely.

Yuki Asano, for discussions that change my mind, your intrinsic curiosity towards life, and a unique friendship.

To my INET friends.

Blas Kolic, Penny Mealy, Kieran Murray, Marco Pangallo, Anton Pichler, Andrea Bacilieri, Torsten Heinrich, Linus Mattauch, Luca Mungo, Valentina Semanova, Juan Palomino, Rupert Way, and Julian Winkler, for the drinks, parties, planking, football games, and for making many office and non-office hours fun.

To the Oxford University Association Football Club and the Furies, for the fun training sessions and matches, and for the shared love for football.

To my friends from before DPhil times, who even at a distance have remained present.

Anji Li, for your continued friendship throughout the years.

Facultad de Ciencias (UNAM) friends, with special thanks to Miguel Quetzeri for his unwaivering support during the first years of my DPhil and for being part of many cherished moments of my life.

Finally, but most of all, to my parents and brother. Your unconditional love and support mean much more than I can express in writing. You are the ones that make me happiest.

Contents

Introduction **1**

I A data-driven network model of occupational mobility 6

1 Literature review **7**

1.1 Overview 7
1.2 Structural change and the future of work 7
1.2.1 Automation 8
1.2.2 The COVID-19 pandemic 9
1.3 Labour economics 10
1.3.1 Search theory 10
1.3.2 Cyclical and structural unemployment in the Beveridge curve 12
1.3.3 Networks and labour economics 13
1.4 Dynamical systems, stochastic processes, and networks 14
1.4.1 Stochastic processes 14
1.4.2 Dynamical systems 16
1.4.3 Networks 17

2 A data-driven network model of the labour market **19**

2.1 Introduction 19
2.2 The occupational mobility network 20
2.3 A network model of the labour market. 22
2.4 Deterministic approximation for large populations 28
2.4.1 Mathematical derivation 29
2.4.2 Simulation evidence 36
2.5 Stability: steady-state and fixed-point 38
2.5.1 Fixed-points 40
2.5.2 Parameter effects 45

2.5.3 Steady-state and fixed-points 46
2.6 Concluding remarks 46

3 The Beveridge curve and model calibration 49
3.1 Introduction 49
3.2 The characteristics and dynamics of the Beveridge curve 50
3.3 Calibration 52
3.4 Reproducing the behaviour of the Beveridge curve 55
3.5 Parameter effects on the Beveridge curve 57
3.6 Concluding remarks 60

4 The future of employment 62
4.1 Introduction 62
4.2 Automation and labour reallocation 63
4.2.1 Labour reallocation 64
4.2.2 Automation estimates 65
4.2.3 Formulating a time dependent automation shock 68
4.3 The impact of automation on employment 69
4.3.1 Aggregate level outcomes. 69
4.3.2 Occupation level outcomes. 73
4.3.3 Network structure 79
4.4 Concluding remarks 84

5 The COVID-19 pandemic and the labour market 87
5.1 Introduction 87
5.2 Supply shocks 89
5.2.1 How much work can be performed from home? 90
5.2.2 Which industries are "essential"? 94
5.2.3 Supply shock: non-essential industries unable to work from home 96
5.3 Demand shock 98
5.4 Combining supply and demand shocks 100
5.4.1 Industry-level supply and demand shocks. 100
5.4.2 Occupation-level supply and demand shocks. 101
5.4.3 Aggregate shocks 102
5.4.4 Shocks by wage level 104
5.5 Occupational mobility network effects of the COVID-19 pandemic . . 105
5.5.1 COVID-19 as a labour demand shock 107

5.5.2 Employment outcomes . . . 108
5.6 Validation . . . 110
5.7 Concluding remarks . . . 112

II The multiplex nature of global financial contagions 115

6 Literature review 116
6.1 Overview . . . 116
6.2 Multilayer and multiplex networks . . . 116
6.2.1 Centrality measures . . . 117
6.3 Financial contagion on networks . . . 119
6.3.1 Financial monoplex networks . . . 119
6.3.2 Financial multilayer networks . . . 122

7 Multiplex financial contagion 124
7.1 Introduction . . . 124
7.1.1 A global financial multiplex network . . . 126
7.2 Identifying systemically important countries . . . 127
7.2.1 PageRank and MultiRank . . . 129
7.2.2 Systemic importance of nodes . . . 130
7.3 Results . . . 134
7.3.1 MultiRank evolution . . . 135
7.3.2 Threshold contagion models . . . 138
7.4 Discussion . . . 140
7.5 Concluding remarks . . . 143

Conclusion 145

I Appendix 148
I.1 Occupational mobility and automation . . . 148
I.1.1 Self-loop heterogeneity . . . 148
I.1.2 Results for different values of γ . . . 149
I.1.3 Measuring the impact of automation . . . 150
I.1.4 Automation shocks that change the aggregate demand . . . 153
I.1.5 Automation time and adoption rate . . . 153

II Appendix **156**
II.1 Derivation of total shocks 156
II.1.1 Derivation of supply shocks 156
II.1.2 Industry-specific shocks 157
II.1.3 Occupation-specific shocks. 159
II.1.4 Derivation of demand shocks 160
II.1.5 Total immediate (first-order) shocks 161
II.1.6 Aggregate total shocks 161
II.1.7 Aggregate total shocks with growth of the Health sector . . . 162
II.2 Data 163
II.2.1 Italian list of essential industries 163
II.2.2 Data for occupations 164
II.2.3 Data for industries 166
II.3 Occupations most at risk of contracting SARS-CoV-2 169
II.4 Discussion of labour supply shocks which we do not include 171
II.4.1 Labour supply shocks from mortality and morbidity 171
II.4.2 Labour supply shocks from school closure 172
II.5 Additional estimates of demand or consumption shocks 173
II.6 Shock dampening effects 173

Bibliography **179**

List of Figures

2.1 **The occupational mobility network.** Panel (**A**) shows the occupational mobility network where the colour of nodes indicates the community they belong to. In panel (**B**) the colour of the nodes indicates the classification of the occupation according to the ACS. In both panels, the size of the nodes is proportional to the logarithm of the number of employees in each occupation. 22

2.2 **Flow chart illustrating the possible transitions of workers and job vacancies during a given time step.** *Top* panel: transitions of a worker. *Bottom* panel: The transitions of a job vacancy. Note that vacancies created in the current time step do not accept job applications until the following time step. 25

2.3 **Simulations and approximations at the aggregate level.** In the *top* panel we show the aggregate target labour demand, constant at the beginning and then oscillating according to a *sine* wave. The *middle* panel shows the unemployment rate and the *bottom* panel the long-term unemployment. The dashed vertical red lines denote the time at which the target demand starts oscillating. For the last two panels, green lines denote simulations and dashed black lines the numerical approximation. It is difficult to note the fluctuations in the simulations since we run the model with many agents (roughly 1.4M). To help the reader see the fluctuations we zoom into the first time steps of the model in the top left box of both panels. 37

2.4 **Simulations and approximations at the occupation level.** In the *top left* panel motion picture projectionist, in the *top right* brokerage clerks, in the *bottom left* postal service clerks and in the *bottom right* electricians. The panels show the unemployment and long-term unemployment rates. The green lines denote simulations, the solid brown lines the mean across the ten simulations and dashed black lines the numerical approximation. The solid black lines closely match the brown lines. 39

2.5 **Fixed-point unemployment.** A heatmap showing the fixed-point unemployment rate for different parameters values of δ_u and δ_v. In general, unemployment rate increases with δ_u and decreases with δ_v. . 45

2.6 **Simulations and approximations at the aggregate and occupation level.** In the *top left* panel, the aggregate unemployment rate, in the *top right* brokerage clerks, in the *bottom left* motion picture projectionists and in the *bottom right* electricians. The panels show the unemployment and long-term unemployment rates. The transparent lines denote simulations, the solid lines the mean across the ten simulations and dashed black lines the numerical approximation. Each colour denotes different parameter values. 47

3.1 **The Beveridge curve.** The historical Beveridge curve (unemployment and vacancy rates) for the United States, 2000-2018. Different periods are highlighted with different colours. Red and yellow points correspond to recessions, while blue and purple dots to recovery periods. 51

3.2 **The Beveridge curve.** The Beveridge curve generated by our model. The parameters of the model are calibrated to match the empirical Beveridge curve between December 2007 and December 2018. The dashed lines correspond to the deterministic approximation of Eqs. (2.14 – 2.16) and solid green lines to the full stochastic model simulation of Eqs. (2.3 – 2.5). The transparent grey line shows the empirical Beveridge curve between December 2007 and December 2018. 57

3.3 **The Beveridge curve.** The unemployment and vacancy rate for different parameter values and networks. In particular, we plot the difference between a complete network, with no skill mismatch frictions, vs. the empirical occupational mobility network. Each dot corresponds to the fixed-point unemployment and vacancy rate for different values of δ_u and δ_v. The highlighted points correspond to the unemployment and vacancy rate of the model using the calibrated parameters (see Section 3.3 for details on calibration). 58

3.4 **Beveridge curve dynamics for the occupational mobility network.** The *top left* panel shows the aggregate demand. The grey part corresponds to the first part of the trajectory. The purple/blue part corresponds to the recession period, and the green/yellow part to the recovery period. The following 5 panels show the dynamics of the model's Beveridge curve under different parameter choices. We observe that when $\delta_u > \delta_v$ the curve cycles counter-clockwise, while when $\delta_u \leq \delta_v$ the curve cycles clockwise. We also show the empirical Beveridge curve in magenta for reference. 59

3.5 **Beveridge curve dynamics for the complete network.** The *top left* panel shows the aggregate demand. The grey part corresponds to the first part of the trajectory. The purple/blue part corresponds to the recession period, while the green/yellow part to the recovery period. The following 5 panels show the dynamics of the model's Beveridge curve under different parameter choices. We observe that when $\delta_u \geq \delta_v$ the curve cycles counter-clockwise, while when $\delta_u < \delta_v$ the curve cycles clockwise. We also show the empirical Beveridge curve in magenta for reference. 60

4.1 **Probability of computerization and the occupational mobility network.** Panel (**A**) is a histogram of the probability of computerization for different occupations as estimated by Frey & Osborne (2017), suggesting a bimodal distribution. Panel (**B**) shows the occupational mobility network, where nodes represent occupations and links represent possible worker transitions between occupations. Red nodes have high automatability and blue nodes have low automatability. The size of the nodes indicates the logarithm of the number of employees in each occupation. 66

4.2 **Suitability for machine learning and the occupational mobility network.** Panel (**A**) is a histogram of the suitability for machine learning as estimated by Brynjolfsson et al. (2018). Panel (**B**) shows the occupational mobility network, where nodes represent occupations and links represent possible worker transitions between occupations. The colour of the nodes indicates the suitability for machine learning. Red nodes have higher suitability for machine learning and blue nodes have a low one. The size of the nodes indicates the logarithm of the number of employees in each occupation. 67

4.3 **Aggregate labour market outcomes under automation shocks.** The grey area denotes the 30 years during which the automation shock takes place. Panel (**A**) shows the evolution of the target labour demand for two example occupations. The occupations coloured in blue have a low automation level and the occupations coloured in red have a high level. Because it is more heterogeneous across occupations, the Frey and Osborne shock implies a larger change in the target labour demand of most occupations. Panel (**B**) shows the unemployment rate as a function of time. Dashed lines are our approximations of the expected value (solved numerically) and the solid lines are 10 simulations with 1.4 M agents. Panel (**C**) shows the long-term unemployment rate as a function of time. As before, dashed lines correspond to the deterministic approximation of Eqs. (2.14 – 2.16) and solid lines to the full stochastic model simulation of Eqs. (2.3 – 2.5). 70

4.4 **Aggregate labour market outcomes under the Brynjolfsson et al. shock.** The grey area denotes the 30 years during which the automation shock takes place. Panel (**A**) shows the evolution of the target labour demand for two example occupations. The occupation coloured in blue has a low suitability for machine learning and the occupation coloured in red has a high one. Because the distribution of the suitability for machine learning is more evenly distributed across occupations that the probability of computerization, the Brynjolfsson et al. shock implies a small change in the target labour demand of most occupations. Panel (**B**) shows the unemployment rate as a function of time. Dashed lines are our approximations of the expected value (solved numerically) and the solid lines are 10 simulations with 1.5 M agents. Panel (**C**) shows the long-term unemployment rate as a function of time. As before, dashed lines correspond to the deterministic approximation of Eqs. (2.14 – 2.16) and solid lines to the full stochastic model simulation of Eqs. (2.3 – 2.5). 71

4.5 **Randomised and assortative versions of the Probability of Computerization shock.** In the *top* panel, the unemployment rate. In the *bottom* panel, the long-term unemployment rate. The dashed black line corresponds to the original shock, the solid magenta lines to randomised versions of the shock, and the blue line to the alternative assortative shock. 74

4.6 **Impact of the Frey and Osborne shock on unemployment and long-term unemployment at the occupation level.** The green dots are for the occupational mobility network and the orange dots are for the complete network. The size of the green dots is proportional to the employment of the occupation they represent. Panel **(A)** shows the percentage change in the unemployment rate vs the automation level for each occupation, while panel **(B)** shows the same thing for the long-term unemployment rate. The scatter in the results demonstrates that, due to network effects, the automation level only partially explains occupational unemployment. 76

4.7 **Unemployment rate at the occupation level simulations and numerical solution.** We compare the unemployment and long-term unemployment from the average of the simulations (solid line) and the numerical solution (dashed line). We also show in transparent lines the 10 simulations. Each simulation uses 1.4 Million agents and we average over 10 simulations. 77

4.8 **Network effects at the occupation level.** We show the network effects for each occupation i.e. the difference between the green and the black dots in Fig. 4.6. If this difference is positive, network effects are detrimental for the occupation. Panel **(A)** Network effects on percentage change in unemployment rate vs. median wage. Panel **(B)** Network effects on percentage change in long-term unemployment rate vs. median wage. 78

4.9 **Impact of the Brynjolfsson et al. shock on unemployment and long-term unemployment at the occupation level.** The green dots are for the occupational mobility network and the orange dots are for the complete network. The size of the green dots is proportional to the employment of the occupation they represent. Panel **(A)** shows the percentage change in the unemployment rate vs the automation level for each occupation, while panel **(B)** shows the same thing for the long-term unemployment rate. The scatter in the results demonstrates that, due to network effects, the automation level only partially explains occupational unemployment. The right panels show the network effects i.e. the difference between the green and the black dots in the left panels. If this difference is positive, network effects are detrimental for the occupation. Panel **(C)** Network effects on percentage change in unemployment rate vs. median wage. Panel **(D)** Network effects on percentage change in long-term unemployment rate vs. median wage. 80

4.10 **Network structure effect at the aggregate level.** We show the change in unemployment and long-term unemployment due to the automation shock for different networks. The increase in unemployment rates is larger for the occupational mobility network. Both randomisations decrease the automation spike in unemployment rates, the weight reshuffling substantially more. 81

4.11 **Randomised network structure comparison with original network results**. In the *top* panels, we plot the percentage change in unemployment when running the model using the OMN and the reshuffled version of the OMN. In the *bottom* panels, we plot the percentage change in long-term unemployment comparison. In the *left* panels, we show the egde reshuffling, and in the *right* panels, the weight reshuffling. In both scenarios the changes in the unemployment rate are considerably smaller for the reshuffled versions of the OMN. 82
4.12 **Retraining policies.** The aggregate outcome of retraining policies, for either random retraining or targeted retraining. In blue, we plot 10 realisations of the random retraining strategy, all show very similar outcomes. 84
4.13 **Retraining policies at the occupation level.** In the *left* panel, the targeted retraining towards occupation with long-term unemployment percentage change above 50. In the *right* panel, random retraining. We pick one at random one realisation of the random retraining to show in this plot. 85

5.1 **A schematic network representation of supply side shocks.** The nodes to the left represent the list of essential industries at the NAICS 6-digit level. A green node indicates essential, a red node non-essential. The orange nodes (center-left) are more aggregate industry categories (e.g. 4-dig. NAICS or the BLS industry categories) for which further economic data is available. These two sets of nodes are connected through industry concordance tables. The blue nodes (center-right) are different occupations. A weighted link connecting an industry category with an occupation represents the number of people of a given occupation employed in each industry. Nodes on the very right are O*NET work activities. Green work activities mean that they can be performed from home, while red means that they cannot. O*NET provides a mapping of work activities to occupations. 90
5.2 **Distribution of Remote Labour Index across occupations**. Boxplots showing the distribution of RLI for each 4-digit occupation in each 2-digit SOC occupation category. 94

5.3 **Distribution of Remote Labour Index across industries**. Boxplots showing the distribution of RLI for each 4-digit occupation in each 2-digit NAICS Industry category. 95
5.4 **Fraction employed in an essential industry vs Remote Labour Index for each occupation**. Omitting the effect of demand reduction, the occupations in the lower left corner, with a small proportion of workers in essential industries and a low Remote Labour Index, are the most vulnerable to loss of employment due to social distancing. . 97
5.5 **Workers that cannot work**. On the left, the percentage of workers in a non-essential job (33% in total). On the right, the percentage of workers that cannot work remotely (56% in total). The intersection is the set of workers that cannot work, which is 19% of all workers. A remaining 30% of workers are in essential jobs where they can work remotely. 98
5.6 **Supply and demand shocks for industries**. Each circle is an industry, with radius proportional to gross output. Many industries experience exactly the same shock, hence the superposition of some of the circles. Labels correspond to broad classifications of industries. . . 101
5.7 **Supply and demand shocks for occupations.** Each circle is an occupation with radius proportional to employment. Circles are colour coded by the log median wage of the occupation. The correlation between wages and demand shocks is 0.26 (p-value = 2.84×10^{-13}) and between wages and supply shocks is 0.40 (p-value = 1.5×10^{-30}). 102
5.8 **Labour shock vs. median wage for different occupations.** We colour occupations by their exposure to disease and infection. There is a 0.40 correlation between wages and the labour shock (p-value = 4.4×10^{-40}). Note the striking lack of high wage occupations with large labour demand shocks. 106

5.9 **Remote Labour Index and Essential Score in the occupational mobility network.** Panel (**A**) corresponds to the distribution of the Remote Labour Index across occupation in the occupational mobility network, where red indicates that the occupation has a high Remote Labour Index and green indicates a low Remote Labour Index. Panel (**B**) the essential score of occupations, where red indicates a high essential score and green a low one. In both panels red occupations are more exposed to job displacement either due to having a low Remote Labour Index or belonging to an essential industry. 106

5.10 **Distribution of the COVID-19 shocks to labour demand.** We plot the distribution across occupations. Some occupations in the healthcare sector experience an increase in demand, while most occupations experience a decrease in demand. 107

5.11 **Unemployment and long-term unemployment rate at the aggregate level.** In all panels, the grey area denotes the period during which we measure the percentage change in unemployment and long-term unemployment. Panel (**A**) shows the evolution of the target demand for two occupations. The occupation coloured in blue has a positive COVID-19 shock, while the occupation coloured in red a negative COVID-19 shock. Panels (**B**) and (**C**) show the unemployment and long-term unemployment rates through time. The green line corresponds to the occupational mobility network while the red line to the complete network. 109

5.12 **Occupation specific changes in unemployment and long-term unemployment** The green dots are for the occupational mobility network and the orange dots are for the complete network. The size of the green dots is proportional to the employment of the occupation they represent and the colour gradient indicates the median wage of the occupation. 111

7.1 **Global financial networks**. Schematics of the networks that we construct using IMF data. In all networks, nodes represent countries and edges encode cross-country financial assets. The multiplex network has three layers, which correspond to three types of assets: equity (layer 1), debt (layer 2), and banking loans and deposits (layer 3). The multiplex network has no interlayer edges, so it is an edge-coloured multigraph. 128

7.2 **Example network for illustrating DebtRank.** A weighted and directed network with five nodes. The quantities f and g denote the weights of the edges. We use arrow widths to highlight our assumption that $g > f$. 131

7.3 **The global financial multiplex network for 2018.** The size of a node is proportional to its degree. For visualization purposes, we show only edges whose weight is ten or more times larger than the mean weight of the edges in the layer. We colour the edges based on the natural logarithm of their weight; darker colours indicate larger weights. In the *left* panel, we show the edge-coloured multigraph on which we run MultiRank. In the *right* panel, we also include interlayer edges that encode the vertical contagion mechanism in our threshold model. All layers have the same layout; in other words, the position of a country is the same in each layer. We label the countries on which we focus our discussion. To avoid overcrowding the figure with labels and because countries have the same position in all layers, we only label each country in one layer. [We produced the network plots using PYMNET (Kivelä 2014) and additional labels added with INKSPACE.] 135

7.4 **Results of computing MultiRank on the multiplex global financial network versus computing PageRank on the aggregated financial network**. In the *left* panel, the temporal evolution of MultiRank, and in the *right*, PageRank for several countries. We show the dominant financial countries in purple, mainland European countries in orange, and Asian countries in blue. 137

7.5 **Layer influence (i.e. MultiRank) $z^{[\alpha]}$ and assets in layers.** In the *left* panel, the temporal evolution of the layers with respect to the influence of layers. In the *right* panel, with respect to the fraction of assets in each layer. 137

7.6 **The number of systemically important countries (which we abbreviate as "systemic countries" in the axis label)** over time in the multiplex global financial network and in its associated monolayer networks. We show simulations on the multiplex global financial network (which we abbreviate by "Mx" in the legend) using dashed curves and simulations on the monolayer networks using solid curves. For the multiplex network, blue, orange, purple, and black correspond to vertical thresholds of 0.33, 0.50, 0.66, and 0.80, respectively. For the aggregated network, brown, green, blue, and black correspond to the equity, debt, banking, and aggregated networks, respectively. . . . 139

7.7 **Systemically important countries and horizontal thresholds.** We show the number of systemically important (i.e. "systemic") countries over time for different values of the horizontal threshold (i.e. capital buffer) for the multiplex global financial network (in the *left* panel) and aggregate network (in the *right* panel). The solid curve is for the baseline values $b_i = \mathrm{CAR}_i$, the dot–dashed curves are for values that below the baseline, and the dashed curve are for values above the baseline. The colour of a curve indicates a specific value of the threshold. 139

7.8 **Buffer size and systemic importance of countries in our multilayer threshold model.** The coloured entries signify that a country is systemically important for a given year and horizontal threshold, and a white entry signifies that a country is not systemically important. The blue entries correspond to a country being systemically important for a horizontal threshold that is smaller than the baseline value $b_i = \mathrm{CAR}_i$, and the red entries indicate systemically important countries for a horizontal threshold that is larger than the baseline. For this plot, we select all countries that are systemically important in at least four years for the baseline horizontal threshold. 141

I.1 **Occupation's self-loop distribution.** In the *left* panel, self-loops of occupations before setting minimum value and linear re-scaling. In the *right* panel, self-loop of occupations after setting minimum value and re-scaling. 149

I.2 **Heterogenous self-loops effect** We compare the model results between using homogenous and heterogenous self-loops, for both percentage change in unemployment and long-term unemployment change . . 150

I.3 **Change in unemployment and long-term unemployment with different values of γ.** The *top* panels show the change in unemployment rates vs the automation probability. The *bottom* panels show the change when $\gamma = 5\delta_u$ and $\gamma = 20\delta_u$ on the y axis and on the x-axis when $\gamma = 10\delta_u$. 151

I.4 **Average and alternative unemployment and long-term unemployment rates.** In the *left* panel, the percentage change in the average unemployment and long term unemployment rates in green and the percentage change in the alternative unemployment and long term unemployment rates in cyan. In the *right* panel, the percentage change of the average unemployment and long-term unemployment rate vs the percentage change of the alternative unemployment and long-term unemployment rate 152

I.5 **Frey and Osborne shock with different post-automation target demand scenarios**. In each panel we plot on the x-axis the percentage change in the period unemployment rates when the aggregate demand does not change and on the y-axis the percentage change in the period unemployment rate when the aggregate demand does change. In the *left* panels we assume the aggregate demand increases by 5%, and in the *right* we assume it decreases by 5%. 154

I.6 **Shock duration and measuring windows effect on the measurements on unemployment rates.** In the *top.* panelf, for different duration of the automation shock (20, 30 and 40 years) we show target demand of two occupations with similar demand level (Childcare workers and officer clerks), the unemployment, and long-term unemployment rates. The grey area denotes the whole period of automation, meaning that the target demand has reached the automation level within a 1×10^{-4} tolerance. The coral area denotes the sharp transition period which is middle steepest part of the Sigmoid shock. In the *bottom* panels, for each occupation we plot the percentage change during the whole transition period vs the percentage change during the sharp transition period. Occupations are colored by their automation probability. 155

II.1 **Schematic network representation.** The same schematic network representation of supply side shocks as in the main text, but now also including mathematical notation. The K-dimensional vector u below the NAICS 6-dig. (left nodes) encodes essential industries with binary elements. This set of nodes is connected to relevant industry categories by concordance tables (incidence matrix S). Matrix M connects the N industry categories with J occupations where an element represents the corresponding employment number. The ability to perform work activities (right nodes) from home is represented in vector v, also by binary elements. We use occupation-activity mappings provided from O*NET, represented as incidence matrix T. The grey arrows show the direction of shocks to industries and employment. The shock originating from the list of essential industries is mapped directly onto the broader industry categories, before it can be computed for occupations. Conversely, the Remote-Labour-Index is first mapped onto occupations and then projected onto industries. 157

II.2 **Distribution of remote labor index and essential score.** In the *left* panel, the relationship between remote labour index and median wage. The Pearson correlation is 0.42 (p-value = 7.3×10^{-33}). In the *right* panel, the relationship between fraction of workers in essential industries and wage. The Pearson correlation is 0.35 (p-value = 1.12×10^{-22}). 165

II.3 **Distribution of remote labor index and essential score.** In the *left* panel, the distribution of the remote labour index for the 740 occupations. In the *right* panel, the distribution of the share of essential jobs within each of the 740 occupations. 165

II.4 **Distribution of shocks.** In the *left* panel, the supply shock distribution across occupation. In the *right* panel, the demand shock distribution across occupations. 166

II.5 **Distribution of remote labor index and essential score.** In the *left* panel, the distribution of the remote labour index, aggregated to 169 industries. In the *right* panel, the Fractions of essential sub-industries per industry category. 167

II.6 **Distribution of shocks.** In the *left* panel, the supply shock distribution across industries. In the *right* panel, the demand shock distribution across industries. 167

II.7 **Shock distribution across industries.** 168
II.8 **Probability of infection and wage.** Relationship between wage and probability of infection in linear and log scale. The Pearson correlation is 0.36 (p-value = 1.7×10^{-21}). However we this correlation is mostly driven by high salaries in the health sector, but there are many low-wage occupations with a significant exposure to infection. 170
II.9 **Unemployment and long-term unemployment rate at the aggregate level.** In all panels, the grey area denotes the period during which we measure the percentage change in unemployment and long-term unemployment. The *left* and *right* panels correspond to a 0.5 and 0.1 shock diminishing factor respectively. *Top* panels show the evolution of the target demand for two occupations. The occupation coloured in blue has a positive COVID-19 shock, while the occupation coloured in red a negative COVID-19 shock. *Middle and bottom* panels show the unemployment and long-term unemployment rates through time. The green line corresponds to the occupational mobility network while the red line to the complete network. 174
II.10 **Occupation specific changes in unemployment and long-term unemployment.** The *left* and *right* panels correspond to a 0.5 and 0.1 shock diminishing factor respectively. The green dots are for the occupational mobility network and the orange dots are for the complete network. The size of the green dots is proportional to the employment of the occupation they represent and the colour gradient indicated the median wage of the occupation. 175

Introduction

Many, if not most, macroeconomic models assume that the economy is at equilibrium, i.e. a state where no agent has an incentive to modify their chosen actions. For a long time, researchers have known that this assumption is only an approximation and that the economy is not at equilibrium. Nonetheless, since equilibrium allows one to solve models analytically, it has been a widely adopted assumption. However, as the 2009 financial crisis portrayed, the economy can be far from equilibrium, and as the recent COVID-19 pandemic showed, it can take a long time for the world to reach a "new normal" and for the economy to converge to a new equilibrium (if an equilibrium is reached at all[1]). Therefore, developing non-equilibrium models seems crucial for answering policy-relevant questions in a timely manner.

There are multiple definitions of equilibrium in economics. In this thesis, we refer to equilibrium as a state of the economy where no agent has an incentive to modify their chosen actions (Arthur 2014). Agents are assumed to be rational in their decision making, meaning that they have clear preferences, use expected values or functions of variables to consider uncertainty, and always choose the strategy or action that has the optimal expected outcome for themselves. Equilibrium is often referred to as the state where supply equals demand because if demand exceeds supply, the seller would have an incentive to produce and sell more goods or increase prices, while if supply exceeds demand, the agent would have wasted resources in producing said good, and would have been better off choosing to produce less or lowering prices.

Assuming that the economy is in a state of equilibrium can explain many economic phenomena. For example, in Labour Economics, spatial general equilibrium models have explained why economic activity agglomerates and consequently why there are enormous differences in worker earnings across metropolitan areas (Moretti 2010, Greenstone et al. 2010); studies using imperfect competition under monopsony have explained why unemployment does not necessarily rise after the introduction of a

[1]Recent research suggests that many economic systems rarely converge to an equilibrium (Pangallo et al. 2019).

minimum wage policy (Alan 2011, Bachmann & Frings 2017); and partial equilibrium search and matching models have explained why there is positive unemployment when there are both workers searching for jobs and employers looking for employees (Mortensen 1982, Diamond 1982).

Although equilibrium analysis has substantially increased our understanding of the economy, renowned economists have questioned where the limits of this assumption lie (Arrow 1986, Arthur 2014). As Arthur (2014) pointed out, assuming equilibrium (i.e. assuming a state where there is no incentive for further behavioural adjustments) rules out any transitory phenomena including adaptation, innovation, and structural change. An alternative is complexity (or non-equilibrium) economics, which focuses on how individuals behave and react in an economic environment and how their reactions may, in turn, alter the economic environment (Farmer et al. 2012). Instead of solving complex optimisation problems to maximise their utility, agents take decisions based on simple rules or heuristics and there is usually some degree of randomness in their decisions. Complexity economics is also characterised by modelling heterogeneous agents whose interactions (often modelled through a network) play an essential role in the economy (Farmer & Foley 2009). Additionally, complexity economics also has a preference for modelling transient states and for using fine-grained data.

To better portray the difference between standard equilibrium approaches and complexity economics, let us discuss automation in the labour market, which has raised concerns about increasing unemployment and inequality (Frey & Osborne 2017). Historically, automation has both destroyed and created new jobs. However, even if a job vacancy opens for every displaced worker, labour market frictions, such as a skill mismatches between the recently unemployed workers and opened job vacancies, can impede some workers from securing a job after being displaced (Restrepo 2015, Acemoglu & Restrepo n.d.).

Equilibrium models consider firms and workers to be rational, so firms would decide whether or not to open job vacancies by maximising the expected values of their profits throughout time. Similarly, workers decide how much to invest in training, how often to search for jobs, and whether take a job by solving a system of equations that maximises their utility. However, several studies have shown that people often make irrational decisions and questioned the validity of assuming agents are rational (Kahneman 2011, Thaler & Sunstein 2009). A complexity economics approach would instead define simple behavioural rules for workers and firms. Another difference is that equilibrium studies focus on computing a new equilibrium state, not on how the actual transition may play out or how long it would take. Instead, from a

complexity economics perspective, one can analyse the transition period between a pre-automation and post-automation economy, which can be helpful for understanding how automation may displace workers in the near future.

Developing non-equilibrium models is challenging. By assuming equilibrium one can define the initial state of the economy, but once we drop this assumption, we must rely on data to initialise the model. These data are not always available or are available only at an aggregate level. Similar problems arise with calibration. Another challenge comes in solving or exploring the outcomes of the models. In non-equilibrium models agents behaviour may not be easily translated into equations. Instead, agents' decisions may depend on heuristic such as imitating their neighbors (Farmer & Foley 2009) or randomly choosing an option from a set. When a non-equilibrium model cannot be translated into equations, the model is usually solved via simulations, which often require vast computational resources. The lack of data and computational power previously constrained the development of non-equilibrium models. However, with the rise of data availability and computational resources, we can now model an economy without assuming equilibrium.

In this thesis, we build non-equilibrium models, which focus on transient dynamics, acknowledge complex network interactions between agents, and are data-driven. The aim is to contribute to the development of models that can provide accurate answers to policy-relevant questions in an efficient manner. This thesis is split into two parts.

In the first part, we build a labour market model designed to understand how the labour market behaves during transition periods. In this model, workers react to external displacement shocks (e.g. automation shocks) and may apply for jobs in different occupations. They choose which jobs to apply to based on their current occupation and the number of jobs available at the time, but there is also randomness in their decision, and they do not solve complex optimisation problems. This model's overall aim is to understand how the employment prospects of workers are affected by automation and the behaviour of other workers.

In the second part, we build a multilayer network model of global financial contagion. We assume a financial system out of equilibrium, where a country (or a set of countries) is financially distressed. Taking a core concept from complex systems, we assume agents' (in this case countries) interactions are encoded via a complex network (Farmer & Foley 2009). Furthermore, we model interactions of various types using a multiplex network structure. Originally, edges in networks only modelled one type of interaction between nodes; however, network theory has expanded to develop multilayer networks, where multiple types of interaction can be modelled through

layers (Kivelä et al. 2014). Recent studies in financial contagion have remarked the importance of modelling the financial system as a multilayer network (Poledna et al. 2015). The aim of this part of the thesis is to study how the multilayer structure of the global financial network can affect financial contagion.

We now summarise the content of each part of the thesis.

Part I. A data-driven network model of occupational mobility. Chapter 1 reviews the literature: we discuss the concept of structural change and review the literature on labour economics and complex systems. This chapter introduces most of the terminology and notation from the networks and dynamical systems literature used in the first part of this thesis.

Chapter 2 introduces our data-driven model, where workers move through an empirically derived occupational mobility network in response to labour reallocation shocks such as automation or a pandemic. The chapter explains how we build the occupational mobility network from U.S. census data. Then, the chapter introduces the model, defines the agents and their behaviour, and describes how we model a labour market environment that reacts to external labour reallocation demand shocks. The model is agent-based, and simulations are computationally costly to run. To explore the model's behaviour more comprehensively, we derive deterministic equations that can be solved numerically in seconds and approximate the model's behaviour in the limit of a large number of agents. These approximations are the main result of this chapter and are vital for deriving the results from the subsequent chapters.

Chapter 3 explores the macroeconomic behaviour of the model. We focus on the Beveridge curve, which is the graphical representation of the relationship between unemployment and vacancy rates. One of the most important empirical regularities of macroeconomics is that the Beveridge curve is downward sloping (i.e. there is a negative relationship between unemployment and vacancy rates). This chapter explains how we use the Beveridge curve to calibrate the parameters in our model. The main result of this chapter is that the model not only fits the Beveridge curve considerably well, but that the model also reproduces the empirically observed counter-clockwise cyclicality of the Beveridge curve.

Chapter 4 explores how automation may impact the employment prospects of workers of different occupations using the model we developed in Chapter 2 and calibrated in Chapter 3. To calibrate automation shocks, we use the estimates of the probability of computerization by (Frey & Osborne 2017) and suitability for machine

learning by (Brynjolfsson et al. 2017). We find that the occupational mobility network structure plays an important role in workers' prospects of different occupations. Furthermore, when we use the probability of computerization (Frey & Osborne 2017) to calibrate the automation shocks, we find that the network effects are most likely to increase the unemployment rate for workers in low-skill occupations.

Chapter 5 focuses on predicting the effects of the COVID-19 pandemic on employment. We provide quantitative predictions of first-order supply and demand shocks for the U.S. economy associated with the COVID-19 pandemic at the level of individual occupations and industries. We find that occupations related to the transport, entertainment, and restaurant industry are severely affected by these shocks. Furthermore, high-wage occupations are relatively immune from adverse supply- and demand-side shocks, while low-wage occupations are much more vulnerable. Later, we use our occupational mobility model to understand how workers may be affected by second-order network displacement effects.

Part II. Multiplex global financial contagion. Chapter 6 reviews the literature on multilayer networks and their applications to model global financial contagions. This chapter introduces notation and terminology. Chapter 7 presents our work in collaboration with the International Monetary Fund (IMF). This work focuses on identifying systemically important countries. Using data on bilateral financial positions, which are split by asset type, we build a global financial multiplex network in which nodes represent countries, edges encode cross-country financial assets of various types, and layers represent asset types. We then develop a multiplex threshold model of financial contagion in which a shock can propagate either within a layer or between layers. We find that the number of systemically important countries can be twice as large when we take into account heterogeneity of financial exposures (i.e. when using a multiplex network) in comparison to a contagion on an associated aggregated global financial network (i.e. on a monolayer network), as often done in other studies.

Part I

A data-driven network model of occupational mobility

Chapter 1

Literature review

1.1 Overview

In the first section of this chapter, we explain the concept of *structural change* and discuss how automation and the COVID-19 pandemic may cause or have caused structural changes in the economy. In the second section, we review the relevant literature in labour economics. In the last section, we explain some tools from complex systems useful for our modelling framework.

1.2 Structural change and the future of work

The nature of work, that is, the tasks and activities performed during working hours, has changed substantially over time. In earlier centuries, most of the tasks involved manual labour such as lifting and carrying materials; today, many tasks involve cognitive work such as programming or coordinating with others. This changing nature of work is closely related to *structural change*, which refers to a change in the way an economy functions. For example, an employment reallocation across industries is a structural change that impacts many people's work. This happened during the industrial revolution, when employment reallocated from agricultural to manufacturing industries (Krüger 2008) and when several tasks were automated by machines, which in turn displaced some workers.

Innovation and structural change may destroy some jobs, but they also create new jobs. A recent study (Dachs et al. 2017) found that innovation's net contribution to employment growth has been mostly positive. While some occupations, such as elevator operators, have disappeared due to automation, new occupations have also

emerged. A recent study suggested that each year roughly 0.56% of new jobs in the U.S. are in new occupations (Lin 2011).

From a Schumpeterian perspective, structural change is a consequence of creative destruction (Schumpeter 1942), that is, the process by which the economy revolutionises its structure from within by having more productive firms entering the market and displacing less productive firms. This means that there are winners and losers in a structural change: some firms increase their profits, but other firms go bust. Similarly, some workers increase their wages while others may face *structural unemployment*, i.e. they may be unemployed due to a mismatch between their skills and the skills demanded of workers by potential employers (Restrepo 2015). However, as Schumpeter himself pointed out, it is important to mitigate creative destruction's adverse effects. In his own words, "there is certainly no point in trying to conserve obsolescent industries indefinitely; but there is a point in trying to avoid their coming down with a crash and in attempting to turn a rout which may become a centre of cumulative depressive effects into orderly retreat" (Schumpeter 1962).

Today's economy is evolving, and there are several potential drivers of structural change. The threat of climate change means that labour demand may increase in carbon-neutral industries and decrease in emission-intensive ones (Garrett-Peltier 2017, Vona et al. 2018). The rise of AI and software technologies could destroy jobs (Vermeulen et al. 2018), and recent changes in the economy due to the COVID-19 pandemic have displaced a significant amount of workers (Chetty et al. 2020). In the following subsections we discuss the latter two, since they are relevant for Chapters 4 and 5 respectively.

1.2.1 Automation

Automation in the labour market is not new, but recent developments in AI threaten to destroy jobs previously considered beyond reach. For example, driving cars was not considered an automatable task in 2003 (Autor et al. 2003). Today, self-driving vehicles are thought to be close to commercial distribution, with the potential to replace 2 million truck and taxi drivers in the U.S. alone (Brynjolfsson et al. 2017). Frey & Osborne (2017) estimated that approximately half of the jobs in the U.S. are at risk of automation. Other studies (Arntz et al. 2016, Brynjolfsson & Mitchell 2017) have nuanced these numbers by arguing that jobs will evolve rather than disappear but have agreed that AI will have a considerable effect on the economy and the workforce. Most of the literature agrees that the biggest challenge lies in the reallocation of

workers and the development of new skills (Arntz et al. 2016, Brynjolfsson et al. 2017, Frank et al. 2019).

Automation affects occupations differently. For example, in the 2000s, digital technologies affected middle-wage occupations disproportionately. This happened because computerization replaced not only manual labour (which tend to be low-wage occupations) but also cognitive routine occupations such as clerical work (which tend to be middle-wage occupations) (Autor 2015*a*). Some studies suggest that low-wage and low-skill occupations are most at risk of automation (Arntz et al. 2016, Frey & Osborne 2017). To design policies that can help displaced workers reallocate in new jobs it is crucial to understand which workers will be the most affected.

Automation has different types of direct effects on the labour market. First, automation may cause some workers' skills to become obsolete and cause jobs to disappear (Mokyr et al. 2015). Second, since new technologies usually require new skills to operate, new types of jobs will emerge (Lin 2011). Third, automation improves productivity and raises aggregate income, causing labour demand to increase in high consumer demand sectors. These three effects will lead to a significant reallocation of labour demand, with some occupations increasing and others decreasing their labour demand.

1.2.2 The COVID-19 pandemic

The COVID-19 has drastically changed how the economy functions. At the time of writing, the pandemic is still ongoing, and much more research is expected to come. This section gives a brief overview of the early literature on the economic consequences of the COVID-19 pandemic.

Supply chains have been disrupted, causing immense losses to economic output (Inoue & Todo 2020). People's consumption also changed, early in the pandemic (Muellbauer 2020) estimated a fall of quarterly consumption of 20%. More recent studies found a dramatic consumption decrease in the restaurants and leisure sectors and some increase in groceries consumption (Chetty et al. 2020, Chen et al. 2020, Carvalho et al. 2020). Finally, the nature of work changed; many workers stopped on-site work and instead worked from home. A survey in February (Zhang et al. 2020) found that 27% of the labour force continued working at the office, 38% worked from home, and 25% stopped working in China. Adams-Prassl et al. (2020) surveyed U.S. and U.K. citizens in late March, and reported that the share of tasks that can be performed from home varies widely between occupations (from around 20 to 70%), and that higher wage occupations tend to be more able to work from home.

Several research efforts have focused on predicting who can work from home (Dingel & Neiman 2020, Hicks et al. 2020, Koren & Pető 2020). Many of these studies have drawn on occupation-level data from the Occupational Information Network (O*NET). For example, Hicks et al. (2020) considered the degree to which an occupation is required to 'work with others' or involves 'physical proximity to others' to assess which occupations are likely to be most impacted by social distancing. Similarly, to identify which workers were likely to work from home, (Dingel & Neiman 2020) used O*NET data on the extent to which occupations require daily 'work outdoors' or needed to engage with 'operating vehicles, mechanized devices or equipment'.

With governments mandating the closure of non-essential businesses, several studies have focused on quantifying essentialness. Bullard (2020) provides an undocumented estimate that around a half of the U.S. economy would be considered either essential or able to operate without creating risks of diffusing the virus. Another study by Barrot et al. (2020) estimated industry-level shocks by considering the list of essential industries, the closure of schools, and an estimate of the ability to work from home (based on ICT use surveys). Italy, Germany, and Spain have also released a list of essential industries (Fana et al. 2020).

1.3 Labour economics

1.3.1 Search theory

Why can there be unemployed workers when there are job vacancies? Search theory has provided a convincing explanation to these questions by acknowledging that it takes time and resources (from both workers and employees) to fill job vacancies. In other words, there are *labour market frictions*, which refer to the time a worker spends between jobs either searching for a job or transferring from one to another. These frictions can be caused by different reasons, including incomplete information and skills mismatch.

Labour market frictions gave rise to the *matching function* which is central to search theory (Petrongolo & Pissarides 2001). In its simplest form, the matching function takes the number of unemployed workers U and job vacancies V as inputs and outputs the number of successful matches m, i.e. the number of jobs created by pairing a worker with a vacancy. Mathematically, we denote this function by $m = M(U, V)$.

Usually, one assumes that the matching function is increasing in both arguments (i.e. the more workers looking for jobs or jobs vacancies, the higher the number of

jobs created) and concave (i.e. the marginal increase of the number of matches with respect to U and V decreases with U and V). Furthermore, it is usually desirable that

$$M(U,V) \leq \min(U,V), \tag{1.1}$$

and that

$$M(U,0) = M(0,V) = 0. \tag{1.2}$$

In other words, if there are no job vacancies nor workers looking for jobs, then there are no matches, and there can never be more jobs created than job vacancies posted or workers looking for jobs. Finally, when there are no search frictions $M(U,V) = \min(U,V)$ (Petrongolo & Pissarides 2001).

In this thesis, we use the micro-founded *urn-ball* matching function (Hall 1979, Pissarides 1979). The urn-ball matching function receives its name from the following narrative. Imagine each worker has a ball (or job application) with her name on it, and each job vacancy is an urn. Each worker places their ball in an urn. After all workers have placed their balls, a ball is drawn from each urn with uniform probability and the corresponding worker is hired.

We can calculate the expected number of job matches as follows. An unemployed worker will apply to a particular vacancy with probability $\frac{1}{V}$. Thus, the probability that the worker does *not* send her application to that vacancy is $1 - \frac{1}{V}$. For U unemployed workers sending one job application, the probability that a particular vacancy does not receive an application is $\left(1 - \frac{1}{V}\right)^U$. Since each vacancy receiving an application hires one worker, the expected number of matches is the expected number of job vacancies that received a job application so that

$$M(U,V) = V\left(1 - \left(1 - \frac{1}{V}\right)^U\right). \tag{1.3}$$

It is common to simplify the above expression using the approximation $(1 - \frac{1}{x})^y \approx e^{-y/x}$ for large x and y. This leads to

$$M(U,V) \approx V\left(1 - e^{-U/V}\right). \tag{1.4}$$

The urn-ball matching function was one of the first proposed, we use it in this thesis due to its simplicity and micro-foundations. However, many more matching functions have been proposed, see Petrongolo & Pissarides (2001) for a review.

1.3.2 Cyclical and structural unemployment in the Beveridge curve

Historically, the economy has exhibited *business cycles*, i.e. fluctuations of economic activity, measured by the Gross Domestic Product (GDP), around its long-term growth trend. The business cycle has booms or recovery periods when GDP is above the long-term growth trend and bust or recessions periods when GDP is below the long-term growth trend. As Okun observed for the U.S. economy, an increase in GDP corresponds to a decline in unemployment (Okun & Potential 1962). This means that unemployment oscillates during business cycles, during recessions unemployment increases, and during recoveries unemployment decreases. These oscillations happen (at least in part), because there is not enough aggregate demand in the economy to hire every person who wants to work, and receives the name of *cyclical unemployment*.

The vacancy rate also oscillates during business cycles, following an inverse pattern than unemployment. The graphical representation of the relationship between the unemployment and vacancy rate is known as the *Beveridge curve* and is one of the most well-known macroeconomic stylised facts (Diamond & Şahin 2015, Beveridge 1944): When more vacancies open, unemployment goes down. The intuition is that when there are many vacancies, unemployed workers get a job faster, so the unemployment rate is low. Similarly, when there are few vacancies, unemployed workers are less likely to find jobs, so the unemployment rate is high.

The Beveridge curve's dynamics are not only driven by cyclical unemployment but also by structural changes and labour reallocation. If there is an increase in the mismatch between the skills of the workers searching for jobs and the skills employees demand then, for same vacancy rate, there can be an increase in structural unemployment. To study labour reallocation and structural unemployment several works have used the task approach, where the economy produces a final good made of a combination of continuous tasks (Autor 2013, Acemoglu & Autor 2011, Acemoglu & Restrepo 2018, Autor 2013). Workers are usually classified into two or three types, low, (medium), and high skilled according to the tasks they can perform. Depending on the state of the economy, the demand for low-skill workers may be less than the supply, resulting in structural unemployment. Labour reallocation has also been studied as a result of economic shocks to particular sectors (Neffke et al. 2018, Diodato & Weterings 2014), while others have studied how labour reallocation changes during business cycles (Haltiwanger et al. 2018). Although structural unemployment has been studied extensively, there is still an ongoing debate on whether it is structural

or cyclical unemployment which drives the shifts of the Beveridge curve to a greater extent (Diamond 2013).

1.3.3 Networks and labour economics

Networks have been used to better understand how skills, knowledge, and work activities are distributed across occupations (Alabdulkareem et al. 2018, Mealy et al. 2018, Nedelkoska et al. 2018, Dworkin 2019). The Skillscape is a network where nodes are occupations and links denote similarity in terms of skills (Alabdulkareem et al. 2018). This network clusters into two, one cluster relates to social-cognitive skills and high-wage occupations, while the other on sensory-physical skills and low wage occupations. Occupation networks can also acknowledge that worker's skills, tasks, and work activities are not only of different difficulty but also of different nature. For example, both mathematicians and surgeons perform difficult skills and are usually categorized as high skills workers, but the nature of these two occupations are very different. In the Job Space network made by Mealy et al. (2018), where nodes are occupations and links denote similarity between occupations in terms of shared work activities, mathematicians and doctors are far apart (Mealy et al. 2018). In other words, this network can distinguish the different nature of work of these two occupations.

Networks have been used to study how workers move between jobs. Schmutte (2014) and Nimczik (2017) constructed a network of job transitions for the U.S. and Austria, respectively. These studies found a clustered structure, with high mobility within clusters and little mobility between clusters. Neffke et al. (2017) defined the inter-industry labour mobility network for Germany and used labour flows to measure inter-industry relatedness and found that this measure is remarkably stable. Guerrero & Axtell (2013) constructed the labour flow network, where nodes are firms and links denote transition of workers between firms and found that 90% of employment growth is concentrated on 10% of firms. Using this network, Axtell et al. (2016) built a model where workers moved through the labour flow network as they search for new jobs and found that, in equilibrium, the behaviour of the labour market and the Beveridge curve depends on the topology of the network.

1.4 Dynamical systems, stochastic processes, and networks

This section gives a brief overview of some mathematical concepts and tools. This section starts with the general definition of stochastic processes and dynamical systems. We then discuss the concept of stability. The last part of this section focuses on network theory.

1.4.1 Stochastic processes

A stochastic process $\underline{X} = \{X_t, t \in T\}$ is a collection of random variables X_t, where t is an element of the index set T. We often interpret t as time and refer to x_t as the state of the process at time t. If T is countable we say that $\underline{X}$ is a discrete-time stochastic process while if T is a continuum we say that $\underline{X}$ is a continuous-time process (Ross et al. 1996). In this thesis, we focus on discrete-time stochastic processes.

If a stochastic process has a finite or countable number of possible states, we denote these states by non-negative integers and the set of possible states by $S = \{0, 1, 2, ...\}$. We say that the process is in state i at time t, when $X_t = i$. A stochastic process is *stationary* if the finite-dimensional distribution of X_t remains the same even when time is shifted by any amount k. A stationary process is said to be in a steady-state or equilibrium (Meyn & Tweedie 2012). In this thesis, we reserve the word equilibrium to refer to economic equilibrium and use steady-state to refer to a stationary process.

Markov chains. Markov chains are a particular type of stochastic processes that have been used to study physical, biological and economic systems such as Brownian motion, population dynamics, and financial asset pricing among others (Serfozo 2009). A Markov chain is a stochastic process that takes a finite or countable number of possible states, where the conditional distribution of any future state X_{t+1} given past states $X_0, X_1, ...X_{n-1}$ depends only on the previous state X_t; that is

$$P\{X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, ..., X_0 = i_0\} = P\{X_{t+1} = j | X_t = i\} \equiv P_{ij}, \quad (1.5)$$

for all state $i_0, ..., i_{t-1}, i, j$ and $\forall\ t \geq 0$. P_{ij} is called the probability or transition matrix. The probability that a process in state i reaches state j in t time steps is P_{ij}^t (Ross et al. 1996).[1]

[1] This result is known as the Chapman-Kolmogorov equation.

Stability. To discuss stability in Markov chains we will use two theorems on stationary and limiting distributions respectively, as well as the following definitions. A Markov chain is *irreducible* if states i and j are accessible to each other. In other words, if $\forall i, j \in S$ there exists $t \geq 1$ for which $P_{ij}^t > 0$. A state i of a Markov chain is *positive recurrent* if the probability of returning to that state within an infinite number of time steps is one. A Markov chain is *ergodic* if it is irreducible and all of its states are positive recurrent (Serfozo 2009). A *stationary distribution* π of a Markov chain satisfies the equation $\pi = \pi P$, where π_i is $i-$th entry of π and gives the probability of being on state i. The *limiting distribution* of a Markov chain is given by

$$\lim_{t \to \infty} P\{X_t = i\} = \pi_i, \ \forall i \in S. \tag{1.6}$$

This distribution does not depend on the initial state X_0 (Serfozo 2009).

The Stationary Distribution Theorem for Markov chains states that an irreducible Markov chain has a positive stationary distribution if, and only if, all if its states are positive recurrent. Furthermore, this stationary distribution is unique. The Limiting Distribution Theorem for Markov chains states that if the Markov chain is ergodic, the limiting distribution is the stationary distribution.

Intuitively, a system is stable if the system's general properties do not change significantly over time. There is no unique definition of stability in Markov chains, but the concepts of recurrence and irreducibility can help provide some definitions (Meyn & Tweedie 2012). In order of restrictiveness, Meyn & Tweedie (2012) provide the following three concepts of stability.

- **Irreducibility.** In its weakest sense, we can say that stability requires that the Markov chain is really just one Markov chain and not multiple and require the Markov chain to be irreducible. An irreducible Markov chain rules out the possibility that a state is so unstable that given a small perturbation, the system changes dramatically and can never return to the previous state.

- **Recurrence.** To increase the degree of stability, one can ask for recurrence in a Markov chain. This means that stability not only requires that reaching any state is possible but that the system will eventually reach all states.

- **Ergodicity.** A third level would require ergodicity. In this sense, the Markov chain has an "invariant regime" or "centre". If the Markov chain is perturbed and leaves this regime, it will return in a strong probabilistic sense to the invariant regime.

Strictly speaking, a Markov chain is only at a steady-state when it started at this state or, if the Markov state is ergodic, it will be at the steady-state after an infinite amount of time. However, for practical purposes, in this thesis, we say that a system reaches a steady-state when there is no observable difference in the system's stochastic behaviour.

1.4.2 Dynamical systems

Dynamical systems allow us to describe how a system evolves in time with respect to a set of rules. A dynamical system is composed of (i) a state-space S made of all the possible states (i.e. coordinates) of the system. The state-space may be discrete or continuous; (ii) a set of times T which may be discrete or continuous; and (iii) an evolution rule or function, which can be deterministic or stochastic, $\Phi : S \times T \to S$, and maps the system between consecutive states. For example, let $\mathbf{x}_t \in S$ be the state of the system at time t, then $\mathbf{x}_{t+1} = \Phi(\mathbf{x}_t)$ and $\mathbf{x}_{t+k} = \Phi^k(\mathbf{x}_t)$; Φ plays a similar role in dynamical systems than the transition matrix P takes in Markov chains. Both Φ and P represent, to some extent, the set of rules under which the system evolves in time. The rest of this section focuses on discrete dynamical systems.

In a deterministic dynamical system, a fixed-point $\mathbf{x}^* \in S$ satisfies the fixed-point equation

$$\mathbf{x}^* = \Phi(\mathbf{x}^*). \tag{1.7}$$

This equation is similar to the balance equation for Markov chains. Fixed-points in dynamical systems are analogous to stationary distributions in Markov chains.

Stability. A fixed-point may be stable or unstable. Intuitively, when a system is at a fixed-point and returns to this fixed-point after experiencing a small perturbation, the fixed-point is stable. Mathematically, the stability of a discrete dynamical system given by a continuous and differentiable map Φ can be determined using the Jacobian matrix J, with components $J_{ij} = \frac{\partial \Phi_i}{\partial x_j}$ (Brown 2007). In particular, a fixed-point is stable if all the eigenvalues of the Jacobian matrix J (evaluated at the fixed-point) have absolute value strictly less than one. If at least one of the eigenvalues has an absolute value strictly greater than one, then the fixed-point is unstable. The stability of the fixed-point is undefined when all eigenvalues have an absolute value equal to one. For a discrete-time dynamical system, there are different notions of stability. We discuss the following two, which are presented in Meyn & Tweedie (2012).

1. A dynamical system is *Lyapunov stable* if for any two initial conditions $x, y \in S$ the following is satisfied

$$\forall \epsilon > 0 \; \exists \delta > 0 \;\; \left[||x, y|| < \delta \Rightarrow \forall t \in T \; ||\Phi^t(x), \Phi^t(y)|| < \epsilon \right], \tag{1.8}$$

where $||\cdot||$ denotes the distance in S. In other words, we require that two initial conditions that are initially close have a comparable behaviour in time. This notion of stability is similar to irreducibility, where we require that the system does not end in completely different states when starting from two close initial conditions.

2. A more strict sense of stability is *global assymptotic stability*, which requires that there exists a unique fixed-point x^*, and all initial conditions $x \in S$ converge to this fixed-point. That is

$$\lim_{t \to \infty} ||\Phi^t(x), \Phi^t(y)|| = 0. \tag{1.9}$$

This notion of stability is similar to ergodicity for Markov Chains, where we require that the system returns to an "invariant regime".

1.4.3 Networks

Networks have been used to study systems of multiple fields, including biology, sociology, and economics (Newman 2018*b*). A network $G = (V, E)$ has a set of vertices V and a set of edges E. Edges are pairs of vertices; if a network is directed, these pairs are ordered. The first node corresponds to the edge's start and the second node to the end of the edge. The out-degree of node i is the number of edges that start in node i, while the in-degree of node i is the number of edges that end in node i. The strength of a node is the weighted degree. A network is weighted if the edges have a number, i.e. a weight, associated with them. We denote the weight of an edge (i, j) by w_{ij}. We can represent a network of n nodes by its adjacency matrix A whose entries are given by

$$A_{ij} = \begin{cases} \omega_{ij} & \text{if (i,j)} \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{1.10}$$

where ω_{ij} is the weight of the edge (i, j) if the network is weighted, and $\omega_{ij} = 1$ if the network is unweighted.

The structure of a network can help us understand the system under study. Several structural measurements can help us characterise a network. Centrality measures, such as Katz centrality and PageRank, identify important nodes in the network. The

clustering and transitivity coefficients measure the extent to which two adjacent nodes share the same neighbors. Networks may also present a core-periphery structure, where nodes in the core are highly interconnected while peripheral nodes connect loosely to both nodes in the core and in the periphery (Newman 2018*a*).

A network is said to have a community structure if nodes can be grouped into sets so that nodes are densely connected within their sets. Several community detection algorithms have been proposed (Fortunato & Hric 2016) including modularity (Newman 2006) maximisation algorithms, the map equation (Rosvall et al. 2009) and stochastic block modelling (Abbe & Sandon 2015). The latter two have been widely used for directed networks.

Null models have also been used ubiquitously in network science to better understand how particular network characteristics affect a system (Newman 2006). In the Erdos-Renyi model $G(n, p)$ a network on n nodes is constructured by connecting nodes at random, where each edge is included in the network with probability p. The configuration model for unweighted networks generates a random network from a given degree sequence, and keeps the degree distribution constant (Newman 2006). A common approach for weighted and directed networks is to keep a network structure fix and rewire edges (i.e. change the start and/or endpoint). There are some more refined extensions of the configuration model for weighted networks that focus on both the degree and strength distribution (Serrano & Boguñá 2005).

Finally, many systems can be modelled as a dynamical system on a network Gleeson (2013). For example, several social science studies have focused on voter and threshold models on networks. Epidemiological models on networks have been studied extensively with compartmental models. While some of these models can be solved analytically, with increasing complexity and non-linearities of the systems, simulations and numerical analysis are often used to study these systems.

Chapter 2

A data-driven network model of the labour market

2.1 Introduction

Workers are affected by short and long-term changes in the economy. For example, in the short-term, economic recessions increase unemployment and in the long term, structural change reallocates labour demand. This chapter presents a new agent-based model of the labour market aimed at understanding transitory dynamics. In particular, this model considers continuous changes in the demand for workers of different occupations and how workers may adapt to these changes by changing occupations.

We build upon the literature on labour economics and networks (see Chapter 2) and go further by modelling labour market during a transitory period during which workers are displaced by external shocks. Central to our model is an empirically derived *occupational mobility network*, in which nodes are different occupations and edges correspond to the probability that workers transition between them. The overall structure of this network influences the efficiency with which workers are reallocated across occupations.

We model the labour market as a stochastic process with discrete time steps. During each time step, occupations open vacancies and separate (fire) workers, unemployed workers search for a new job, and vacancies and job applications are matched. A key restriction is that the last occupation in which a worker is employed determines the possible jobs that she can apply for.

We analyse the dynamics of our model using master equations. In the limit where the number of jobs and workers are large, the behaviour of the model can be described by a deterministic dynamical system, which dramatically speeds up the analysis with

little loss of accuracy. We use empirical data both to create the occupational mobility network and to calibrate the parameters of the model.

The model we develop is general and can potentially be used to study network-induced labour frictions in any setting. As we explain in the next chapter, we calibrate the model by fitting it to the empirical Beveridge curve, which relates vacancies and unemployment rates. We show that the effects of network structure on aggregate unemployment are substantial. The mobility constraints imposed by the network structure increase unemployment by roughly 29%. In Chapter 4, we also show that changes in the distribution of the labour demand across the network can cause significant shifts in the aggregate unemployment rate even when the total supply and demand of jobs is held constant.

This chapter is structured as follows. First, we explain how we build the occupational mobility network from census data. In Section 2.3, we introduce our labour market model. In Section 2.4, we propose a deterministic dynamical system to describe the behaviour of the agent-based model and show with computer simulations that the dynamical system accurately reproduces the dynamics of the agent-based model in the limit of a large number of agents. In Section 2.5, we discuss the existence of a steady-state and fixed-point in the agent-based model and deterministic dynamical system, respectively. Finally, in Section 2.6, we conclude with remarks.

2.2 The occupational mobility network

We first construct an occupational mobility network representing the ease with which a worker can transition between occupations. Here we follow Mealy et al. (2018) and construct a network using empirical data on occupational transitions in the United States between 2010 and 2017. The nodes are occupations in the 2010 American Classification System (ACS); and the weights of the edges are proportional to the probability that a worker transitions between occupations (see Fig. 2.1).

We build the occupational mobility network as follows. We use monthly panel data from the U.S. Current Population Survey (CPS) to count the number of workers T_{ij} who transitioned from occupation i to occupation j during the period of study. Letting $T_i = \sum_j T_{ij} \quad \forall i \neq j$, we estimate that if a worker of occupation i changes occupation, the probability of transitioning to occupation j is

$$P'_{ij} = \frac{T_{ij}}{T_i} \quad \forall i \neq j. \tag{2.1}$$

The CPS has scarce data on workers who switched jobs but remained in the same occupation. We assume that $P_{ii} = r \; \forall i$, where $r \in [0, 1]$. In Appendix I.1.1, we discuss alternative assumptions and show that the overall results do not change significantly. To estimate r we use the annual occupational mobility rate (i.e. the percentage of workers that switch occupations within a year (Groes et al. 2014)). We discuss this calibration in more detail in the next chapter. We use P'_{ij} and r to write the adjacency matrix of the occupational mobility network as follows

$$A_{ij} = \begin{cases} r & \text{if } i = j, \\ (1-r)P_{ij} & \text{if } i \neq j. \end{cases} \tag{2.2}$$

The occupational mobility network is weighted and directed with $n = 464$ nodes. Figure 4.1B contains a visualisation of the network which suggests that there is a rich structure underlying occupational transitions (Mealy et al. 2018), with occupations of similar classification clustering together. As a result, occupational mobility may be significantly more restricted than is commonly assumed in the most basic labour market models, as noted in Autor (2013), Petrongolo & Pissarides (2001).

In our model, A_{ij} reflects the relative preference with which a worker from occupation i chooses to apply to a vacancy of occupation j, rather than any other occupation j' (see Eq. (2.8)). However, this preference is not directly observable from data. To overcome this issue, we use the empirical network of occupational mobility, where A_{ij} is the probability that a worker from occupation i is hired in occupation j, conditional on the worker switching jobs (note that this implies the worker was hired).

While the empirical mobility network allows us to calibrate the heterogeneity in occupational mobility at a detailed level, the concepts are different; the relative preference with which a worker from occupation i applies to a job vacancy in j is different from the probability that a worker from occupation i, who is switching jobs, transitions to occupation j. However, the former is not directly observable from data. To overcome this issue, we use the empirical network of occupational mobility as indicative of the preference with which a worker from occupation i applies to a job vacancy in j. A caveat is that since the odds of a worker being hired do not uniquely depend on the preference with which workers apply to job vacancies, the transitions of workers observed in our model do not perfectly match the empirically observed transitions. Though the matching between the transitions in our model and the empirical ones is not perfect, they are significantly similar – the Pearson correlation between them is 0.97.

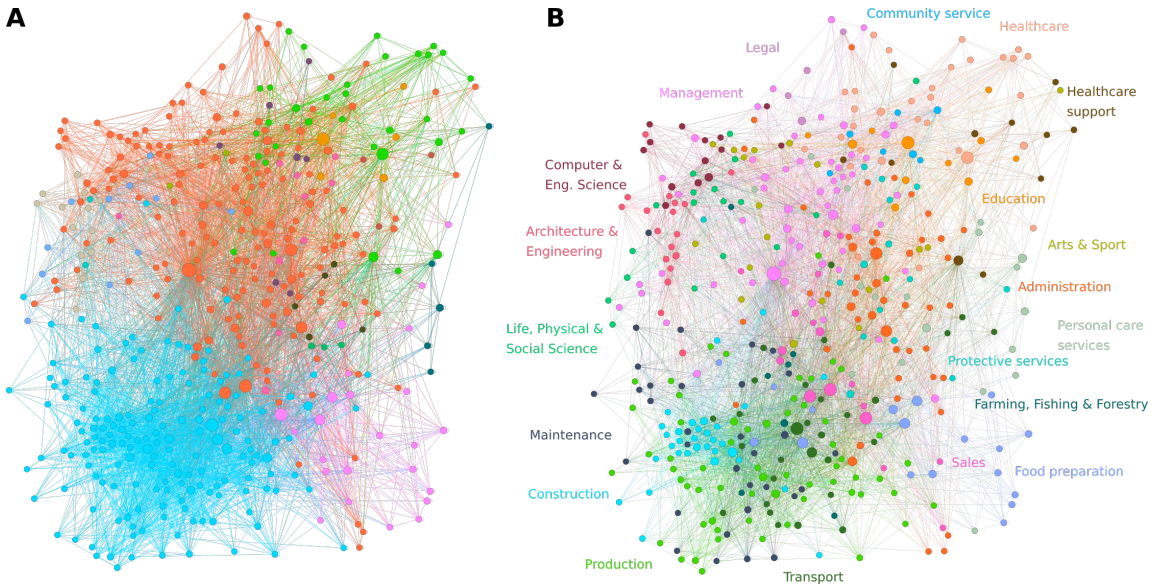


Figure 2.1: **The occupational mobility network.** Panel (**A**) shows the occupational mobility network where the colour of nodes indicates the community they belong to. In panel (**B**) the colour of the nodes indicates the classification of the occupation according to the ACS. In both panels, the size of the nodes is proportional to the logarithm of the number of employees in each occupation.

2.3 A network model of the labour market.

Our model is designed to understand the dynamics of unemployment at the occupation level. The flow of workers on the network (i.e. workers changing occupations) is described by a set of discrete-time stochastic processes for employment, unemployment, and vacancies in each occupation i. The agents in our model are workers, who can be either employed or unemployed. All workers have only one occupation in each time step, but they can switch occupations. We assume that workers are perfectly geographically mobile, and we neglect wage pressure. The set of possible occupations is fixed, and the occupation of a worker is defined as the occupation in which she was last employed. At any given time t, the number of workers employed in occupation i is $e_{i,t}$, the number of unemployed workers is $u_{i,t}$, and the number of open job vacancies is $v_{i,t}$. The number of workers that are separated (i.e. fired) is $\omega_{i,t}$ and the number of newly opened vacancies is $\nu_{i,t}$. The labour flow $f_{ij,t+1}$ is the number of workers hired in occupation j who were previously in occupation i. Using these notations, we can

write

$$e_{i,t+1} = e_{i,t} - \underbrace{\omega_{i,t+1}}_{\text{separated workers}} + \underbrace{\sum_j f_{ji,t+1}}_{\text{hired workers}}, \tag{2.3}$$

$$u_{i,t+1} = u_{i,t} + \underbrace{\omega_{i,t+1}}_{\text{separated workers}} - \underbrace{\sum_j f_{ij,t+1}}_{\text{transitioning workers}}, \tag{2.4}$$

$$v_{i,t+1} = v_{i,t} + \underbrace{\nu_{i,t+1}}_{\text{opened vacancies}} - \underbrace{\sum_j f_{ji,t+1}}_{\text{hired workers}}. \tag{2.5}$$

These equations express conservation laws stating that the change in each variable is equal to the difference between inflow and outflow. Equation (2.3) states that the change in employment is equal to the number of workers that are hired minus the number of workers that are separated. Similarly, Eq. (2.4) states that the change in unemployment is equal to the number of workers who are separated minus the number who are hired (transitioning workers). Finally, Eq. (2.5) states that the change in the number of vacancies is equal to the number of vacancies created minus the number of workers hired to fill existing vacancies. Figure 2.2 is a flow chart that makes the transitions explicit from the perspective of a worker and from the perspective of a job vacancy.

Our model builds on Mortensen and Pissarides' work (Pissarides 1985, Mortensen 1982), who proposed the use of a matching function between workers and vacancies. In our model, the labour flow $f_{ij,t+1}$ corresponds to a new matching function, which has the advantage of being more granular and motivated by the stochastic process illustrated in Fig. 2.2.

We denote occupation-specific stochastic variables by lower-case letters and aggregate quantities by upper case letters (e.g. total unemployment is $U_t = \sum_i u_{i,t}$) and use bold font for vectors (i.e. the ith element of $\mathbf{u}_t$ is $u_{i,t}$). We denote realised values of stochastic variables by a hat above e.g. $\hat{e}_{i,t}$ is a realisation of the stochastic variable $e_{i,t}$. We assume that the time step duration is long enough for workers to transition between occupations, but too short for workers to change their employment status more than once (see Chapter 3 for more details on time step calibration). That is, a worker is not allowed to switch her status from employed to unemployed and then back to employed in a single time step. Likewise, a vacancy cannot be opened and filled within the same time step.

We assume that the number of separated workers $\omega_{i,t+1}$ and the number of opened vacancies at $t+1$, $\nu_{i,t+1}$, follow a binomial processes of the form,

$$\omega_{i,t+1} \sim \text{Bin}(\hat{e}_{i,t}, \pi_{u,i,t}), \tag{2.6}$$

$$\nu_{i,t+1} \sim \text{Bin}(\hat{e}_{i,t}, \pi_{v,i,t}), \tag{2.7}$$

where $\text{Bin}(m, p)$ denotes a binomial distribution with m trials and success probability p. The success probabilities $\hat{\pi}_{u,i,t}$ and $\hat{\pi}_{v,i,t}$ depend on the imbalance of supply and demand for labour and play a key role in the dynamics. We will first complete an overview of the search and matching process and return in a moment to specify $\pi_{u,i,t}$ and $\pi_{v,i,t}$.

Search and matching. We denote the number of workers that switch occupation from i to j at time $t+1$ by $f_{ij,t+1}$ and refer to it as the flow of workers. The flow of workers depends on the structure of the occupational mobility network, the number of vacancies and unemployed workers, and the processes of job search and job matching. For the search and matching process we use the urn-ball matching function discussed in Chapter 1. To facilitate the analysis of the model, we make the simplifying assumption that only unemployed workers apply for jobs, and that each worker makes exactly one application. We denote the probability that a worker in occupation i applies to a vacancy in occupation j at time $t+1$ by $q_{ij,t+1}$, and assume that it is proportional to the number of vacancies in occupation j and the empirical transition probability between the two occupations A_{ij}. This leads to the following functional form

$$q_{ij,t+1} = \frac{\hat{v}_{j,t} A_{ij}}{\sum_k \hat{v}_{k,t} A_{il}}, \tag{2.8}$$

where in the numerator we have the number of vacancies of occupation j multiplied by the empirical transition probability A_{ij}. In the denominator we have a normalizing factor so that $\sum_j q_{ij,t+1} = 1$. After all workers have placed their application, for each vacancy one applicant chosen uniformly at random is hired. Some vacancies may not receive applications, in which case no one is hired, and the job vacancy remains open. At the next time step, the model repeats itself as illustrated in Fig. 2.2. In the next Section 2.4, we present analytical approximations to the dynamics of the model and make more explicit how the flow of workers depends on the structure of the occupational mobility network.

Note that we use the empirical transition probability as a proxy for the relative preference with which a worker from occupation i applies to a job vacancy in j. We

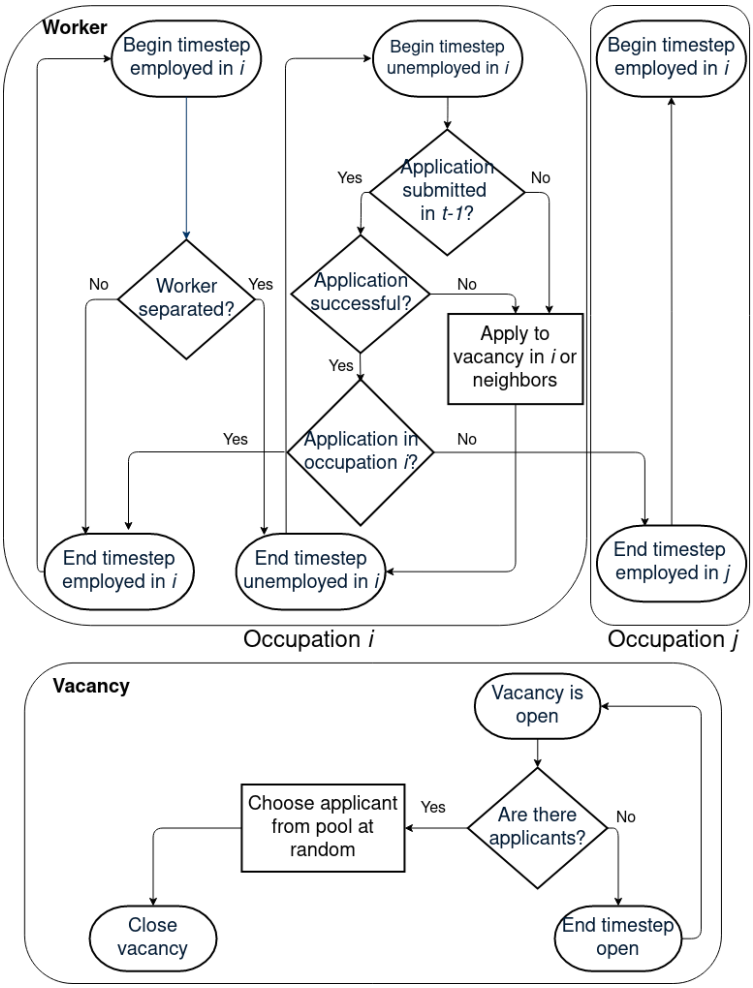


Figure 2.2: **Flow chart illustrating the possible transitions of workers and job vacancies during a given time step.** *Top* panel: transitions of a worker. *Bottom* panel: The transitions of a job vacancy. Note that vacancies created in the current time step do not accept job applications until the following time step.

cannot use the actual probability with which a worker *applies* to a job vacancies of another occupation since this is not recorded in the census data.

Supply and demand for labour. Workers move across the occupational mobility network in response to shifts in labour demand. These shifts are determined by the success probability $\pi_{u,i,t}$ of the binomial process for separating workers in Eq. (2.6) and the success probability $\pi_{v,i,t}$ for the binomial process for creating vacancies in Eq. (2.7). We break each of these into two separate random processes. The first is a *spontaneous process* (or state-independent) and the second is a *state-dependent process*.

In the spontaneous process, workers are separated and vacancies are opened at random, independent of the state of the system. For simplicity here we assume that the separation and opening rates are the same for all occupations. For any given occupation, the spontaneous probability that a given worker is separated at any given time is δ_u, and the spontaneous probability that a vacancy opens is δ_v times the number of workers in that occupation.

The state-dependent process drives the labour demand reallocation by adjusting the *realised* labour demand towards the *target* labour demand. The target labour demand $d^\dagger_{i,t}$ is the desired quantity of labour for occupation i at time t. This is imposed externally and allows us to impose automation shocks (or other reallocation shocks) as a function of time. The realised labour demand, in contrast, is a time-dependent variable corresponding to the sum of the number of employed workers plus the number of job vacancies in a given occupation, i.e.

$$d_{i,t} = e_{i,t} + v_{i,t}.$$

The difference between the realised and target demand can be attributed to supply factors, e.g. when there are more employed workers in an occupation than the target demand requires, or to demand factors, e.g. when vacancies are scarcer than what the target demand dictates. Opening vacancies and separating workers allow the realised labour demand to adjust: The realised labour demand for an occupation i increases when vacancies of occupation i open and decreases when workers of occupation i are separated.

In this state-dependent process we equilibrate the realised labour demand $d_{i,t}$ to the target demand $d^\dagger_{i,t}$ by opening more vacancies if $d_{i,t} < d^\dagger_{i,t}$ or separating workers if $d_{i,t} > d^\dagger_{i,t}$. To do so, we define an additional occupation-specific probability $\alpha_{u,i,t}$ that a worker from occupation i is separated at time t and an additional occupation-specific probability $\alpha_{v,i,t}$ that a vacancy in occupation i opens. Both of these probabilities are functions of time and depend on the difference between the realised labour demand. We assume the following functional forms

$$\alpha_{u,i,t} = \gamma_u \frac{\max\left\{0, d_{i,t} - d^\dagger_{i,t}\right\}}{e_{i,t}}, \tag{2.9}$$

$$\alpha_{v,i,t} = \gamma_v \frac{\max\left\{0, d^\dagger_{i,t} - d_{i,t}\right\}}{e_{i,t}}, \tag{2.10}$$

where γ_u and γ_v are parameters that determine the speed of adjustment and are in the interval $[0, 1]$. The α's are probabilities, and so must satisfy $0 \leq \alpha_{u,i,t} \leq 1$ and

$0 \leq \alpha_{v,i,t} \leq 1$[1]. For the purposes of this thesis, we let the rates for separations and vacancies be the same, i.e. $\gamma_u = \gamma_v = \gamma$. For a full description of how we calibrated parameters and set initial conditions we refer the reader to Chapter 3.

All of the processes corresponding to δ_u, δ_v, $\alpha_{u,i,t}$ and $\alpha_{v,i,t}$ are independent. Thus the probability that a worker in occupation i is *not* separated from her job is $(1-\delta_u)(1-\alpha_{u,i,t})$. This means that the probability that a worker *is* separated is given by

$$\pi_{u,i,t} = 1 - (1-\delta_u)(1-\alpha_{u,i,t}) = \delta_u + \alpha_{u,i,t} - \delta_u \alpha_{u,i,t}, \tag{2.11}$$

where the negative term on the right hand side avoids counting a worker as separated twice. Similarly, for each employed worker in occupation i, the probability that a vacancy opens is

$$\pi_{v,i,t} = \delta_v + \alpha_{v,i,t} - \delta_v \alpha_{v,i,t}. \tag{2.12}$$

This completes our specification of the model. In Table 2.1, we show a summary of all the variables and parameters.

Main variables	**Description**
$e_{i,t}$	Number of employed workers at time t in occupation i.
$u_{i,t}$	Number of unemployed workers at time t which were last employed in occupation i.
$v_{i,t}$	Number of job vacancies at time t of occupation i.
Other variables	
$d_{i,t}$	Realised labour demand at time t of occupation i. ($d_{i,t} = e_{i,t} + v_{i,t}$).
$q_{ij,t}$	The probability that an unemployed worker from occupation i applies to a job vacancy of occupation j at time t.
$p_{j,t}$	The probability that a job application sent to a vacancy of occupation j is successful.
$s_{ij,t}$	The number of job applicants occupation j receives from workers of occupation i at time t.
$s_{j,t}$	The number of job applicants occupation j receives at time t.
$f_{ij,t}$	Flow of unemployed workers of i to be employed at occupation j.
$u_{i,t}^{(k)}$	The number of unemployed workers of occupation i that have spent exactly k time steps unemployed at time t.

[1]Strictly speaking, one could impose a reallocation shock so large that $|d_{i,t} - d_{i,t}^{\dagger}| > e_{i,t}/\gamma$ making one of the α's greater than one. This would only happen in exceptional circumstances and is very unlikely, nevertheless in such a case we simply set $\alpha = 1$.

$\pi_{u,i,t}$	Occupation-specific probability that a worker of occupation i.
$\pi_{v,i,t}$	Occupation-specific probability that a vacancy of occupation i opens, per worker employed in occupation i.
$\omega_{i,t}$	Number of workers from occupation i that separated from their jobs at time t. Drawn from Binomial distribution $\text{Bin}(e_{i,t}, \pi_{u,i,t})$.
$\nu_{i,t}$	Number of vacancies of occupation i opened at time t. Drawn from Binomial distribution $\text{Bin}(e_{i,t}, \pi_{v,i,t})$.
Parameters	
δ_u	Rate at which employed workers are separated due to the spontaneous process.
δ_v	Rate at which employed vacancies are opened due to spontaneous process.
γ	Rate at which employed workers and vacancies are separated or opened due to the market adjusting towards the target demand.
τ	Number of time steps after which an unemployed worker is considered long-term unemployed.
r	Weights of the self-loops of the occupational mobility network.
A	Adjacency matrix of the occupational mobility network.
$d_i^\dagger$	Post-shock target labour demand. Number of workers needed at occupation i after the shock is complete.
$d_{i,t}^\dagger$	Target labour demand of occupation i at time t.
Δt	Duration of a time step in units of weeks.

Table 2.1: Variables and parameters.

2.4 Deterministic approximation for large populations

To understand the system we focus on the expected value of the stochastic variables $e_{i,t}$, $u_{i,t}$ and $v_{i,t}$. Although the workers and employers follow simple rules in our model, when the number of workers L is large, running the simulation and computing the expected value is computationally costly. However, when L is large, we can take advantage of the law of large numbers and multivariate Taylor expansions to approximate the system's behaviour in terms of expected values. This approximation is good for most purposes, makes it easier to understand the mechanics of the model, and is faster to simulate which is useful for exploring the parameter space.

To keep the notation compact, we often denote approximations to the expected values by a bar above the variable, e.g.

$$\bar{u}_{i,t+1} \approx \mathbb{E}\left[u_{i,t+1} | \mathbf{u}_{i,t} = \hat{\mathbf{u}}_{i,t}, \mathbf{v}_{i,t} = \hat{\mathbf{v}}_{i,t}, \mathbf{e}_{i,t} = \hat{\mathbf{e}}_{i,t}\right]. \tag{2.13}$$

We propose the following deterministic dynamical as an approximation to the expected value of $e_{i,t}$, $u_{i,t}$ and $v_{i,t}$.

$$\bar{e}_{i,t+1} = \bar{e}_{i,t} - \underbrace{\left(\delta_u \bar{e}_{i,t} + (1-\delta_u)\gamma \max\left\{0, \bar{d}_{i,t} - d^{\dagger}_{i,t}\right\}\right)}_{\text{separated workers}} + \underbrace{\sum_j \bar{f}_{ji,t+1}}_{\text{hired workers}}, \tag{2.14}$$

$$\bar{u}_{i,t+1} = \bar{u}_{i,t} + \underbrace{\left(\delta_u \bar{e}_{i,t} + (1-\delta_u)\gamma \max\left\{0, \bar{d}_{i,t} - d^{\dagger}_{i,t}\right\}\right)}_{\text{separated workers}} - \underbrace{\sum_j \bar{f}_{ij,t+1}}_{\text{transitioning workers}} \tag{2.15}$$

$$\bar{v}_{i,t+1} = \bar{v}_{i,t} + \underbrace{\left(\delta_v \bar{e}_{i,t} + (1-\delta_v)\gamma \max\left\{0, d^{\dagger}_{i,t} - \bar{d}_{i,t}\right\}\right)}_{\text{opened vacancies}} - \underbrace{\sum_j \bar{f}_{ji,t+1}}_{\text{hired workers}}, \tag{2.16}$$

where

$$\bar{f}_{ij,t+1} = \frac{\bar{u}_{i,t}\bar{v}^2_{j,t}A_{ij}(1 - e^{-\bar{s}_{j,t+1}/\bar{v}_{j,t}})}{\bar{s}_{j,t+1}\sum_k \bar{v}_{k,t}A_{ik}}, \tag{2.17}$$

and

$$\bar{s}_{j,t+1} = \sum_i \frac{\bar{u}_{i,t}\bar{v}_{j,t}A_{ij}}{\sum_k \bar{v}_{k,t}A_{ik}}, \tag{2.18}$$

is the expected number of applications submitted to vacancies of occupation j. Since the subindex i denotes occupations, this dynamical system is composed of $3n$ equations.

In the following subsection we present a mathematical analysis that motivates these approximations. In the next subsection we show with simulations that our approximations work considerably well in the limit of a large number of agents. We conclude that we can use these approximations to explore the behaviour of our model.

2.4.1 Mathematical derivation

In this section, we derive approximations for the dynamics of the agent-based model; these approximations are based on deriving expectations for time step $t+1$, conditional on the state of the system at time t. A more in-depth mathematical analysis would then need to show that the approximation can work iteratively between time

steps. The mathematical arguments provided in this section are not a mathematical proof that the expected value of the variables $e_{i,t}$, $u_{i,t}$, and $u_{i,t}$ converge to the proposed approximations. However, as we show in the following subsection with computer simulations, these approximations work well in the limit of a large number of agents.

We focus on approximations within a time step, and condition calculations on the realisation of the random variables at the previous time step (i.e. we have hats on all variables at time t denoting realised values). We split this derivation into two. First, we discuss the approximation for the separation and opening of vacancies (the terms with the maximum functions) in Eqs. (2.14 – 2.16). We then discuss the approximation for the flow of workers.

Separation and opening of vacancies. Here we show that in the limit of a large number of agents

$$\bar{\omega}_{i,t+1} = \delta_u \hat{e}_{i,t} + (1-\delta_u)\gamma \max\left\{0, \hat{d}_{i,t} - d^{\dagger}_{i,t}\right\} \approx \mathbb{E}[\omega_{i,t+1} | e_{i,t} = \hat{e}_{i,t}, d_{i,t} = \hat{d}_{i,t}], \quad (2.19)$$

and that

$$\bar{\nu}_{i,t+1} = \delta_v \hat{e}_{i,t} + (1-\delta_v)\gamma \max\left\{0, d^{\dagger}_{i,t} - \hat{d}_{i,t}\right\} \approx \mathbb{E}[\nu_{i,t+1} | e_{i,t} = \hat{e}_{i,t}, d_{i,t} = \hat{d}_{i,t}]. \quad (2.20)$$

The separation and opening of vacancies depends on the difference between the realised demand $d_{i,t+1} = e_{i,t+1} + v_{i,t+1}$ and the target demand $d^{\dagger}_i$. It follows from Eqs. (2.3 – 2.5) that the realised demand is given by

$$d_{i,t+1} = \hat{d}_{i,t} - \omega_{i,t+1} + \nu_{i,t+1}, \quad (2.21)$$

where $\omega_{i,t}$ and $\nu_{i,t}$ are binomial random variables of $\hat{e}_{i,t}$ draws and success probability $\pi_{u,i,t}$ and $\pi_{v,i,t}$ respectively. In the limit of a large number of agents we can approximate their distributions to Normal distributions, so that

$$\nu_{i,t+1} - \omega_{i,t+1} = \hat{e}_{i,t}(\pi_{v,i,t} - \pi_{u,i,t}) - \eta_{i,t+1}, \quad (2.22)$$

where

$$\eta_{i,t+1} \sim \mathcal{N}\left(0, \left(\pi_{u,i,t}(1-\pi_{u,i,t}) + \pi_{v,i,t}(1-\pi_{v,i,t})\right) e_{i,t}\right). \quad (2.23)$$

Since we are conditioning on realised values of the variables for time t, it follows from Eqs. (2.21), (2.22), and (2.23) that in the limit of a larger number of agents $d_{i,t}$ is normally distributed. This means that

$$\mathbb{E}\left[\max\left\{0, d^{\dagger}_i - \hat{d}_{i,t}\right\}\right] = \max\left\{0, d^{\dagger}_i - \hat{d}_{i,t}\right\} + \epsilon, \quad (2.24)$$

where

$$\epsilon = \sqrt{\frac{2}{\pi}\left(\pi_{u,i,t}(1-\pi_{u,i,t}) + \pi_{v,i,t}(1-\pi_{v,i,t})\right) e_{i,t}}. \tag{2.25}$$

To obtain the above equations we have used the fact that $\max\left\{0, b\right\} = \frac{b}{2} + \frac{|b|}{2}$, that the expected value of the absolute value of a normally distributed function with variance σ^2 is $\sigma\sqrt{\frac{2}{\pi}}$, and Eq. (2.23). Similarly,

$$\mathbb{E}\left[\max\left\{0, \hat{d}_{i,t} - d_i^\dagger\right\}\right] = \max\left\{0, \hat{d}_{i,t} - d_i^\dagger\right\} + \epsilon. \tag{2.26}$$

It follows from Eqs. (2.6 – 2.7), (2.9–2.12) and (2.24–2.26) that

$$\mathbb{E}\left[\omega_{i,t+1} | e_{i,t} = \hat{e}_{i,t}, d_{i,t} = \hat{d}_{i,t}\right] = \delta_u \hat{e}_{i,t} + (1-\delta_u)\gamma \max\left\{0, \hat{d}_{i,t} - d_{i,t}^\dagger\right\} + (1-\delta_u)\gamma\epsilon, \tag{2.27}$$

and that,

$$\mathbb{E}\left[\nu_{i,t+1} | e_{i,t} = \hat{e}_{i,t}, d_{i,t} = \hat{d}_{i,t}\right] = \delta_v \hat{e}_{i,t} + (1-\delta_v)\gamma \max\left\{0, d_{i,t}^\dagger - \hat{d}_{i,t}\right\} + (1-\delta_v)\gamma\epsilon. \tag{2.28}$$

The first two terms scale linearly with the employment of occupations, while ϵ scales with the square root of the employment of occupations. This means that in the limit of a large number of agents we can make the approximation

$$\mathbb{E}\left[\omega_{i,t+1} | e_{i,t} = \hat{e}_{i,t}, d_{i,t} = \hat{d}_{i,t}\right] \approx \bar{\omega}_{i,t+1} = \delta_u \hat{e}_{i,t} + (1-\delta_u)\gamma \max\left\{0, \hat{d}_{i,t} - d_{i,t}^\dagger\right\}$$

and that

$$\mathbb{E}\left[\nu_{i,t+1} | e_{i,t} = \hat{e}_{i,t}, d_{i,t} = \hat{d}_{i,t}\right] \approx \bar{\nu}_{i,t+1} = \delta_v \hat{e}_{i,t} + (1-\delta_v)\gamma \max\left\{0, d_{i,t}^\dagger - \hat{d}_{i,t}\right\}.$$

Note that since we are interested in calculating unemployment, vacancy, and employment rates (i.e. we will divide by a factor proportional to L) we are only interested in the terms that scale linearly with L.

Flow of workers approximation. In this subsection we derive Eq. (2.17), which is

$$\bar{f}_{ij,t+1} = \frac{\hat{u}_{i,t}\hat{v}_{j,t}^2 A_{ij}(1 - e^{-\bar{s}_{j,t+1}/\hat{v}_{j,t}})}{\bar{s}_{j,t+1}\sum_k \hat{v}_{k,t}A_{ik}} \approx \mathbb{E}\left[f_{ij,t+1} | \mathbf{u}_{i,t} = \hat{\mathbf{u}}_{i,t}, \mathbf{v}_{i,t} = \hat{\mathbf{v}}_{i,t}, \mathbf{e}_{i,t} = \hat{\mathbf{e}}_{i,t}\right],$$

where

$$\bar{s}_{j,t+1} = \sum_i \frac{\hat{u}_{i,t}\hat{v}_{j,t}A_{ij}}{\sum_k \hat{v}_{k,t}A_{ik}}. \tag{2.29}$$

We denote the number of workers from occupation i that apply to occupation j at time t by $s_{ij,t}$. Given that only unemployed workers apply for jobs and that $q_{ij,t}$ is

the probability that a worker from occupation i applies to a vacancy of occupation j the expected number of applications submitted from occupation i to occupation j is

$$\mathbb{E}\left[s_{ij,t+1}|u_{i,t} = \hat{u}_{i,t}\right] = \hat{u}_{i,t}q_{ij,t+1}. \tag{2.30}$$

The labour flow $f_{ij,t+1}$ is equal to the number of workers from occupation i applying to occupation j, $\hat{s}_{ij,t+1}$, multiplied by the probability $p_{j,t+1}$ that each application is successful. (All applications are accepted with uniform probability, so p does not depend on i). The expected value is

$$\mathbb{E}\left[f_{ij,t+1}|\mathbf{u}_{i,t} = \hat{\mathbf{u}}_{i,t}, \mathbf{v}_{i,t} = \hat{\mathbf{v}}_{i,t}, \mathbf{e}_{i,t} = \hat{\mathbf{e}}_{i,t}\right] = \mathbb{E}\left[s_{ij,t+1}p_{j,t+1}|\mathbf{u}_{i,t} = \hat{\mathbf{u}}_{i,t}, \mathbf{v}_{i,t} = \hat{\mathbf{v}}_{i,t}, \mathbf{e}_{i,t} = \hat{\mathbf{e}}_{i,t}\right]. \tag{2.31}$$

Letting the total number of applications $s_{j,t+1}$ to occupation j be

$$s_{j,t+1} = \sum_k s_{kj,t+1}, \tag{2.32}$$

the fraction $p_{j,t+1}$ of successful applications is the ratio of the number of vacancies $m_{j,t+1}$ that successfully match to the total number of applications, i.e.

$$p_{j,t+1} = m_{j,t+1}/s_{j,t+1}. \tag{2.33}$$

Matching. To simplify upcoming calculations that require derivatives, it is convenient to express the number of matches $m_{j,t+1}$ in terms of the exponential function. This is a standard approximation, and the derivation we present is based on Stevens (2007). Recall that $m_{j,t+1}$ is the number of vacancies that successfully match with a job applicant. Since employers hire a worker uniformly at random from the pool of applicants, then $m_{j,t+1}$ is equal to the number of job applications that receive at least one job application.

An unemployed worker who applies for a job in occupation j with $\hat{v}_{j,t}$ vacancies, will apply to a particular vacancy with probability $\frac{1}{\hat{v}_{j,t}}$. Thus the probability that the worker does *not* send her application to that vacancy is $1 - \frac{1}{\hat{v}_{j,t}}$. For $\hat{s}_{j,t+1}$ unemployed workers sending applications to occupation j, the probability that a particular vacancy does not receive an application is $(1 - \frac{1}{\hat{v}_{j,t}})^{\hat{s}_{j,t+1}}$. Since each vacancy receiving an application hires one worker, the expected number $\bar{m}_{j,t+1}$ of successful job applications is

$$\bar{m}_{j,t+1} = \hat{v}_{j,t}\left(1 - \left(1 - \frac{1}{\hat{v}_{j,t}}\right)^{\hat{s}_{j,t+1}}\right). \tag{2.34}$$

Using the approximation that $(1 - \frac{1}{x})^y \approx e^{-y/x}$ for large x and y we obtain

$$\bar{m}_{j,t+1} = \hat{v}_{j,t}(1 - e^{-\hat{s}_{j,t+1}/\hat{v}_{j,t}}). \tag{2.35}$$

Taylor approximation for flow of workers. It follows from Eqs. (2.31) and (2.33) that the flow of workers depends on $s_{ij,t+1}$, $m_{j,t+1}$ and $s_{j,t+1}$. These variables are not independent, but here we show that, in the large L limit, we can approximate $\mathbb{E}\left[f_{ij,t+1}|\mathbf{u}_{i,t}=\hat{\mathbf{u}}_{i,t},\mathbf{v}_{i,t}=\hat{\mathbf{v}}_{i,t},\mathbf{e}_{i,t}=\hat{\mathbf{e}}_{i,t}\right]$ by

$$\bar{f}_{ij,t+1}=\bar{s}_{ij,t+1}\hat{v}_{j,t}\frac{\left(1-e^{-\bar{s}_{j,t+1}/\hat{v}_{j,t}}\right)}{\bar{s}_{j,t+1}}. \tag{2.36}$$

To derive this approximation it is useful to define

$$s_{j\backslash i,t+1}\equiv\sum_{k\neq i}s_{kj,t+1}, \tag{2.37}$$

which is the number of applications occupation j receives from all unemployed workers except those from occupation i. Note that $s_{j,t+1}=s_{j\backslash i,t+1}+s_{ij,t+1}$. Using this fact, and Eq. (2.31) we define the following multivariate form of the flow of workers

$$\begin{aligned}\mathbb{E}\left[f_{ij,t+1}|\mathbf{u}_{i,t}=\hat{\mathbf{u}}_{i,t},\mathbf{v}_{i,t}=\hat{\mathbf{v}}_{i,t},\mathbf{e}_{i,t}=\hat{\mathbf{e}}_{i,t}\right]&=g(s_{ij,t+1},s_{j\backslash i,t+1})\\&\equiv s_{ij,t+1}\hat{v}_{j,t}\frac{1-\exp\left\{-\left(\frac{s_{ij,t+1}+s_{j\backslash i,t+1}}{\hat{v}_{j,t}}\right)\right\}}{s_{ij,t+1}+s_{j\backslash i,t+1}}.\end{aligned} \tag{2.38}$$

This definition will allow us to do a multivariate Taylor expansion of the function g around the expected value of $s_{ij,t+1}$ and $s_{j\backslash i,t+1}$.

Recall that, for fixed i, the random variables $s_{ij,t+1}$ follow a multinomial distribution with $\hat{u}_{i,t}$ trials and probabilities $q_{ij,t+1}$ for $j=1,...,n$. This means that $s_{ij,t+1}$ and $s_{ik,t+1}$ are drawn from the same realisation of the multinomial distribution, and are therefore correlated. However, $s_{ij,t+1}$ and $s_{kj,t+1}$ are drawn from different realisations (and distributions); thus they are independent.

It follows from the fact that $s_{ij,t+1}$ and $s_{kj,t+1}$ are independent and from Eq. (2.37), that $s_{ij,t+1}$ is independent from $s_{j\backslash i,t+1}$. Furthermore, in the limit of a large number of agents, $\hat{u}_{i,t}$ is large, and the standard deviation of $s_{ij,t+1}$ is small in comparison to the average. The same is true for $s_{j\backslash i,t+1}$; we therefore expand $g(s_{ij,t+1},s_{j\backslash i,t+1})$ in a Taylor series around the expected value of $s_{j,t+1}$ and $s_{j\backslash i,t+1}$ as follows,

$$\begin{aligned}\bar{g}(s_{ij,t+1},s_{j\backslash i,t+1})&=g(\bar{s}_{ij,t+1},\bar{s}_{j\backslash i,t+1})\\&+\frac{1}{2}\frac{\partial^2}{\partial s_{ij,t+1}^2}\left(g(\bar{s}_{ij,t+1},\bar{s}_{j\backslash i,t+1})\right)\mathrm{Var}[s_{ij,t+1}]\\&+\frac{1}{2}\frac{\partial^2}{\partial s_{j\backslash i,t+1}^2}\left(g(\bar{s}_{ij,t+1},\bar{s}_{j\backslash i,t+1})\right)\mathrm{Var}[s_{j\backslash i,t+1}]+\ldots.\end{aligned} \tag{2.39}$$

Next, we now show that, in the limit of a large number of agents, the second and third terms are negligible in comparison to the first term. Exclusively for this derivation, we introduce the notation $v \equiv \hat{v}_{j,t}$, $x \equiv s_{ij,t+1}$ and $y \equiv s_{j\setminus i,t+1}$. We denote the expected value of a variable x by μ_x and the variance by σ_x^2, so that

$$\begin{aligned}
\mu_x &= \hat{u}_{i,t} q_{ij,t+1}, && (2.40)\\
\mu_y &= \sum_{k\neq i} \hat{u}_{k,t} q_{kj,t+1}, && (2.41)\\
\sigma_x^2 &= \hat{u}_{i,t} q_{ij,t+1}(1 - q_{ij,t+1}), && (2.42)\\
\sigma_y^2 &= \sum_{k\neq i} \hat{u}_{k,t} q_{kj,t+1}(1 - q_{kj,t+1}), && (2.43)
\end{aligned}$$

where we have used the formula for the expected value and variance of (independent) multinomial distributions. Note that μ_x, μ_y, σ_x^2, and σ_y^2 are proportional to the unemployment variables and therefore scale linearly with the number of agents L.

Using this notation and taking partial derivatives from Eq. (2.38), we obtain

$$\begin{aligned}
\bar{h}(x,y) &= h(\mu_x,\mu_y) + \sigma_x^2 \left[\left(\frac{v\mu_x}{(\mu_y+\mu_x)^3} - \frac{v}{(\mu_y+\mu_x)^2} \right) \left(1 - e^{-(\mu_x+\mu_y)/v}\right) \right.\\
&+ \left. \left(\frac{1}{\mu_y+\mu_x} - \frac{\mu_x}{(\mu_y+\mu_x)^2} \right) e^{-(\mu_x+\mu_y)/v} - \frac{1}{2}\frac{\mu_x}{v(\mu_y+\mu_x)} e^{-(\mu_x+\mu_y)/v} \right]\\
&+ \sigma_y^2 \left[\frac{v\mu_x(1 - e^{-(\mu_x+\mu_y)/v})}{(\mu_x+\mu_y)^3} - \frac{\mu_x}{(\mu_x+\mu_y)^2} e^{-(\mu_x+\mu_y)/v} \right.\\
&- \left. \frac{1}{2}\frac{\mu_x}{v(\mu_y+\mu_x)} e^{-(\mu_x+\mu_y)/v} \right] + O\left(\frac{1}{L}\right), \qquad (2.44)
\end{aligned}$$

Since μ_x, μ_y, σ_x^2, σ_y^2 and v scale linearly with L, in the limit of a large number of agents these five variables are of the same order of magnitude. The higher moments terms in the expansion are at most of the order of $\frac{1}{L}$ given that the n-th order partial derivatives are proportional to $\frac{1}{L^{n-1}}$. It follows from these observations and from Eqs. (2.38) and (2.44) that, in the limit of large number of agents, the first term of Eq. (2.44) scales with L, while the other terms do not scale with L. In other words, in the limit of a large number of agents, we can approximate

$$\bar{f}_{ij,t+1} = \bar{s}_{ij,t+1} \frac{\bar{v}_{j,t}^2 (1 - e^{-\bar{s}_{j,t+1}/\bar{v}_{j,t}})}{\bar{s}_{j,t+1}}, \qquad (2.45)$$

where we have recovered our original notation. The relative error of this approximation is inversely proportional to the number of agents i.e. the relative error of our approximation is

$$\left| \frac{\mathbb{E}\left[f_{ij,t+1} | \mathbf{u}_t, \mathbf{v}_t; A\right] - \bar{f}_{ij,t+1}}{\mathbb{E}\left[f_{ij,t+1} | \mathbf{u}_t, \mathbf{v}_t; A\right]} \right| \propto \frac{\kappa_{ij,t+1}}{L + \kappa_{ij,t+1}}, \qquad (2.46)$$

where

$$\begin{aligned}\kappa_{ij,t+1} &= \sigma_x^2\left[\left(\frac{v\mu_x}{(\mu_y+\mu_x)^3}-\frac{v}{(\mu_y+\mu_x)^2}\right)\left(1-e^{-(\mu_x+\mu_y)/v}\right)\right.\\ &+ \left.\left(\frac{1}{\mu_y+\mu_x}-\frac{\mu_x}{(\mu_y+\mu_x)^2}\right)e^{-(\mu_x+\mu_y)/v}-\frac{1}{2}\frac{\mu_x}{v(\mu_y+\mu_x)}e^{-(\mu_x+\mu_y)/v}\right]\\ &+ \sigma_y^2\left[\frac{v\mu_x(1-e^{-(\mu_x+\mu_y)/v})}{(\mu_x+\mu_y)^3}-\frac{\mu_x}{(\mu_x+\mu_y)^2}e^{-(\mu_x+\mu_y)/v}\right.\\ &- \left.\frac{1}{2}\frac{\mu_x}{v(\mu_y+\mu_x)}e^{-(\mu_x+\mu_y)/v}\right]+O\left(\frac{1}{L}\right), \end{aligned} \tag{2.47}$$

depends on the network structure, but does not scale with L. Therefore, when $L\to\infty$ the relative error is negligible.

Substituting Eqs. (2.8) and (2.30) into Eq. (2.45), we can write $\bar{f}_{ij,t+1}$ in terms of the adjacency matrix and the expected values of the state variables as

$$\bar{f}_{ij,t+1}=\frac{\hat{u}_{i,t}\hat{v}_{j,t}^2A_{ij}(1-e^{-\bar{s}_{j,t+1}/\hat{v}_{j,t}})}{\bar{s}_{j,t+1}\sum_k\hat{v}_{k,t}A_{ik}}, \tag{2.48}$$

where

$$\bar{s}_{j,t+1}=\sum_i\frac{\hat{u}_{i,t}\hat{v}_{j,t}A_{ij}}{\sum_k\hat{v}_{k,t}A_{ik}}. \tag{2.49}$$

We have derived our approximations within a time step, conditioning on the previous time step. In practice we do these approximations at each time step, i.e. we substitute $\hat{e}_{j,t}$ by $\bar{e}_{i,t}$, $\hat{u}_{j,t}$ by $\bar{u}_{i,t}$, and $\hat{v}_{j,t}$ by $\bar{v}_{i,t}$. However, the mathematical analysis needed to derive the error between time steps is out of the scope of this thesis. Instead, in Section 2.4.2 we show that this approximation works using simulations.

Unemployment spells and long-term unemployment. Finally, we use the dynamical system given by Eqs.(2.14 – 2.16) to derive the long-term unemployment. The expected number of workers with an unemployment spell of k steps for occupation i at time t is the expected number of workers with an unemployment spell of $k-1$ steps at the previous time step times the probability that a worker of occupation i is not hired. Thus the expected number of unemployed workers of occupation i with an unemployment spell of k time steps $u_{i,t+1}^{(k)}$ is given by the recursive relationship

$$\bar{u}_{i,t+1}^{(k)}=\bar{u}_{i,t}^{(k-1)}\left(1-\frac{\sum_j\bar{f}_{ji,t}}{\bar{u}_{i,t}}\right), \tag{2.50}$$

with $\bar{u}_{i,1}^{(1)}=\bar{\omega}_{i,1}=\bar{e}_{i,0}\pi_{u,i,t}$.

The U.S. Bureau of Labor Statistics defines long-term unemployed workers as those who have an unemployment spell of 27 or more weeks. Similarly, in our model the long-term unemployed workers are those who have been unemployed for τ or more time steps. In the next chapter we explain how we calibrate the time step duration so that $\tau = 27$ weeks. Using Eq. (2.50), we compute the expected number of long-term unemployed workers ($\bar{u}_{i,t+1}^{(\geq\tau)}$) by summing over all workers with an unemployment spell of τ or more time steps

$$\bar{u}_{i,t+1}^{(\geq\tau)} = \sum_{k=\tau}^{\infty} \bar{u}_{i,t+1}^{(k)}. \tag{2.51}$$

2.4.2 Simulation evidence

We show how our approximations compare with ten different simulations of the agent-based model at the aggregate and occupation level. We run these tests using the occupational mobility network, parameter values $\delta_u = 0.016$, $\delta_v = 0.012$, $\gamma = 0.16$, and $\tau = 4$ (i.e. the time step duration is roughly of 6.75 weeks), and an oscillating target demand. For the aggregate target demand, we assume that it is equal to the labour force L for the first 100 time steps (roughly 13 years), and then starts oscillating with an amplitude of $0.065L$ and period of 103 time steps (see top panel of Fig. 2.3). We assume no labour reallocation for this test, i.e. the share of the target demand across occupations is constant. In this chapter, we are interested in showing that our approximations match the simulations and do not delve into how we choose the target demand and the parameter values. In Chapter 3, we explain these choices in more detail and explore alternatives. Chapter 3 shows that our approximations also work for a target labour demand with labour reallocation.

The approximations we propose are for the limit of a large number of agents and would expect them to be smaller at the aggregate level than at the occupation level. Therefore, for our approximations to work at the occupation level, we need the approximations to work well at the aggregate level. In Fig. 2.3, we show that this is the case. In the bottom two panels of Fig. 2.3, we show that the unemployment and long-term unemployment rates match the simulations almost flawlessly at the aggregate level. At this level, it is difficult to see the fluctuations in the simulations, and to help the reader see these small fluctuations, we zoom into the first time steps (see top left corners of the bottom two panels in Fig. 2.3).

At the occupation level, the approximations also match the expected value of the simulations closely. Since we have 464 it is difficult to show the outcome for all occupations, so we focus the analysis on four example occupations. Bearing in mind

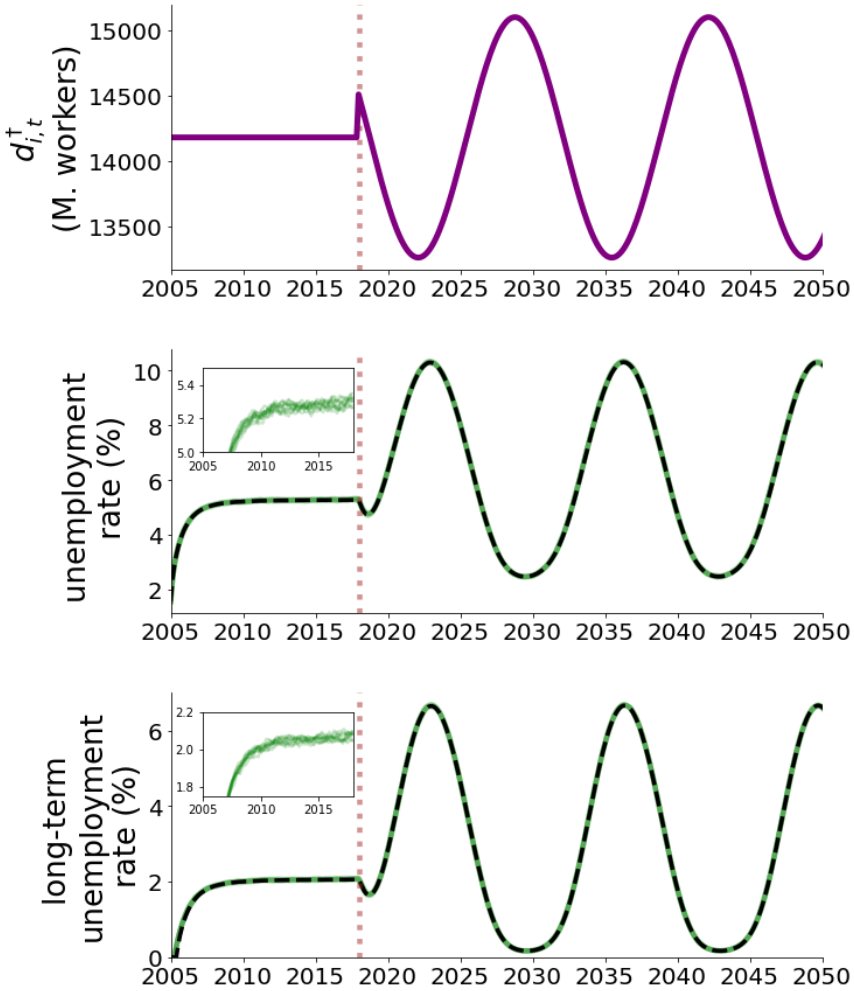


Figure 2.3: **Simulations and approximations at the aggregate level.** In the *top* panel we show the aggregate target labour demand, constant at the beginning and then oscillating according to a *sine* wave. The *middle* panel shows the unemployment rate and the *bottom* panel the long-term unemployment. The dashed vertical red lines denote the time at which the target demand starts oscillating. For the last two panels, green lines denote simulations and dashed black lines the numerical approximation. It is difficult to note the fluctuations in the simulations since we run the model with many agents (roughly 1.4M). To help the reader see the fluctuations we zoom into the first time steps of the model in the top left box of both panels.

that the errors and fluctuations depend on the number of employed workers in each occupation, we select two occupations among the five occupations with the lowest employment and two occupations with above-average employment. Namely, we focus on motion picture projectionists, brokerage clerks, postal service clerks, and electricians, whose employment is roughly 5300, 5900, 124, 200 million, and 632, 900 according to the 2016 Census (Flood et al. 2020). Since we run the numerical simulations with a hundredth of the real labour force, each occupation's target demand is 53, 59, 12, 400, and 60, 300, respectively. As shown in Fig. 2.4, our approximations match the average unemployment and long-term unemployment rate of each occupation. Noticeably, the fluctuations are much larger for motion picture projectionists and brokerage clerks, with low employment. Since our approximations work well at the occupation level when simulating 1.4 million workers, it gives us confidence that our approximations can work well for cities.

2.5 Stability: steady-state and fixed-point

In our model, the labour market interacts with the rest of the economy via the target demand $\mathbf{d}_t^\dagger$. In recessions, the target demand of most occupations decreases, while in recovery periods the target demand of most occupations increases. When a structural change happens, some occupations will decrease their target demand and others will increase it. If the rest of the economy is in a stable state and the target demand is constant (i.e. $\mathbf{d}_t^\dagger = \mathbf{d}^\dagger \ \forall t$), then we expect the labour market to be in a stable state too. In this section, we argue that this is indeed the case.

The agent-based model we derived in this section is a Markov chain. The state of the system at time $t+1$ is uniquely defined by the vectors of integer numbers $\mathbf{e}_{t+1}, \mathbf{u}_{t+1}$ and $\mathbf{u}_{t+1}$, which means that the system has a countable number of states. Furthermore, the stochastic evolution rules depend only on predefined parameters and on the state of the system at the previous time step given by $\mathbf{e}_t, \mathbf{u}_t$ and $\mathbf{u}_t$. Since the occupational mobility network is strongly connected, there is reason to believe that there is a set of states for which the Markov chain is ergodic and that the system converges to a steady-state. Although proving ergodicity is out of the scope of this thesis, we explore the model's behaviour using simulations and show that for the cases we study for all practical purposes, the system converges to a steady-state.

Eqs. (2.14 – 2.16) define a discrete-time dynamical system, which has fixed-point equations. We can derive a close-form solution to the fixed-point equations for the

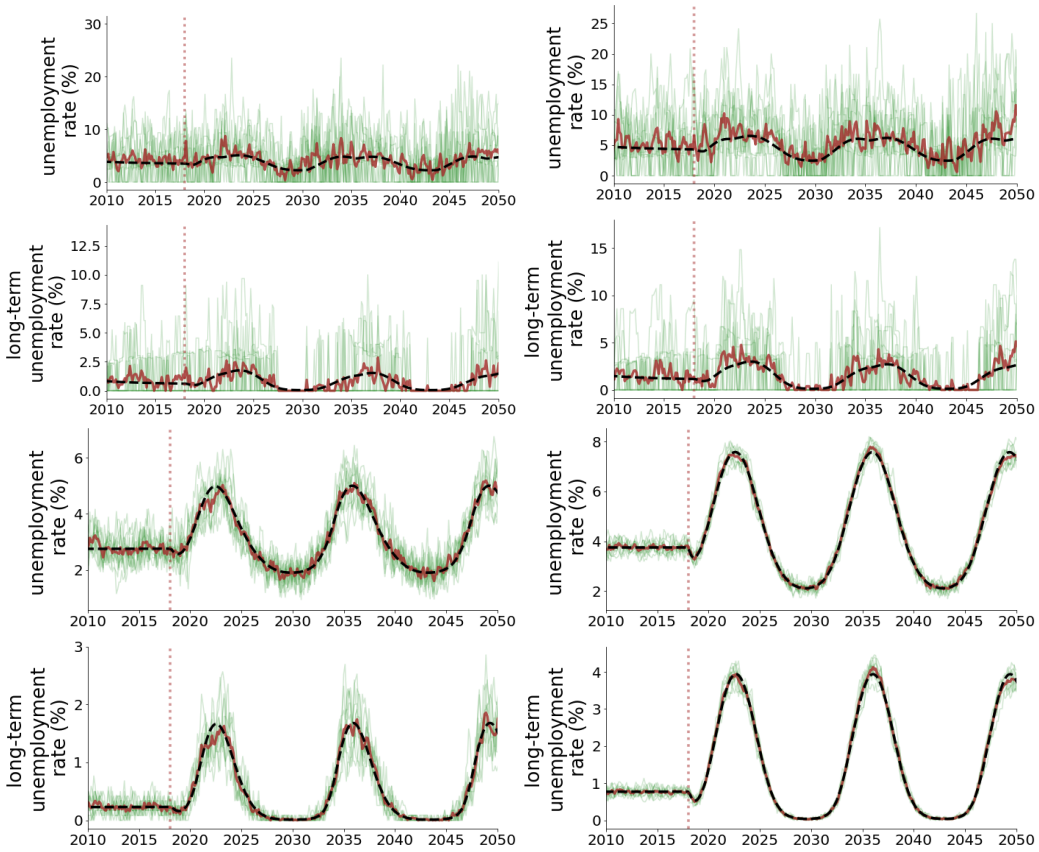


Figure 2.4: **Simulations and approximations at the occupation level.** In the *top left* panel motion picture projectionist, in the *top right* brokerage clerks, in the *bottom left* postal service clerks and in the *bottom right* electricians. The panels show the unemployment and long-term unemployment rates. The green lines denote simulations, the solid brown lines the mean across the ten simulations and dashed black lines the numerical approximation. The solid black lines closely match the brown lines.

simple case of a connected network, i.e. a fully connected network where each component is the adjacency matrix is equal to $1/n$. For more complicated cases, we solve equations numerically to find a solution. Determining the stability of a fixed-point in our system using analytical techniques is impractical since the system is composed of more than a thousand variables (464×3 to be precise) and since Eqs. (2.14 – 2.16) are not differentiable. However, we explore the stability numerically. Our exploration suggests that, for all initial conditions we consider, the fixed-point solution we find is stable. Furthermore, the fixed-point solution corresponds to the expected value of the steady-state the system reaches.

We also find that the fixed-point equations depend on parameters δ_u and δ_v, as well as the network structure and the target labour demand. Thus the network structure and the distribution of labour demand can substantially influence the fixed-point unemployment at both the occupational and the aggregate level.

2.5.1 Fixed-points

In this section, we show that the dynamical system given by Eqs. (2.14 – 2.16) has a fixed-point when the target demand is constant. In other words, we find that when $d^\dagger_{i,t} = d^\dagger_i \quad \forall i, t$, there exist values $\bar{e}^*_i, \bar{u}^*_i$ and $\bar{v}^*_i$ such that when $\bar{e}_{i,t} = \bar{e}^*_i$, $\bar{u}_{i,t} = \bar{u}^*_i$ and $\bar{v}_{i,t} = \bar{v}^*_i$, then Eqs. (2.14 – 2.16) satisfy that $\bar{e}_{i,t+1} = \bar{e}_{i,t} = \bar{e}^*_i$, $\bar{u}_{i,t+1} = \bar{u}_{i,t} = \bar{u}^*_i$, and $\bar{v}_{i,t+1} = \bar{v}_{i,t} = \bar{v}^*_i$.

We split this derivation into two, first we derive a fixed-point equation for the realised demand. We then show that this fixed-point solution is consistent with a fixed-point solution for employment, vacancies, and unemployment. We also discuss the implications for long-term unemployment and solve the fixed-point equations for a simplified version of the model considering a complete network.

Realised demand fixed-point. Since the realised demand is the sum of employment and vacancies, when we sum Eqs. (2.14 – 2.16) we obtain the dynamic equation the realised demand follows,

$$\bar{d}_{i,t+1} = \bar{d}_{i,t} + (\delta_v - \delta_u)\bar{e}_{i,t} + \begin{cases} \gamma(1-\delta_u)(d^\dagger_{i,t} - \bar{d}_{i,t}) & \text{if } \bar{d}_{i,t} \geq d^\dagger_{i,t}, \\ \gamma(1-\delta_v)(d^\dagger_{i,t} - \bar{d}_{i,t}) & \text{if } \bar{d}_{i,t} < d^\dagger_{i,t}. \end{cases} \tag{2.52}$$

We simplify this expression by defining $\gamma'_u = \gamma(1-\delta_u)$ and $\gamma'_v = \gamma(1-\delta_v)$ as follows,

$$\bar{d}_{i,t+1} = \bar{d}_{i,t} + (\delta_v - \delta_u)\bar{e}_{i,t} + \gamma'_u(d^\dagger_{i,t} - \bar{d}_{i,t}) + (\gamma'_v - \gamma'_u)\max\left\{0, d^\dagger_{i,t} - \bar{d}_{i,t}\right\}. \tag{2.53}$$

We now use that the target demand is constant and apply the fixed-point condition $\bar{d}_{i,t+1} = \bar{d}_{i,t}$. This gives,

$$(\delta_u - \delta_v)\bar{e}_{i,t} = \gamma'_u(d_i^\dagger - d_i^*) + (\gamma'_u - \gamma'_v)\max\left\{0, d_i^\dagger - \bar{d}_i^*\right\}. \tag{2.54}$$

We now assume that the employment is at the fixed-point value $\bar{e}_{i,t} = \bar{e}_i^*$. Later on we will show this fixed-point value for employment is consistent with the equations we derive here. If the realised demand has a fixed-point solution d_i^*, this value must satisfy

$$(\delta_u - \delta_v)\bar{e}_i^* = \gamma'_u(d_i^\dagger - d_i^*) + (\gamma'_u - \gamma'_v)\max\left\{0, d_i^\dagger - \bar{d}_i^*\right\}. \tag{2.55}$$

We consider the two cases $d_i^\dagger < d_i^*$ and $d_i^\dagger > d_i^*$ and assume that $\delta_u > \delta_v$. If $d_i^\dagger < d_i^*$, Eq. (2.55) yields

$$d_i^\dagger - d_* = \frac{\delta_u - \delta_v}{\gamma'_u}\bar{e}_{i,t},$$

which is a contradiction since the right hand side is positive, but the left hand side negative (since $\delta_u > \delta_v$). In the case when $d_i^\dagger > d_i^*$, we find the following fixed-point solution

$$d_i^* = d_i^\dagger - \frac{\delta_u - \delta_v}{\gamma'_v}\bar{e}_i^*. \tag{2.56}$$

Performing an analogous analysis for the case $\delta_u < \delta_v$ we obtain the following fixed-point solution for the realised demand,

$$\bar{d}_i^* = \begin{cases} d_i^\dagger - \frac{\delta_u - \delta_v}{\gamma'_v}\bar{e}_i^* & \text{if } \delta_u \geq \delta_v, \\ d_i^\dagger - \frac{\delta_u - \delta_v}{\gamma'_u}\bar{e}_i^* & \text{if } \delta_u < \delta_v. \end{cases} \tag{2.57}$$

In other words, when $\delta_u > \delta_v$ the realised demand is lower than the target demand. This happens because when $\delta_u > \delta_v$ (i.e. the probability of separation is higher than the probability of opening a vacancy at random) the adjustment towards the target demand does not fully compensate for asymmetry between the opening and separation rates; thus the fixed-point realised demand is lower than the target demand. Similarly, when $\delta_u < \delta_v$, the fixed-point realised demand is higher than the target demand. In both cases, the difference between the fixed-point realised and the target demand is proportional to $|\delta_u - \delta_v|$ and inversely proportional to the adjustment rate γ. Except for the case when $\delta_u = \delta_v$ the fixed-point solution we derived for the realised demand depends on the employment having a fixed-point. We discuss this more in the next section.

Fixed-point for employment, unemployment and vacancies. We use Eq. (2.57) to derive a fixed-point equation for vacancies. We focus on the case $\delta_u \geq \delta_v$ since the other case can be solved analogously. It follows from Eq. (2.57) and the fact that the realised demand is the sum of employment and vacancies that any fixed-point solution for the vacancies must satisfy that

$$\bar{v}_i^* = d_i^\dagger - \left(1 - \frac{\delta_u - \delta_v}{\gamma_v'}\right) \bar{e}_i^*. \tag{2.58}$$

When $\delta_u \geq \delta_v$, $d_i^\dagger < d_i^*$, so the maximum term is zero in the employment and unemployment equations (Eqs. (2.14) and (2.15)). It follows that, under the fixed-point condition $e_{i,t+1} = e_{i,t}$, Eq. (2.14) leads to

$$\bar{e}_i^* = \frac{1}{\delta_u} \sum_j \bar{f}_{ji}^*, \tag{2.59}$$

where $\bar{f}_{ji}^*$ is the flow of workers at the fixed-point and is given by

$$\bar{f}_{ij}^* = \frac{\bar{u}_i^* (\bar{v}_j^*)^2 A_{ij} (1 - e^{-\bar{s}_j^*/\bar{v}_j^*})}{\bar{s}_j^* \sum_k \bar{v}_k^* A_{ik}}, \tag{2.60}$$

with

$$\bar{s}_j^* = \sum_i \frac{\bar{u}_i^* \bar{v}_j^* A_{ij}}{\sum_k \bar{v}_k^* A_{ik}}. \tag{2.61}$$

Using Eq. (2.59) and imposing the fixed-point condition $u_{i,t+1} = u_{i,t}$ we find that

$$\frac{1}{\delta_v} \sum_j \bar{f}_{ji}^* = \frac{1}{\delta_u} \sum_j \bar{f}_{ij}^*. \tag{2.62}$$

In other words, at the fixed-point the total inflow of workers into an occupation is proportional to the total outflow of workers of that same occupation. When $\delta_u = \delta_v$ these two flows are equal.

Equations (2.59) and (2.62) are non-linear functions of $\mathbf{e}^*$, $\mathbf{u}^*$, and $\mathbf{v}^*$. We have not found a general closed-from solution for these equations. However, there are three important observations we can make. Equation (2.58) shows that a fixed-point solution depends on the target demand, which is an external input to the model. Equation (2.59) shows that the fixed-point employment depends on the flow of workers between occupations, which depends on the network structure (see Eq. (2.60)). Therefore the network structure plays a role in determining fixed-points and steady-states. Equation (2.62) shows that a fixed-point solution will also depend on the parameters δ_u and δ_v. Finally, even though we do not have an analytical solution for

Eqs. (2.58), (2.59), and (2.62), we have found a fixed-point solution using computer calculations for all the networks and parameters we have studied.

Analytically determining the stability of a fixed-point in our system is out of the scope of this thesis. The system has thousands of variables and is not differentiable. This means that we cannot use the standard Jacobian analysis. However, for all the parameter values and scenarios we study, we find a stable behaviour.

Long-term unemployment. Equation (2.51) gives the number of long-term unemployed workers for time t. When the system is at a fixed-point with unemployment of occupation i given by $\bar{u}_i^*$ then, the number of unemployed workers with a job spell of k time steps is

$$\bar{u}_i^{*(k)} = \delta_u \bar{e}_i^* \left(1 - \frac{\sum_j \bar{f}_{ij}(\bar{\mathbf{u}}^*, \bar{\mathbf{v}}^*; A)}{\bar{u}_i^*}\right)^k, \tag{2.63}$$

which decays with k.

Analytical solution for a complete network. We now show that our model has an analytically computable fixed-point solution under the following assumptions:

1. A complete network of n nodes, i.e. $A_{ij} = \frac{1}{n} \quad \forall i, j$.
2. $\delta_u = \delta_v = \delta$ and that $\gamma_u = \gamma_v = \gamma$.
3. The target labour demand is constant, equal to the labour supply, and distributed homogenously among all occupations i.e. $\bar{d}_i^\dagger = \frac{L}{n} \quad \forall i$.

As before, we denote the fixed-point value of the variables with a star superindex (e.g. x^*).

In the previous section we showed that fixed-points depend on the target demand and the network structure. Since in this scenario all occupations have equal target demand and are positioned indistinguishably in the network, all occupations have the same fixed-point. Therefore, we lose the i subindex in our notation. Since $\delta_u = \delta_v$, Eq. (2.62) yields $\sum_j \bar{f}_{ji}^* = \sum_j \bar{f}_{ij}^* \equiv F^*$. Using the full expression for the flow of workers in Eq. (2.60) we obtain

$$F = \sum_{j=1}^{n} \frac{1}{n} \frac{u^* v^{*2}(1 - e^{s^*/v^*})}{s^* \sum_{k=1}^{n} \frac{1}{n} v^*} = \frac{u^* v^* (1 - e^{-s^*/v^*})}{s^*}. \tag{2.64}$$

Similarly, from Eq. (2.30) we have that

$$s^* = \sum_{i=1}^{n} \frac{1}{n} \frac{u^* v^*}{\sum_{k=1}^{n} \frac{1}{n} v^*} = u^*. \tag{2.65}$$

We then substitute s^* in F and obtain,

$$F = v^*(1 - e^{-\bar{u}^*/\bar{v}^*}), \tag{2.66}$$

It follows from assumption 2 and Eq. (2.57) that at the fixed-point $\bar{d}^* = d^\dagger$. With this in mind, using Eq. (2.66) and Eqs. (2.14 – 2.16), we obtain the following dynamic equations for the total number of employed and unemployed workers and job vacancies,

$$\begin{aligned} \bar{e}^* &= \bar{e}^* - \delta\bar{e}^* + v^*(1 - e^{-u^*/v^*}), \\ u^* &= u^* + \delta\bar{e}^* - v^*(1 - e^{-u^*/v^*}), \\ v^* &= v^* + \delta\bar{e}^* - v^*(1 - e^{-u^*/v^*}). \end{aligned} \tag{2.67}$$

We know that the number of unemployed and employed workers equals the labour, i.e. $U^* + E^* = L$. Since all occupations have the same fixed-point, then $\bar{u}^* + \bar{e}^* = \frac{L}{n}$. It follows from this observation, the fact that $d^* = \frac{L}{n}$, and Eq. (2.57) that

$$\bar{u}^* = \frac{L}{n} - \bar{e}^* = \bar{v}^*. \tag{2.68}$$

We then substitute Eq. (2.68) into Eq. (2.67) and obtain

$$\bar{e}^* = \frac{L}{n}\frac{(1 - e^{-1})}{\delta + (1 - e^{-1})}, \tag{2.69}$$

and

$$u^* = v^* = \frac{L}{n}\frac{\delta}{\delta + (1 - e^{-1})}. \tag{2.70}$$

These equations along with the condition of constant labour force imply that the unemployment rate is

$$\frac{U^*}{L} = \frac{\delta}{\delta + 1 - e^{-1}}. \tag{2.71}$$

In other words, in this case the unemployment rate is an increasing function of δ.

Zero fixed-point. Equations (2.14 – 2.16) accept a trivial fixed-point $e_i^* = u_i^* = v_i^* = d_i^\dagger = 0$. We neglect this fixed-point since it is uninteresting for our analysis. However, when running the agent simulation there is a non-zero probability that the number of employed workers and vacancies of an occupation is zero, i.e. $\hat{e}_{i,t} = \hat{v}_{i,t} = 0$. At this point, even if $d_{i,t}^\dagger > 0$, no vacancies would open and therefore employment would be zero for the rest of the simulation. To avoid this, we introduce the additional rule that if $\hat{e}_{i,t} = \hat{v}_{i,t} = 0$ but $d_{i,t}^\dagger > 0$, then a vacancy opens. When running the simulation with a large number of agents (which is the case of the labour market) the probability that $\hat{e}_{i,t} = \hat{v}_{i,t} = 0$ is negligible and this additional rule is very unlikely to be used.

2.5.2 Parameter effects

In this subsection, we briefly discuss how the parameters δ_u and δ_v affect the fixed-point unemployment rate. For this analysis, we consider sets of parameters $\delta_u, \delta_v, \gamma$ that satisfy that $\frac{\delta_u - \delta_v}{\gamma} < 0.06$. In this case, the fixed-point realised demand is approximately equal to the target demand (see Eq. (2.57)). This means that the reallocation process plays a minor role in determining fixed-points, and we can focus on the spontaneous separations and opening of vacancies to understand how fixed-points change.

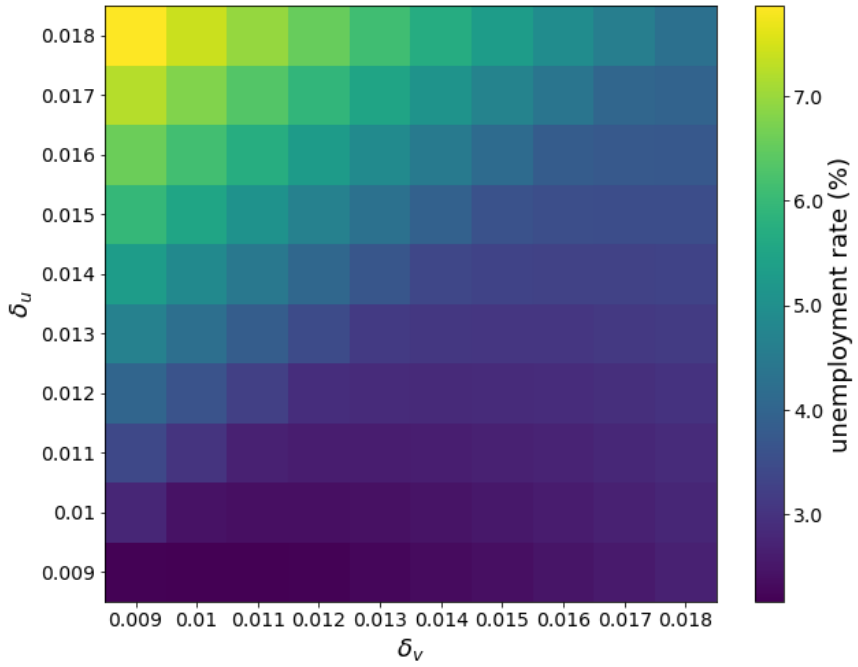


Figure 2.5: **Fixed-point unemployment.** A heatmap showing the fixed-point unemployment rate for different parameters values of δ_u and δ_v. In general, unemployment rate increases with δ_u and decreases with δ_v.

The parameters governing the spontaneous separation of workers and opening of vacancies are δ_u and δ_v, where δ_u is the probability of workers being separated and δ_v the probability of a vacancy opening. Intuitively, one would expect the unemployment rate to increase with δ_u and decrease with δ_v. As shown in Fig. 2.5 this is almost always the case. There are mild exceptions when an increase in δ_v can increase unemployment. For example, when $\delta_u = 0.009$ and $\delta_v = 0.018$ the unemployment rate is 2.58, but when $\delta_u = 0.009$ and $\delta_v = 0.017$ the unemployment rate increases to 2.68. This happens when the difference between δ_u and δ_v is not very small, and when many vacancies open, the reallocation process separates additional workers so

that the realised demand equalizes with the target demand. However, there are few of these exceptions, and the differences in the unemployment rate are small and are even difficult to spot from Fig. 2.5. Therefore, we conclude that, in general, the unemployment rate increases with δ_u and decreases with δ_v.

2.5.3 Steady-state and fixed-points

In the subsections above we studied the fixed-point behaviour of the dynamical systems defined by Eqs.(2.14 – 2.16). In this subsection, we argue that the fixed-point we find corresponds to the expected value of a steady-state of the agent-based model. We keep the target demand constant and study the behaviour of the unemployment rate with simulations. Fig. 2.6 shows the agent-based model's unemployment rate through time at the aggregate and occupation level. This figure shows that, for several parameter values, when the target demand is constant, the steady-state unemployment fluctuates around a fixed value, showing convergence. Furthermore, the expected value of different simulations closely matches the fixed-point found using Eqs.(2.14 – 2.16). The size of the fluctuations is larger for occupations with lower demand (e.g. motion picture projectionists) and smaller for occupations with higher employment (e.g. electricians). This suggests that as one increases the number of agents in the model, the approximation works better.

From the remainder of this thesis, we refer to the steady-state and fixed-point interchangeably. Although we do not have mathematical proof that the agent-based model converges to a steady-state corresponding to a fixed-point from the dynamical system, the simulation evidence supports this hypothesis. In the following chapters, we also explore different scenarios and find that the dynamical system defined by Eqs. (2.14 – 2.16) accurately describes the behaviour of the agent-based model.

2.6 Concluding remarks

In this chapter, we presented a data-driven network model of the labour market. This model has the advantage that it can explore how the labour market reacts to external economic variations, for example, recessions or labour reallocation. This chapter presented a mathematical analysis of the model, where we derived dynamic equations that approximate the expected value of the number of employed, unemployed, and long-term unemployed workers, as well as the number of vacancies, all of them at the occupation level. We showed that these approximations work well in the limit of a large number of agents. These dynamic equations have the advantage that they can

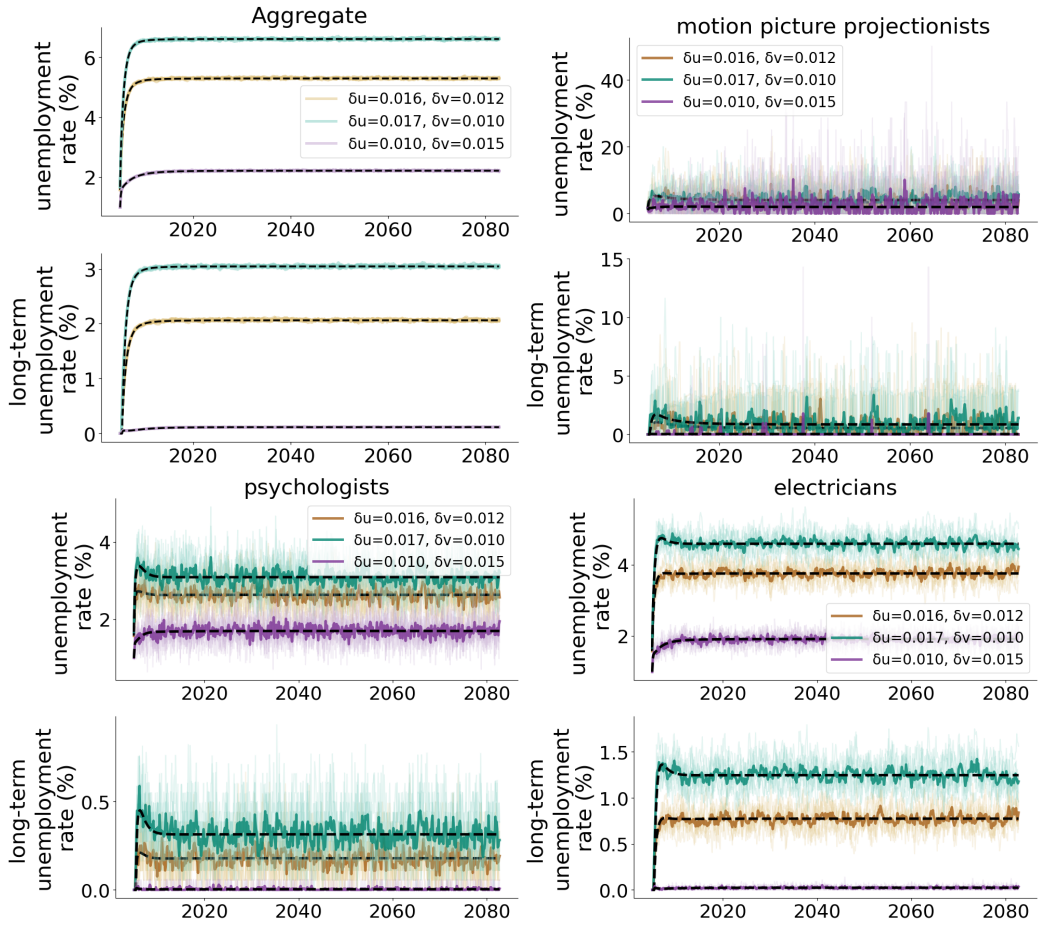


Figure 2.6: **Simulations and approximations at the aggregate and occupation level.** In the *top left* panel, the aggregate unemployment rate, in the *top right* brokerage clerks, in the *bottom left* motion picture projectionists and in the *bottom right* electricians. The panels show the unemployment and long-term unemployment rates. The transparent lines denote simulations, the solid lines the mean across the ten simulations and dashed black lines the numerical approximation. Each colour denotes different parameter values.

be solved numerically relatively fast, unlike the agent-based model simulations that are computationally costly to run. Having these equations allows us to explore the parameter space more quickly, which helps us understand the behaviour and calibrate the model.

The dynamic equations also allowed us to derive fixed-point equations when the target demand is constant. For particular parameter choices, we solved these equations numerically and found a fixed-point with stable behaviour. Further work could involve a more detailed mathematical analysis of the stability of this system. Re-

cent work (La Guardia & Miranda 2018) has shown that Lyapunov stability can be extended to non-differentiable maps as long as they are Lipschitz continuous (which is the case of our dynamical system). One could use the techniques developed in La Guardia & Miranda (2018) to the stability of the system analytically.

In the next chapters, we will work with the dynamic equations we derived to describe the behaviour of the agent-based model to understand the economics of the labour market.

Chapter 3

The Beveridge curve and model calibration

3.1 Introduction

The unemployment and vacancy rates are among the most important economic variables for understanding the state of the labour market. Their relevance has been such that their relationship has its own name: the Beveridge curve. The Nobel Prize in Economics for 2010 was awarded to Diamond, Mortensen, and Pissarides for their work on search theory (Mortensen 1982, Pissarides 1979, Diamond 1982), which among other things, explains some of the behaviour of the Beveridge curve. This chapter discusses how our model can help understand some of the unexplained behaviour behind this curve, and how we calibrate our model to fit the Beveridge curve.

Our model has the advantage that it outputs economic variables both at the macro and micro levels. At the macro level, the model provides unemployment, long-term unemployment, and vacancy rates. At the micro-level, the model outputs occupation-specific rates. Since the model outputs economic variables at the occupation level, it would be ideal to use fine-grained occupational data to calibrate all model parameters. However, these data are not always available; in particular, there is no publicly available data of vacancies at the occupation level. Therefore, we use the unemployment and vacancy rate (which form the Beveridge curve) to calibrate our model.

This chapter is structured as follows. The next section explains the behaviour of the Beveridge curve and discusses the related literature. Section 3.3 presents our calibration methodology, where we show that we can calibrate our model to replicate the empirical behaviour of the curve for the U.S labour market. Section 3.4 discusses

the empirical behaviour of the Beveridge curve and how our model reproduces it. Section 3.5 discusses how the dynamics of the curve change with parameter values and their implications for explaining the cycling direction of the curve. Finally, in Section 3.6, we conclude with further work and important remarks.

3.2 The characteristics and dynamics of the Beveridge curve

The Beveridge curve (i.e. the graphical representation of the relationship between unemployment and the job vacancy rate) has shown empirical regularities through time that portray the state of the economy. In this thesis, we focus on discussing the characteristics of the U.S. Beveridge curve, but we note that most of these characteristics are present in the Beveridge curve of several other countries (Arpaia et al. 2014).

In Fig. 3.1, we plot the Beveridge curve for the U.S. between January 2001 and September 2018. We observe that in recession periods, and particularly from December 2007 to June 2009, the unemployment and vacancy rates move downward along the curve. In contrast, from July 2009 onward, the economy recovers and the unemployment and vacancy rates move upward along the curve. While the slope is negative for almost all periods, the curve itself also shifts as the cycle unfolds. After the 2009 financial crisis the Beveridge curve shifted away from the origin, with unemployment increasing for all vacancy rates. Throughout time, the Beveridge curve exhibits the following three important features.

1. **Shifts of the Beveridge curve**. The curve can shift away or toward the origin (Diamond 1982). For example, after the 2009 financial crisis, the Beveridge curve shifted away from the origin, i.e. for the same vacancy rate the unemployment rate was higher after the crisis.

2. **Movements along the curve during recession and recovery periods** During recession periods, unemployment and vacancy rates move downward along the curve, and during recovery periods, the unemployment and vacancy rates move upward along the curve. The recession from December 2007 to June 2009 and the recovery period from 2009 onward are a good example of this feature.

3. **Counter-clockwise cycling direction** Historically, the Beveridge curve has (almost) always shifted outwards after recessions (Diamond & Şahin 2015), i.e. the curve cycles counter-clockwise. In other words, for the same vacancy rate, the unemployment rate has been larger during recoveries than during recessions. This feature is particularly worrying to policymakers and is also known as the *jobless recovery* (Eeckhout & Lindenlaub 2019).

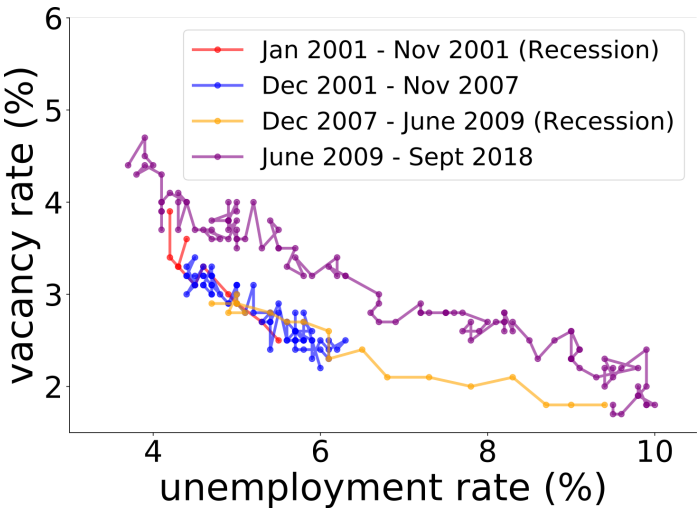


Figure 3.1: **The Beveridge curve.** The historical Beveridge curve (unemployment and vacancy rates) for the United States, 2000-2018. Different periods are highlighted with different colours. Red and yellow points correspond to recessions, while blue and purple dots to recovery periods.

The first two features have been explained by several models (see (Elsby et al. 2015) for a review). Most prominently, the Diamond-Mortensen-Pissarides model predicts that structural changes in the labour market cause the Beveridge curve to shift. For example, in their model, a decrease in worker-vacancy matching efficiency would shift the Beveridge curve further away from the origin.

The first feature was also explained by Axtell et al. (Axtell et al. 2019), who built a model where workers search for jobs through the labour flow network (a network of firms). For unweighted synthetic networks their model shows that the network structure could shift the Beveridge curve closer or further away to the origin. The following Section (3.4) shows that our model also reproduces this feature.

The Diamond-Mortensen-Pissarides framework also explains that during booms of the business cycle, more vacancies open and unemployment decreases, while in recessions, fewer vacancies open and more workers are separated. Thus, they correctly

predict the second feature: points on the upper left of the curve correspond to booms and points on the lower right to recessions (Pissarides 2011, Diamond 1982).

The third feature, the counter-clockwise cyclicality of the Beveridge curve, is yet to be fully understood. The standard interpretation is that the outwards shifts of the Beveridge curve correspond to a deterioration in the matching/hiring process in the economy (Diamond & Şahin 2015). However, some models (Mortensen 1999, Pissarides 1985, Sniekers 2018) argue that this phenomenon is independent of structural change and due to the business cycle dynamics.

For example, in Pissarides 1985 model vacancies are not a state variable and during a recovery period vacancies increase immediately, while unemployment takes time to decrease since the worker-vacancy matching requires time. This mechanism implies that vacancies increase faster than unemployment decreases during recoveries and causes the counter clockwise cyclicality of the Beveridge curve.

Eeckhout & Lindenlaub (2019) argue that the counter-clockwise cyclicality of the Beveridge curve originates from employed workers increasing their search intensity in recovery periods. The employed workers fill in several job vacancies, decrease the rate at which unemployed workers find jobs, and slow down the decline of the unemployment rate.

In Section 3.4, we show that our model also supports the hypothesis that the counter-clockwise cyclicality of the Beveridge curve results naturally from the business cycle dynamics. Our model shows this feature while incorporating the vacancy rate as a state variable and without the need of modelling time-dependent search efforts. Before discussing this, we first discuss our calibration methodology.

3.3 Calibration

The target labour demand, the network, and the parameters δ_u, δ_v and τ determine the dynamics of our model as described by Eqs. (2.14 – 2.16). To calibrate the target labour demand and the occupational mobility network we can rely almost entirely on occupation-specific data. For the other parameters, we rely on the Beveridge curve, i.e. we use macroeconomic variables.

Initialising target labour demand. To calibrate the target labour demand in the absence of external shocks, we assume that the labour market is initially in a steady-state where the distribution of target labour demand across occupation equals the distribution of employment in that occupation in 2016. Ideally, one would make

the target labour demand of each occupation proportional to the sum of the empirical number of employed workers and job vacancies in that occupation. However, there is no publicly available vacancy data at the occupation level. Since employment constitutes the largest part of the demand (we observe that employment usually corresponds to more than 95% of the demand), we consider it appropriate to make the target labour demand of each occupation proportional to the occupational employment. We then scale the occupation-specific target labour demand so that the aggregate target labour demand equals the labour force (i.e. $D_0 = L$).

Fitting δ_u and δ_v (spontaneous shocks), and Δt (time step duration). We simulate an idealised business cycle and adjust these three parameters to find the best match to the empirical U.S. Beveridge curve from December 2007 to September 2018. To create the artificial business cycle, we assume the aggregate target demand D_t follows a *sine* wave of the form $D_t = D_0(1 + a\sin(t/2\pi T_{\text{cycle}}))$, where D_0 is the initial demand, and T_{cycle} is the period of the business cycle. Based on visual inspection, we assume that the empirical curve has traversed about three-quarters of a business cycle between 2008 and 2018. Thus, January 2008 is about a quarter of a cycle past the previous peak, and December 2018 is the new peak. This gives a period of the oscillation $T_{\text{cycle}} = 10 \times \frac{4}{3} \approx 13.3$ years. (The assumptions about phase do not influence the fit, they only explain our reasoning in choosing T_{cycle}).

To match the time step duration Δt in our model with real-time units, we use the definition of long-term unemployment. The U.S. Bureau of Labor Statistics defines long-term unemployed workers as those who have an unemployment spell of 27 or more weeks[1]. Since τ is the number of time steps after which a worker is considered long-term unemployed, Δt satisfies that $\Delta t = \frac{27\text{weeks}}{\tau}$.

We assume the model is at its steady-state at the beginning of the simulation, with the initial target demand $d_0^\dagger$ of each occupation proportional to the employment in 2016 as explained previously. We then let the target demand $d_{i,t}^\dagger$ of individual occupations move in tandem according to the *sine* wave, so that each occupation makes a pro-rata change tracking D_t, i.e. $d_t^\dagger = d_0^\dagger(1 + a\sin(t/2\pi T_{cycle}))$ and simulate the model.

We run an exhaustive search over possible values of the amplitude a of the *sine* function (which determines the business cycle's amplitude) and the parameters δ_u, δ_v, and τ. Once we determine τ we can calculate the time step duration using $\Delta t = \frac{27\text{weeks}}{\tau}$ and match the business cycle period T_{cycle}. The objective of the search is to minimize

[1] https://data.bls.gov/timeseries/LNS13008636

the discrepancy between the model's Beveridge curse and the empirical one. As a criterion for the goodness of fit, we compare the intersection of the enclosed areas. The objective function is

$$\min_{a,\delta_u,\delta_v,\tau} \frac{A_m \cap A_e}{A_m \cup A_e}, \tag{3.1}$$

where A_m is the area enclosed by the Beveridge curve of the model, A_e is the area enclosed by the empirical Beveridge curve, $A_m \cap A_e$ is the intersection of their areas, and $A_m \cup A_e$ is the union of their areas. This fitting method requires each Beveridge curve to enclose an area. The model's Beveridge curve is closed since the *sine* wave is periodic. We close the empirical Beveridge curve by connecting the starting and endpoints. We use this criterion since it considers the shift in the Beveridge curve during the business cycle which, as we discussed in the previous section, is relevant for the economic literature.

The optimal parameters are $a = 0.065$, $\tau = 4$ (i.e. $\Delta t = 6.75$ weeks), $\delta_u = 0.016$ and $\delta_v = 0.012$. Figure 3.2 shows the fit between the model's Beveridge curve and the empirical one. The model's optimal parameters are reasonably stable concerning the optimal choice $a = 0.065$. For example, when we increase a by 10%, Δt remains constant while δ_u and δ_v increase roughly by 6%, and when we decrease a by 10%, Δt and δ_u remain constant while δ_v increases by less than 5%. Section 3.5 discusses how the Beveridge curve changes with different parameter choices.

Calibrating r, the probability that a worker changing jobs remains within the same occupation. We are handicapped by the fact that this is not directly recorded. Still, we can use data on the annual occupational mobility rate at the aggregate level, which is the percentage of workers that change occupations within a year, to infer this indirectly. A recent study shows that in the Danish economy, the annual occupational mobility rate is 19% (Groes et al. 2014). Therefore, we assume that each year 81% of workers remain in their current occupation and estimate r as follows.

In Chapter 2, we explained that we use the empirical occupational transitions to incorporate the relative preference with which a worker from occupation i applies to a job vacancy in j (for $i \neq j$). Consistent with this approach (and acknowledging the same caveats), here we use the fact that every year 81% of workers remain in their current occupation to calibrate the preference r with which workers chose to apply to job vacancies in their current occupation.

For simplicity, we consider the following abstraction. We assume that the probability that a worker does not change occupation in one time step is time-invariant and

constant across occupations. Then, we observe that in the model, only workers who are unemployed change occupation. Thus, the probability that a randomly chosen worker *does not* change occupation in one time step is the probability that she is employed $(1-u)$ plus the probability that she is unemployed (u) times the probability that she does not change occupation (r), that is $(1-u)+ur$. Then, the probability x that a worker does not change occupation in y time steps is $x = ((1-u)+ur)^y$. Solving for r implies

$$r = \frac{x^{1/y} + u - 1}{u}. \tag{3.2}$$

With the baseline parameters, our model makes roughly $y = 52/6.75 = 7.7$ time steps in one year. Assuming $x = 0.81$, $\delta_u = 0.016$ and $u = 0.06$ (the average U.S. unemployment rate since the year 2000) gives the estimate $r = 0.55$. In Appendix I.1.1, we discuss how we can use some census data to incorporate some heterogeneity into the self-loops. However, this heterogeneity does not change significantly the dynamics of our model. Therefore, out of simplicity, we consider homogeneous self-loops for our baseline.

Calibrating γ, which is the rate at which the realised demand adjusts towards the target demand. Unfortunately, we have no empirical data to calibrate this parameter, but as we demonstrate in Appendix I.1.2, the results of the model are fairly insensitive to γ across a wide range of reasonable parameters. We choose $\gamma = 10\delta_u$.

3.4 Reproducing the behaviour of the Beveridge curve

This section discusses in detail how our model reproduces the three features of the Beveridge curve discussed in Section 3.2.

Shifts of the Beveridge curve. The Beveridge curve produced by our model can shift toward and outward the origin due to structural changes. For example, when the worker-vacancy matching efficiency decreases, the Beveridge curve shifts further away from the origin. To demonstrate this, we hold the target aggregate demand D_t constant, and vary the structure of the network by replacing the empirical occupational mobility network A by a complete network. We do this for different values of δ_u and fixed value of δ_v and trace the fixed-point behaviour in Fig. 3.3. In this figure δ_u varies from 0.009 to 0.0165 in intervals of 0.00025, while we fix

Parameter	Value	Description
δ_u	0.0160	Rate at which employed workers are separated due to the spontaneous process.
δ_v	0.0120	Rate at which employed vacancies are opened due to the spontaneous process.
γ	0.160	Speed at which the realised demand adjust towards the target demand by separating workers or opening vacancies.
Δt	6.75	Duration of a time step in units of weeks.
r	0.55	Probability that a worker stays in the same occupation
τ	4	Time steps after which a worker is considered long-term unemployed

Table 3.1: Parameter values after calibration (rounded)

$\delta_v = 0.012$. The highlighted markers with bold border correspond to the calibrated parameters ($\delta_u = 0.016$ and $\delta_v = 0.012$).

As expected, when we change the network structure from the empirical occupational mobility network to the complete network, the Beveridge curve shifts downwards towards the origin. When we consider parameters calibrated to actual data (highlighted with a bold border), removing the network structure corresponds to an increase in unemployment from 4.1% to 5.3%. This effect is substantial, representing a 29% increase.

Similar to what we observed last chapter, Fig. 2.5 shows that, generally speaking, when δ_u increases, the unemployment rate also increases and the the vacancy rate decreases. However, when the difference between δ_u and δ_v is large the vacancy rate increases with δ_u. This is why the left-most section of the Beveridge curve in Fig. 2.5 is upward sloping for both networks.

Movements along the curve during recession and recovery periods. Our model reproduces the dynamics of the Beveridge curve over business cycles. Figure 3.2 shows that the unemployment rate increases in the recession while the vacancy rate decreases. In the recovery, the vacancy rate increases and the unemployment rate decreases. This feature is independent of the cycling direction of the Beveridge curve.

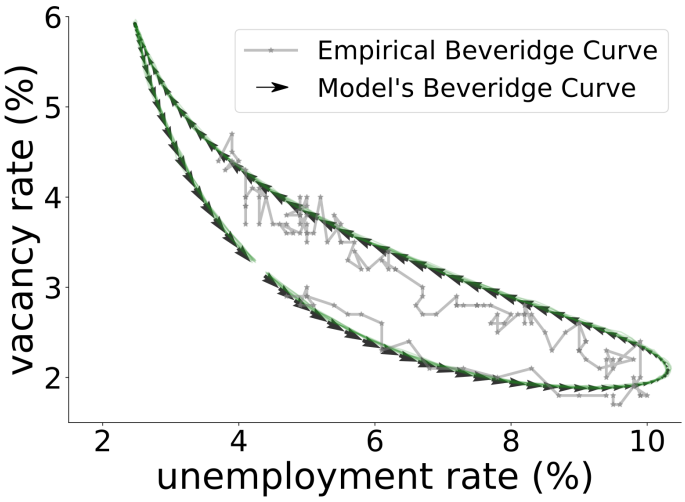


Figure 3.2: **The Beveridge curve.** The Beveridge curve generated by our model. The parameters of the model are calibrated to match the empirical Beveridge curve between December 2007 and December 2018. The dashed lines correspond to the deterministic approximation of Eqs. (2.14 – 2.16) and solid green lines to the full stochastic model simulation of Eqs. (2.3 – 2.5). The transparent grey line shows the empirical Beveridge curve between December 2007 and December 2018.

In Fig. 3.4, we show some examples where the curve cycles clockwise. However, even in this case, the unemployment and vacancy rates move downward along the curve during the recession and upward along the curve during the recovery.

Counter-clockwise cycling direction. The structural changes are not the only mechanism with which our model shows shifts in the Beveridge curve. An externally imposed business cycle, without assuming any structural change or labour reallocation, can cause the curve to cycle counter-clockwise, shifting further away from the origin after recessions. Fig. 3.2 shows that under the calibrated parameters the Beveridge curve cycles in a counter-clockwise direction even though we assume no structural changes. However, unlike the first two features we discussed, the direction of the cycle of the Beveridge curve strongly depends on the parameters δ_u and δ_v. In the next section, we study this in further detail.

3.5 Parameter effects on the Beveridge curve

The parameter values, as well as the type of network used in the model, influence the Beveridge curve's position, direction of cycle, and area it encloses. In this section, we discuss how the direction in which the Beveridge curve cycles is influenced by the

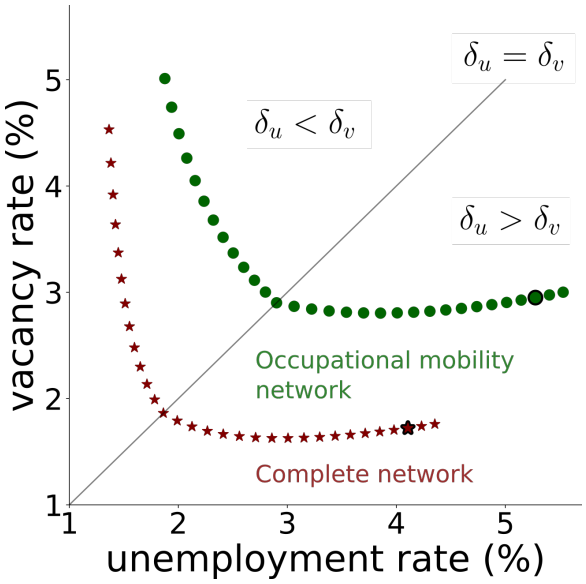


Figure 3.3: **The Beveridge curve.** The unemployment and vacancy rate for different parameter values and networks. In particular, we plot the difference between a complete network, with no skill mismatch frictions, vs. the empirical occupational mobility network. Each dot corresponds to the fixed-point unemployment and vacancy rate for different values of δ_u and δ_v. The highlighted points correspond to the unemployment and vacancy rate of the model using the calibrated parameters (see Section 3.3 for details on calibration).

state-independent rates at which workers are separated (δ_u) and vacancies open (δ_v). To do this, we run a number of numerical exercises varying δ_u and δ_v while keeping all other parameters fixed. In these exercises, the aggregate target demand is constant at the beginning and then oscillates according to a *sine* wave (see top left panels of Fig. 3.4 and Fig. 3.5)

We test five different parameter options for δ_u and δ_v. Starting from the calibrated values $\delta_u = 0.016$ and $\delta_v = 0.012$, we gradually decrease δ_u and increase δ_v until $\delta_u = 0.012$ and $\delta_v = 0.016$. During these gradual changes, we find that the Beveridge curve first reduces its enclosed area, then changes its cycling direction from counter-clockwise to clockwise, and finally increases its enclosed area. This behaviour is present for both the occupational mobility network and the complete network's Beveridge curve (see Figs. 3.4 and 3.5).

For the two networks and for the five cases we study, we observe that when $\delta_u > \delta_v$, the curve cycles counter-clockwise. On the contrary, when $\delta_v > \delta_u$ the curve cycles clockwise. However, when $\delta_u = \delta_v$, the network determines the the direction of the cycle – the occupational mobility network shows clockwise cycles while the complete network shows an "8"-shaped curve, where the bottom part cycles counter-clockwise

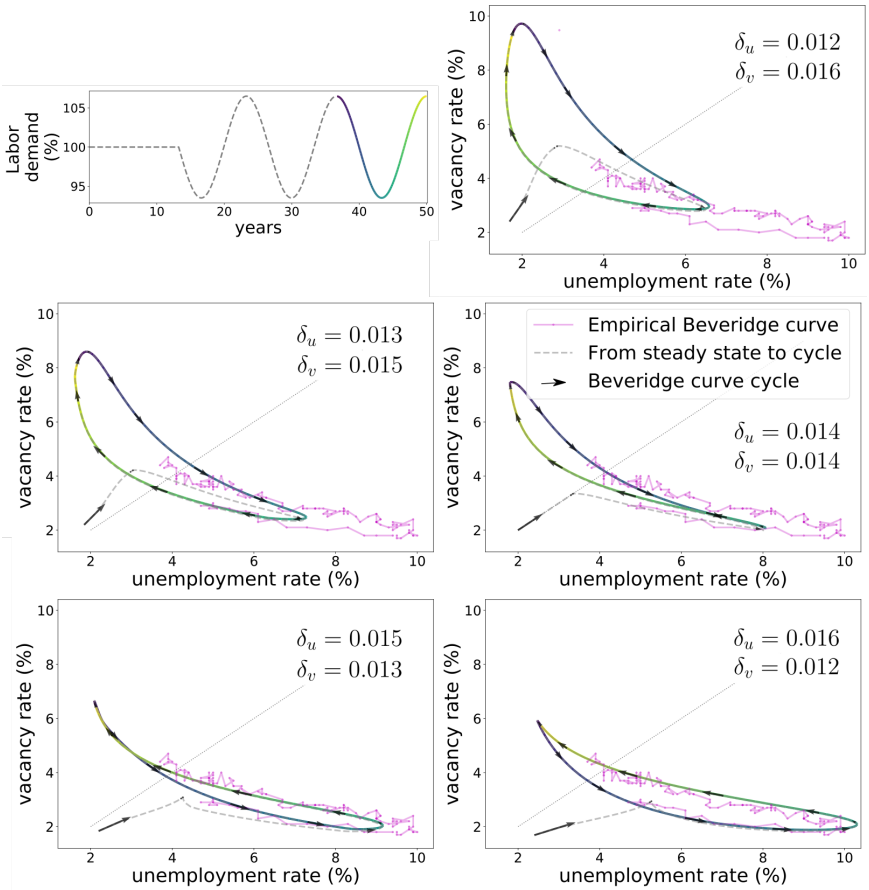


Figure 3.4: **Beveridge curve dynamics for the occupational mobility network.** The *top left* panel shows the aggregate demand. The grey part corresponds to the first part of the trajectory. The purple/blue part corresponds to the recession period, and the green/yellow part to the recovery period. The following 5 panels show the dynamics of the model's Beveridge curve under different parameter choices. We observe that when $\delta_u > \delta_v$ the curve cycles counter-clockwise, while when $\delta_u \leq \delta_v$ the curve cycles clockwise. We also show the empirical Beveridge curve in magenta for reference.

and the upper part clockwise. These results suggest that for some similar values of δ_u and δ_v, which depend on the network structure, the curve flips and starts to exhibit the opposite cycling direction. We also observe that the network structure affects the area enclosed by and the position of the curve (see differences between Fig. 3.4 and Fig. 3.5).

These results suggest that as δ_u takes on greater values than δ_v, the Beveridge curve tends to cycle in a counter-clockwise direction;[2] i.e. the recovery of unemploy-

[2]The difference between δ_u and δ_v needed for the Beveridge curve to cycle counter-clockwise depends on the network.

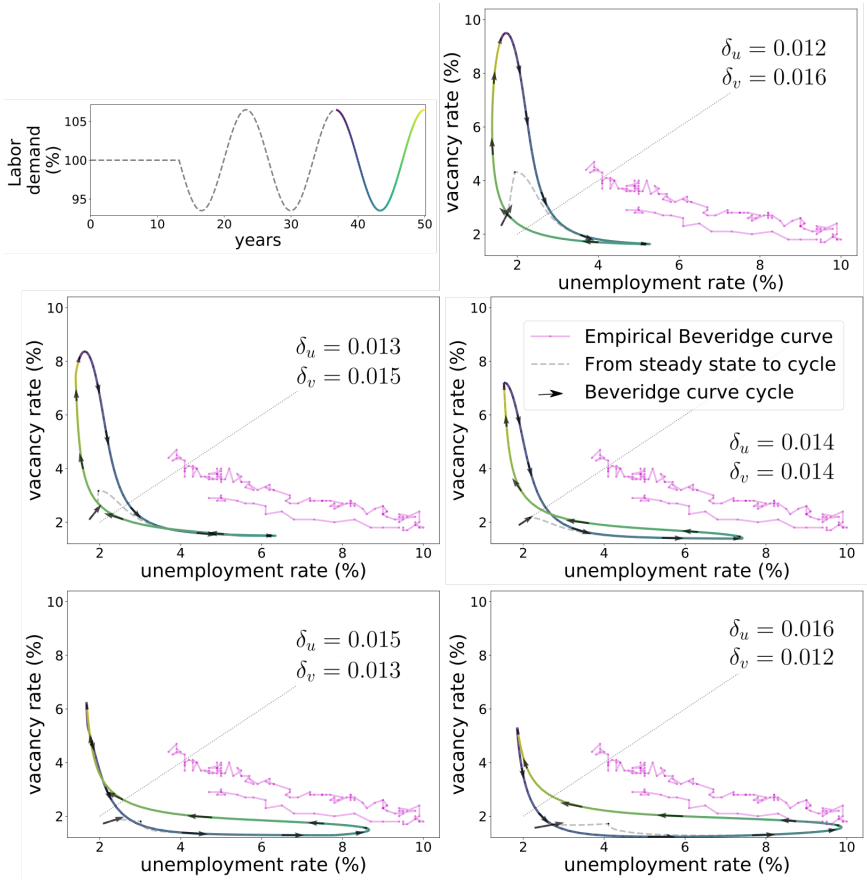


Figure 3.5: **Beveridge curve dynamics for the complete network.** The *top left* panel shows the aggregate demand. The grey part corresponds to the first part of the trajectory. The purple/blue part corresponds to the recession period, while the green/yellow part to the recovery period. The following 5 panels show the dynamics of the model's Beveridge curve under different parameter choices. We observe that when $\delta_u \geq \delta_v$ the curve cycles counter-clockwise, while when $\delta_u < \delta_v$ the curve cycles clockwise. We also show the empirical Beveridge curve in magenta for reference.

ment is slow, meaning that for the same vacancy rate, the unemployment rate will be higher during recoveries than during recessions. In other words, our model suggests that the jobless recovery may occur when the rate at which vacancies open spontaneously (i.e. independently from labour reallocation) is smaller that the spontaneous rate at which workers are separated.

3.6 Concluding remarks

This chapter showed that our model can reproduce important macroeconomic features of the labour market at the steady-state and during business cycles. We found

a good fit between the Beveridge curve of our model and the empirical one. Using an empirical network and calibrating our model to reproduce the Beveridge curve, quantified the labour market frictions imposed by the empirical network: These frictions can account for more than a quarter of the steady-state unemployment rate.

We must note the caveat that we are using four parameters (δ_u, δ_v, τ, and a) to fit the Beveridge curve. This means that our parameter calibration may suffer from overfitting and that we may need to improve our calibration methodology to produce reliable forecasts. However, as we show in Chapter 4, our model can still provide valuable insights into the labour market dynamics. In Appendix I, we show that, even when we use different parameters than the ones calibrated in this section, our main results remain robust.

The findings of this chapter are relevant for the macroeconomic literature on the Beveridge curve. Studies based on search theory and networks have argued that the shifts of the Beveridge curve are caused by structural changes (Petrongolo & Pissarides 2001, Rogerson et al. 2005, Axtell et al. 2019). Meanwhile, other studies suggest that these shifts are part of the Beveridge curve's counter-clockwise cyclicality, which results from business cycles dynamics (Kohlbrecher & Merkl 2016, Sniekers 2018).

This chapter found that structural changes, such as changes in the network structure, cause shifts in the Beveridge curve. However, these structural changes are not necessary for the Beveridge curve to present an outward shift. Instead, business cycles alone are enough to cause the Beveridge curve to cycle counter-clockwise. This result is important, given that a better understanding of the mechanisms underpinning the Beveridge curve could help policymakers mitigate adverse employment impacts of business cycles and accelerate the recovery process.

There is plenty of room for further work. For example, we could use job ads to infer the job vacancies at the occupation level (Turrell et al. 2018) and study occupation-specific Beveridge curves. Alternatively, one could use occupation-specific data to calibrate the model. Another interesting analysis would be to explore how different parameters change the direction of the cycle of the Beveridge curve in more detail. Finally, the Beveridge curve's location and cycle amplitude has changed through time. It would be interesting to see if we can reproduce these changes by introducing time dependent parameters in our model.

Chapter 4

The future of employment

4.1 Introduction

In response to widespread concern about the potential impact of automation on the labour market (Brynjolfsson et al. 2017, Autor et al. 2003, 2006, Autor 2015*b*), significant effort has been devoted towards analysing how susceptible a given occupation is to computerization (Frey & Osborne 2017, Brynjolfsson et al. 2018, Arntz et al. 2016). However, studies that estimate the likelihood of a robot 'stealing' a particular job only provide part of the picture.

Consider, for example, the job security faced by a statistical technician vs a childcare worker. Estimates developed by Frey & Osborne (2017) suggest that statistical technicians are more likely than childcare workers, for example, to be replaced by software technology. However, should such forecasts eventuate, and statistical technicians find themselves out of a job, their existing skills could allow them to transition into a range of 'safer' occupations with lower automation risk and growing demand.

In contrast, while childcare workers may not experience a *direct* threat from computerization, their employment prospects may nonetheless still be impacted. As automation displaces people in other occupations, many of these workers could have the requisite skills to become childcare workers and may consequently provide an *indirect* effect on the job security of existing childcare workers. Thus, even though the immediate risk of automation is predicted to be larger for statistical technicians, accounting for possible occupational transitions and labour demand reallocation could see childcare workers facing a greater risk of unemployment.

In this chapter, we study these important, but overlooked, indirect labour displacement effects using the labour market model we developed in Chapter 2 and calibrated in Chapter 3.

To explore the potential impacts of automation on the labour market, we impose an automation shock to our model that, over the years, decreases demand for labour in some occupations and increases demand in others. We focus on a transition period of around a decade right after the automation shock hits, and the bulk of labour reallocation occurs. We then study the associated aggregate and occupation-specific unemployment and long-term unemployment ($>$ 27 weeks) dynamics. We study two automation shock scenarios, based on estimates by Frey & Osborne (2017) and Brynjolfsson et al. (2018).

Unsurprisingly, we find that occupations at higher risk of automation tend to be affected most, but we also show that restrictions on worker movements imposed by the occupational mobility network generate substantial labour market mismatch. In some areas of the network, many workers can be competing for a small number of vacancies. Simultaneously, occupations in other areas can have vacancies that are left unfilled for a long time.

This chapter is structured as follows. Section 4.2 explains how we determine the automation shocks based on estimates in the literature (Frey & Osborne 2017, Brynjolfsson et al. 2018). Section 4.3 presents our analysis on the impact of automation on employment at the aggregate and occupation level. This section shows that the network structure plays an important role in determining the impact of automation on employment. Finally, in Section 4.4, we discuss future work and the policy implications of our work.

4.2 Automation and labour reallocation

Fears of technological unemployment are neither new, nor unfounded, but they are sometimes misplaced (Mokyr et al. 2015). While it is clear that automation will replace a number of workers, this is an old process that has so far not caused persistent large unemployment rates (Frank et al. 2019). Instead, the average length of the work week has declined substantially (Ramey & Francis 2009) and work has shifted to new occupations (Lin 2011).

Taking this into account, we model automation as a time-dependent reallocation shock that lowers the target demand for some occupations while raising the target demand for others. We assume that the decrease in the target demand of some occupations cancels out with the increase in the target demand of the other occupations, so that the aggregate labour demand remains constant i.e, $D_t^\dagger = D_0^\dagger = L$. As explained later, we will explore two automation scenarios based on two different estimates of

the automatability of occupations (Frey & Osborne 2017, Brynjolfsson et al. 2018). Below, we explain how we determine the automation shocks.

4.2.1 Labour reallocation

We introduce automation shocks into our model by specifying a post-shock reallocated target demand. After an automation period, the target labour demand converges to the post-shock target demand. We denote the time at which the target demand converges to the post-shock target demand by t^* and the post-shock target demand by $\mathbf{d}^\dagger_{t*}$. An automation scenario is given by a particular choice of the vector $\mathbf{d}^\dagger_{t*}$. We determine this vector as follows.

First, for each occupation i an automation scenario gives an automation level $p_i \in [0, 1]$. We explain how we determine the automation level of each scenario in Section 4.2.2. We assume that the automation level is the fraction of total hours worked in an occupation that are no longer needed post-shock. We also assume that working hours are reduced for all workers in the economy, so that the total demand for workers remains constant. This leads to the following calculation.

Let x_0 be the current number of labour hours for the average worker in a given period (say a week). The hours of work each occupation demands are given by the components of the vector

$$\mathbf{h}_0 = x_0 \mathbf{d}_0. \tag{4.1}$$

Letting $\mathbf{p}$ be the vector with the automation level of each occupation, the new number of hours of work $\mathbf{h}_{t^*}$ after automation is

$$\mathbf{h}_{t^*} = \mathbf{h}_0 \odot (\mathbf{1} - \mathbf{p}), \tag{4.2}$$

where $\odot$ denotes the element-wise multiplication of vectors and $\mathbf{1}$ the vector of ones. Assuming that the new aggregate hours of work demanded is split equally among workers, the new number of hours of work per week is

$$x_{t^*} = \frac{\sum_i^n h_{i,t^*}}{L}. \tag{4.3}$$

Finally, we split the hours of labour demanded by occupations equally among workers, so that

$$\mathbf{d}^\dagger_{t^*} \equiv \mathbf{d}^\dagger = \mathbf{h}_{t^*} \frac{1}{x_{t^*}}. \tag{4.4}$$

This final step guarantees that the aggregate labour demand remains constant. It is possible to assume that, rather than keeping the aggregate labour demand constant, automation either increases or decreases labour demand. We explore this further in the Appendix I.

4.2.2 Automation estimates

We study two automation scenarios, one based on a study by Frey & Osborne (2017) and the other based on a study of Brynjolfsson et al. (2018). We refer to these automation scenarios as the *Frey and Osborne shock* and as the *Brynjolfsson et al. shock*.

Frey & Osborne (2017) estimated the probability that each of 702 occupations in the O*NET 6-digit classification system could be computerized soon (Frey & Osborne 2017). To do this they gave experts a description of tasks performed by workers in a restricted sample of 70 occupations and asked them whether the occupations could be automated within the next two decades. Based on the experts' answers and using nine O*NET variables that describe occupations as inputs, they trained a supervised machine learning algorithm and estimated what they called the the *probability of computerization* for the remaining occupations. They found that approximately half of the jobs in the U.S. would be at risk for some degree of automation. Figure 4.1 shows the distribution of the probability of computerization accross occupations and the occupational mobility network. Since the probability of computerization is already bounded 0 and 1, we set the automation level equal to the probability of computerization in this scenario.

Brynjolfsson & Mitchell (2017) took a different approach, taking advantage of the 8-digit level O*NET classification of occupations based on work activities. (This has 974 occupations). They asked workers from a crowd sourcing platform to rate what they called the *suitability for machine learning* for each work activity. They then used the breakdown of work activities for each occupation to estimate the (mean) suitability for machine learning for each occupation, broken down into a five point scale (Brynjolfsson et al. 2018). Most occupations have at least some tasks that are suitable for machine learning, but few, if any, have all tasks suitable for machine learning. This suggests that many jobs will be re-designed rather than destroyed.

Since the suitability of machine learning for occupations lies between 0 and 5, we equal the automation level in this scenario to the suitability for machine leaning divided by 5. This guarantees the automation level lies between 0 and 1. Figure 4.2 shows the distribution of the probability of computerization across occupations and the occupational mobility network.

Both of these studies estimate the probability that an occupation will be *technically automatable*. This is not the probability that an occupation will be automated, which also depends on cost, institutions, etc., and it is not an estimate of the share of jobs in an occupation that will be automated. Nonetheless, for simplicity we use

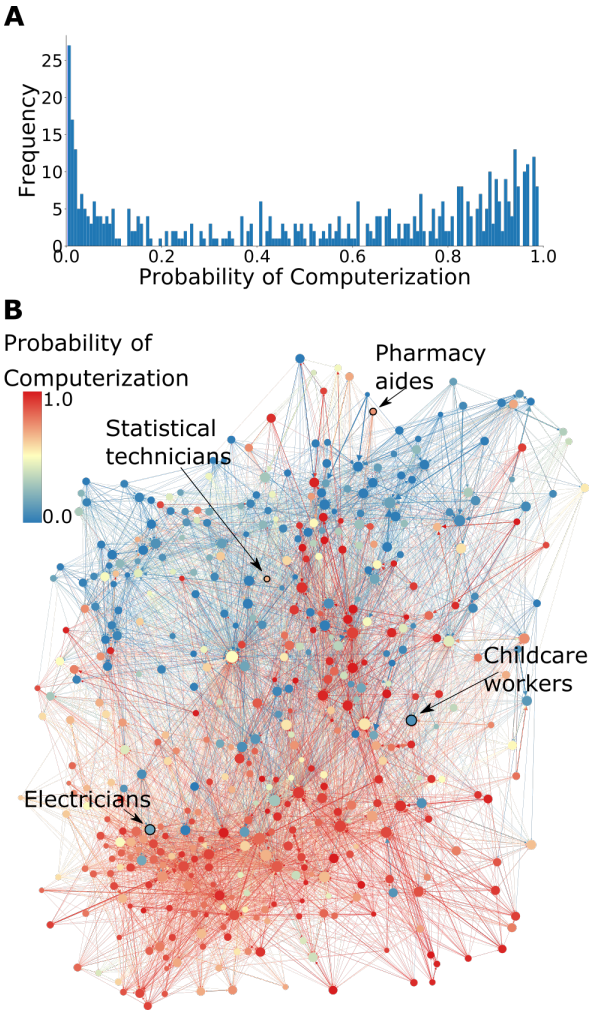


Figure 4.1: **Probability of computerization and the occupational mobility network.** Panel (**A**) is a histogram of the probability of computerization for different occupations as estimated by Frey & Osborne (2017), suggesting a bimodal distribution. Panel (**B**) shows the occupational mobility network, where nodes represent occupations and links represent possible worker transitions between occupations. Red nodes have high automatability and blue nodes have low automatability. The size of the nodes indicates the logarithm of the number of employees in each occupation.

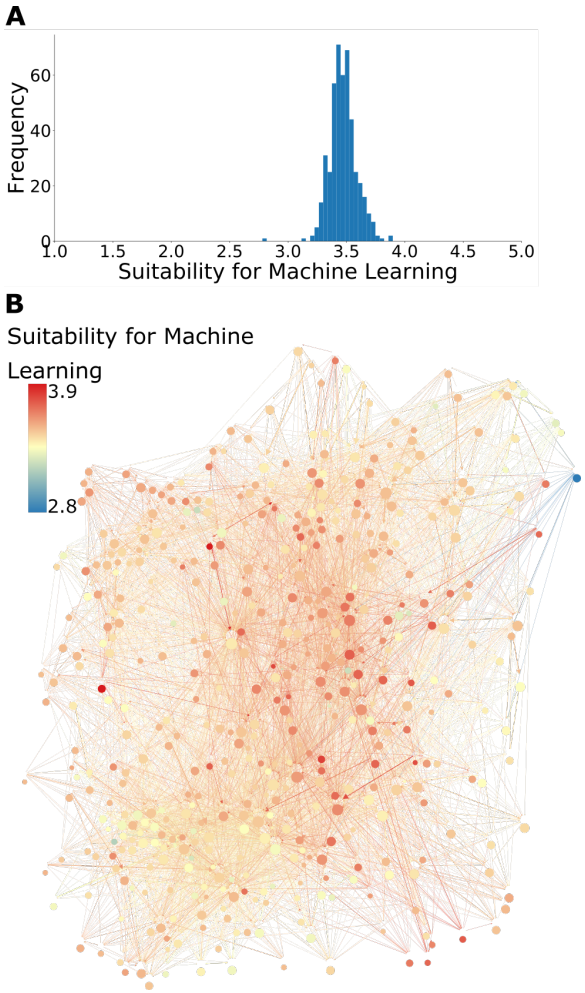


Figure 4.2: **Suitability for machine learning and the occupational mobility network.** Panel (**A**) is a histogram of the suitability for machine learning as estimated by Brynjolfsson et al. (2018). Panel (**B**) shows the occupational mobility network, where nodes represent occupations and links represent possible worker transitions between occupations. The colour of the nodes indicates the suitability for machine learning. Red nodes have higher suitability for machine learning and blue nodes have a low one. The size of the nodes indicates the logarithm of the number of employees in each occupation.

these estimated to calibrate the automation levels, directly determining the share of jobs in an occupation that will be automated. We map the 6 and 8 digit O-NET classifications used in these studies into the the U.S. occupational mobility network (which is based on the 4 digit American Community Survey classification) using the 2016 National Employment Matrix Crosswalk (see (Mealy et al. 2018)).

These two estimates are substantially different and yield different automation shocks. First, the estimates differ in their correlation to wages. The Frey and Osborne estimates are strongly anti-correlated with wages, whereas the Brynjolfsson et al. (2018) estimates have a low correlation with wages. Second, as we see in Fig. 4.1, the distribution of the probability of computerization is wide, whereas the suitability for machine learning distribution has a narrow peak. Since the Frey and Osborne estimates vary substantially between occupations and the Brynjolfsson et al. estimates do not, the corresponding changes in the target labour demand are large for the Frey and Osborne shock but small for the Brynjolfsson et al. shock. (See the top panels of Fig. 4.3 for examples of how the target labour demand changes for different occupations under the two shocks). In Section 4.3, we discuss how the differences between these two automation shocks affect the outcome of our model.

4.2.3 Formulating a time dependent automation shock

Previously, we discussed how automation shocks determine the post-shock target demand. However, the reallocation of labour demand is unlikely to happen immediately, and we must therefore specify a target labour demand that transitions over time from the pre-shock target demand to the post-shock target demand. We follow the innovation literature, which suggests that the adoption of technologies follows a sigmoid function or S-curve over time (Stoneman 2001).

Before introducing the automation shock, we first initialise the model so that the distribution of the target demand $d^{\dagger}_{i,0}$ matches the employment distribution of 2016 (see Chapter 3 for details). After the system reaches the steady-state, we introduce an automation shock by making the target demand $d^{\dagger}_{i,t}$ follow a sigmoid function, which begins at $d^{\dagger}_{i,0}$ and converges to the post-automation target demand (see top panels of Fig. 4.3 for examples). Frey and Osborne say that their estimates are over "some unspecified number of years, perhaps a decade or two." (Frey & Osborne 2017). Based on this, we choose the adoption rate so that the total shock is spread across a 30 year period, though most of the change happens within about 10 years.

We use the following function for the target demand

$$d_{i,t}^{\dagger} = \begin{cases} d_{i,0} & \text{if } t < t_s, \\ d_{i,0} + \frac{d_i^{\dagger} - d_{i,0}}{1+e^{k(t-t_0)}} & \text{if } t \geq t_s. \end{cases} \tag{4.5}$$

where t_s is the time at which the automation shock starts and t_0 is 15 years after t_s. Furthermore $k = 0.079$, which guarantees that the target demand equals the post-shock reallocate demand up to a 0.0001 tolerance within 30 years. In the Appendix I.1.5, we demonstrate the robustness of the results under variations in the time span of the automation shock.

4.3 The impact of automation on employment

We now use the model to assess the impact of automation shocks on employment for the Frey and Osborne shock and the Brynjolfsson et al. shock. We present our results at the aggregate and occupation level.

4.3.1 Aggregate level outcomes.

As seen in Fig. 4.3B and C, even though the aggregate target demand is held constant, the Frey and Osborne shock increases both the aggregate unemployment rate and the aggregate long-term unemployment rate during the period of automation. This increase is caused by the substantial reallocation of labour demand across occupations (see Fig. 4.3A for an example of how the target demand changes at the occupation level). In contrast, the Brynjolfsson et al. shock causes no noticeable change in the aggregate unemployment rates (see Fig. 4.4).

This difference is caused by the different distributions of the two shocks. The Frey and Osborne shock is very heterogeneous across occupations, affecting some occupations a great deal and others very little, so that the changes in target demand at the occupation level are substantial. In contrast, the Brynjolfsson et al. shock affects most occupations similarly, so that the changes in the target demand are smaller and the network effects at the aggregate level are small. Since the Brynjolfsson et al. shock has little effect at the aggregate level, we focus the discussion of this subsection on the Frey and Osborne shock.

We compare the behaviour with the empirical occupational mobility network to the hypothetical behaviour assuming a complete network, in which any worker can transition equally well to any occupation. We use the same parameters for both networks (see calibrated parameter values in Chapter 3). The aggregate unemployment

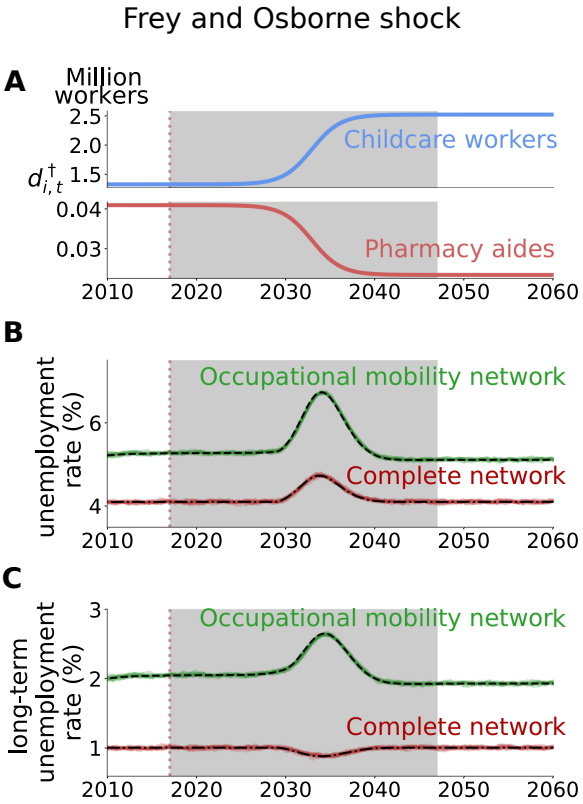


Figure 4.3: **Aggregate labour market outcomes under automation shocks.** The grey area denotes the 30 years during which the automation shock takes place. Panel (**A**) shows the evolution of the target labour demand for two example occupations. The occupations coloured in blue have a low automation level and the occupations coloured in red have a high level. Because it is more heterogeneous across occupations, the Frey and Osborne shock implies a larger change in the target labour demand of most occupations. Panel (**B**) shows the unemployment rate as a function of time. Dashed lines are our approximations of the expected value (solved numerically) and the solid lines are 10 simulations with 1.4 M agents. Panel (**C**) shows the long-term unemployment rate as a function of time. As before, dashed lines correspond to the deterministic approximation of Eqs. (2.14 – 2.16) and solid lines to the full stochastic model simulation of Eqs. (2.3 – 2.5).

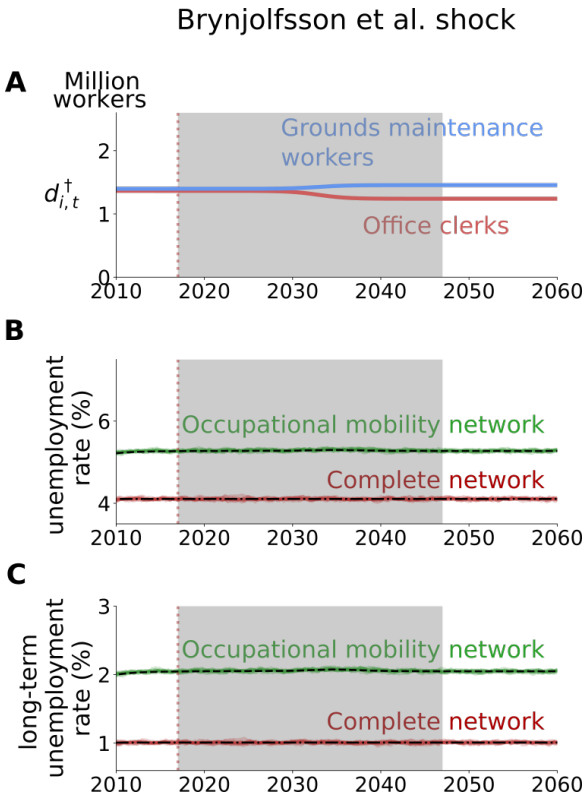


Figure 4.4: **Aggregate labour market outcomes under the Brynjolfsson et al. shock.** The grey area denotes the 30 years during which the automation shock takes place. Panel (**A**) shows the evolution of the target labour demand for two example occupations. The occupation coloured in blue has a low suitability for machine learning and the occupation coloured in red has a high one. Because the distribution of the suitability for machine learning is more evenly distributed across occupations that the probability of computerization, the Brynjolfsson et al. shock implies a small change in the target labour demand of most occupations. Panel (**B**) shows the unemployment rate as a function of time. Dashed lines are our approximations of the expected value (solved numerically) and the solid lines are 10 simulations with 1.5 M agents. Panel (**C**) shows the long-term unemployment rate as a function of time. As before, dashed lines correspond to the deterministic approximation of Eqs. (2.14 – 2.16) and solid lines to the full stochastic model simulation of Eqs. (2.3 – 2.5).

rate is initially about 5.3% for the empirical network and 4.1% for the complete network. When we apply the Frey and Osborne shock, the aggregate unemployment rate for the empirical network rises to 6.7% at its peak and then decays. In contrast, for the complete network, the aggregate unemployment rises to only 4.7% before it decays. Thus the total change in unemployment with the empirical network is more than a factor of two larger, demonstrating the importance of the network structure.

The long-term unemployment rate, for the occupational mobility network, is about 2.1% for the empirical network, substantially smaller than the (short-term) unemployment rate. When we apply the Frey and Osborne shock, the aggregate long-term unemployment rate rises to roughly 2.6% at its peak and then decays. The relative change from the initial value to the peak value is about 27% for both the unemployment and long-term unemployment. The behaviour for the complete network is quite different: First, the initial level of long-term unemployment for the complete network is only 1.0%, more than a factor of two smaller than for the empirical network. Second, when we apply the shock, long-term unemployment for the complete network remains nearly flat.

In the Appendix I.1.4, we also explore what happens when we relax the assumption that the aggregate target labour demand remains constant. Unsurprisingly, we find that when the aggregate demand decreases, the automation shock displaces more workers for all occupations.

A change in the steady-state unemployment. Another interesting result is that the steady-state value of the aggregate unemployment shifts after the shock. The aggregate unemployment rate changes from 5.28% to 5.10%, for a net change of roughly -0.17%. While this is small, bear in mind that we have kept both the total aggregate target demand and all the model parameters constant. This is consistent with our result that the steady-state explicitly depends on the network structure and the target demand in each occupation (see Chapter 2 Eqs. (2.59 – 2.62) for details).

The fact that we see this shift when we change the target demand demonstrates the key role that the network structure plays in determining the steady-state and transient behaviour. Note that there is no noticeable shift in the steady-state for the complete network.

We conjecture that the Frey and Osborne shock causes such persistent effects since automation levels of neighboring occupations tend to be similar. This has two effects: It means that there are some regions of the network where workers easily find new jobs, and others where workers get trapped because there are no good alternatives, causing

a substantial boost to long-term unemployment. The steady-state shift occurs because the post-automation distribution of the target labour demand across occupations is more concentrated on fewer occupations that are more densely connected between each other, reducing worker-vacancy matching frictions.

To test our conjecture, we create a surrogate *reshuffled* Frey and Osborne shock by randomising occupations' automation levels. We do this by randomly shuffling the automation level of each of the 464 occupations, i.e. randomly reassigning each automation level to a new occupation (without replacement). This preserves the distribution of automation levels but removes any correlation between neighboring occupations, i.e. this breaks the assortative distribution of automation levels. When we impose this randomised shock, the aggregate unemployment rate does not decrease, while the long-term unemployment tends to increase slightly (see Fig. 4.5). Thus the most persistent effects disappear when the correlation inherent in the network structure is removed.

In addition, we create another surrogate *assortative* shock where we randomise relative to the Frey and Osborne shock while intentionally creating a correlation between neighboring occupations as follows. As shown in Fig. 2.1 , occupations of the same broad classification typically have high connectivity. We take advantage of this and order occupations by their occupation code while ordering the probabilities of computerization in ascending order. We then create an assortative shock by matching the sorted lists of occupations and probabilities of computerization. When we impose this assortative shock, the post-automation aggregate unemployment rates decrease, supporting our conjecture (see Fig. 4.5).

To summarise, the results at the aggregate level demonstrate how the structure of the occupational mobility network can cause substantial and long-lasting dislocations of the labour force. When the target demand for labour is altered by the automation shock, old jobs are closed in some occupations and new jobs are opened in others. Under the Frey and Osborne shock, which has strong network structure, we see a large transient effect that causes a substantial rise in aggregate unemployment over the decade during which the dislocation takes place. This is also shown in aggregate long-term unemployment and even causes a permanent shift in aggregate unemployment.

4.3.2 Occupation level outcomes.

We now show how automation affects the occupation-specific unemployment rates, where the network plays a crucial role. We measure the *average unemployment rate*

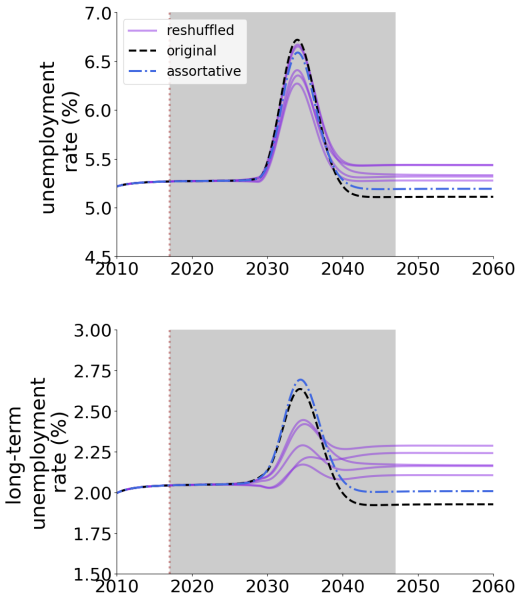


Figure 4.5: **Randomised and assortative versions of the Probability of Computerization shock.** In the *top* panel, the unemployment rate. In the *bottom* panel, the long-term unemployment rate. The dashed black line corresponds to the original shock, the solid magenta lines to randomised versions of the shock, and the blue line to the alternative assortative shock.

and *average long-term unemployment rate* during the shock as

$$\bar{u}_{i,\text{average}}(T) = \frac{\sum_{t\in T} \bar{u}_{i,t}}{\sum_{t\in T}(\bar{u}_{i,t} + \bar{e}_{i,t})}, \tag{4.6}$$

and

$$\bar{u}^{(\geq\tau)}_{i,\text{average}}(T) = \frac{\sum_{t\in T} \bar{u}^{(\geq\tau)}_{i,t}}{\sum_{t\in T}(\bar{u}_{i,t} + \bar{e}_{i,t})}, \tag{4.7}$$

where T is the set of time steps that correspond either to the period before automation or the automation shock duration. (We discuss an alternative way of defining the average unemployment rate in the Appendix I show that our results are robust). For simplicity, from here onward, we refer to the average unemployment rate and the average long-term unemployment rate during the automation period simply as the unemployment rate and the long-term unemployment rate.

Frey and Osborne shock. In Fig. 4.6, we compare the percentage changes in unemployment and long-term unemployment with each occupation's automation level. To highlight the role of the network, we do this both for the occupational mobility network and for the complete network. For the complete network, occupations with

the same automation level have the same percentage change in their unemployment rates. In contrast, for the occupational mobility network, the automation level is *not* a perfect predictor of the occupation-level outcome. Network effects specific to each occupation also affect unemployment (this is also true for Brynjolfsson et al. shock).

It is useful to highlight some specific cases to make the size of the network effects more clear. Both dispatchers and pharmacy aides have a high probability of computerization of 0.72. Still, the automation shock causes a 21% increase in the dispatchers' long-term unemployment, while the pharmacy aides' long-term unemployment *decreases* by roughly 17%.

Some occupations experience the opposite change that one would expect. Statistical technicians and pharmacy aides are likely to be automated (with a probability of computerization above 0.6) while childcare workers and electricians are not (with a probability of computerization below 0.2). However, statistical technicians and pharmacy aides *decrease* their long-term unemployment, while childcare workers and electricians increase theirs. This is due to the fact that it is relatively easy for statistical technicians and pharmacy aides to transfer to jobs in other occupations with increasing demand. In contrast, it is easier for other workers with occupations susceptible to automation to transfer to childcare workers or electricians, thereby increasing workers' supply relative to the demand. This illustrates the importance of network effects.

The unemployment rate of different occupations follows different trajectories (see Fig. 4.7). For example, office clerks, who are likely to be automated, show an increase in their unemployment rates during the shock. However, their unemployment rates return to a steady-state value roughly equal to their pre-shock rate. Figure. 4.7 shows that the fluctuations for the office clerks are considerably larger after the shock than before. This happens because the target demand for office clerks decreases substantially, and so does the number of workers in that occupation.

Dispatchers and pharmacy aides have similar automation levels. However, as Fig. 4.7 shows, dispatchers increase their unemployment rates considerably during the shock. Instead, pharmacy aides only slightly increase their unemployment rate, while the long-term unemployment decreases during and after the shock. Electricians, who are unlikely to be automated, first decrease their unemployment rate, but shortly after increase it. We explain this behaviour as follows. During the first part of the automation shock, more electrician vacancies open, thus decreasing unemployment. Nevertheless, the automation shock also causes workers of nearby occupations to become unemployed. As the automation shock continues to separate workers of

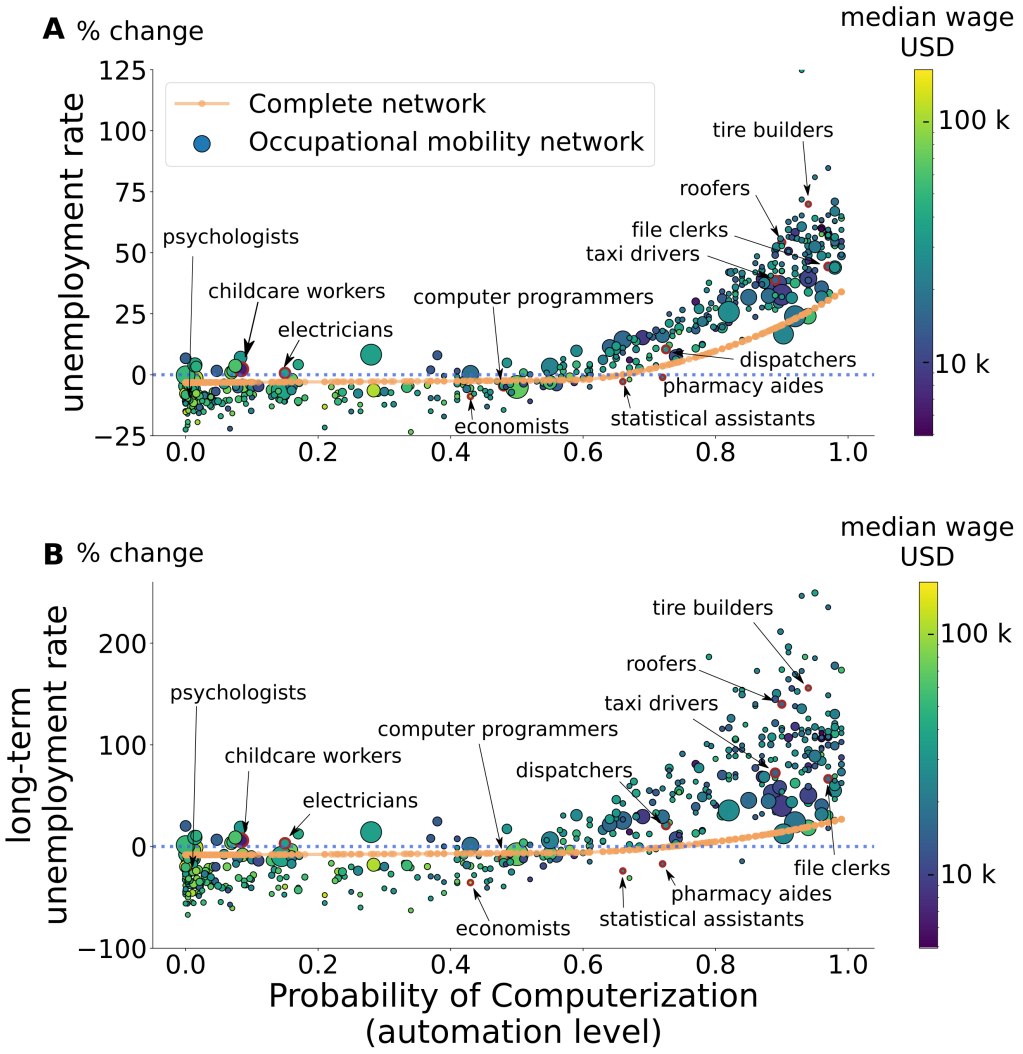


Figure 4.6: **Impact of the Frey and Osborne shock on unemployment and long-term unemployment at the occupation level.** The green dots are for the occupational mobility network and the orange dots are for the complete network. The size of the green dots is proportional to the employment of the occupation they represent. Panel **(A)** shows the percentage change in the unemployment rate vs the automation level for each occupation, while panel **(B)** shows the same thing for the long-term unemployment rate. The scatter in the results demonstrates that, due to network effects, the automation level only partially explains occupational unemployment.

neighboring occupations, many of these unemployed workers apply for the electrician vacancies causing the electricians' unemployment rates to increase.

Finally, we also study if the network effects are beneficial or detrimental, that

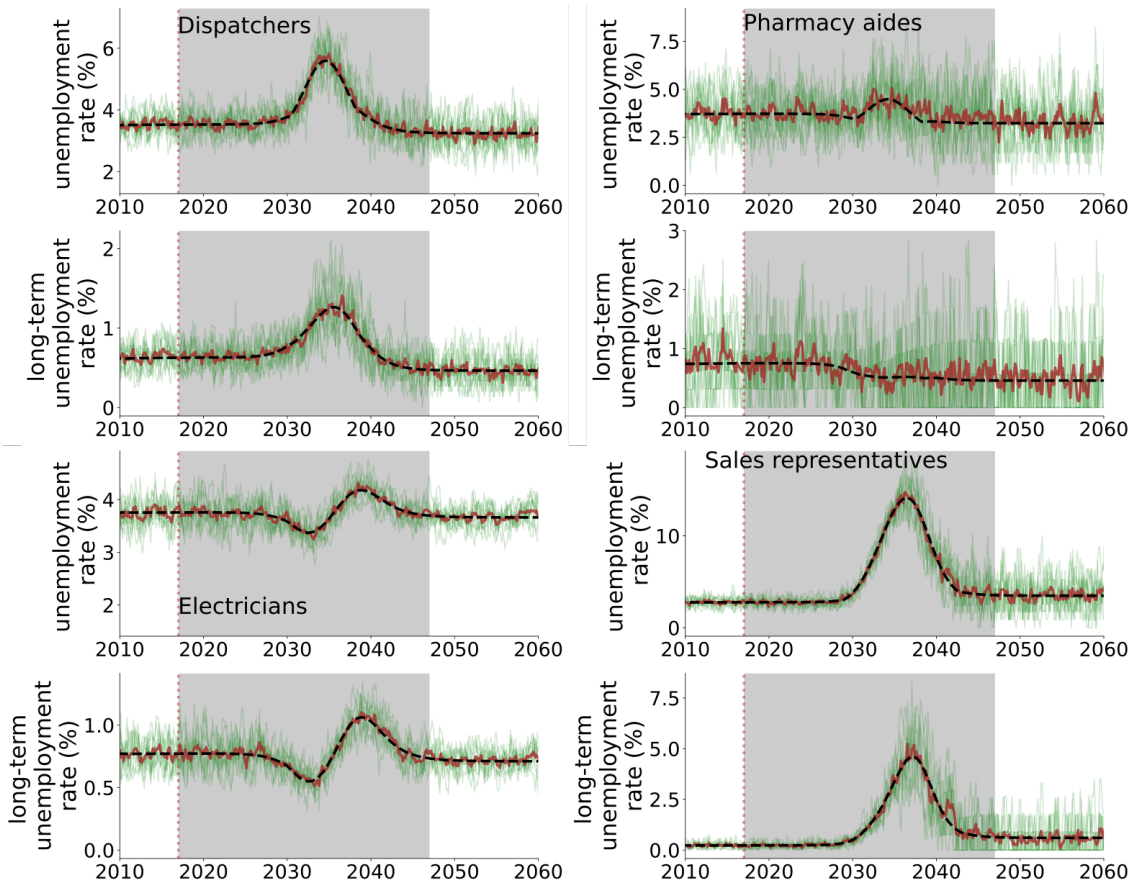


Figure 4.7: **Unemployment rate at the occupation level simulations and numerical solution.** We compare the unemployment and long-term unemployment from the average of the simulations (solid line) and the numerical solution (dashed line). We also show in transparent lines the 10 simulations. Each simulation uses 1.4 Million agents and we average over 10 simulations.

is, whether the occupation-specific unemployment rates decrease or increase when we consider the occupational mobility network instead of the complete network. We measure the network effects by taking the difference between the percentage change in unemployment rates in the two cases (i.e. the difference between the green and the black dots in Fig. 4.6). If the difference is positive, the network effects are detrimental – the occupation faces a larger increase in (long-term) unemployment.

In Fig. 4.8, we compare the network effects with the median wage of occupations. We find that occupations with low median wage are more likely to face detrimental network effects, while almost all high wage occupations have better outcomes with the empirical network than with the complete network.

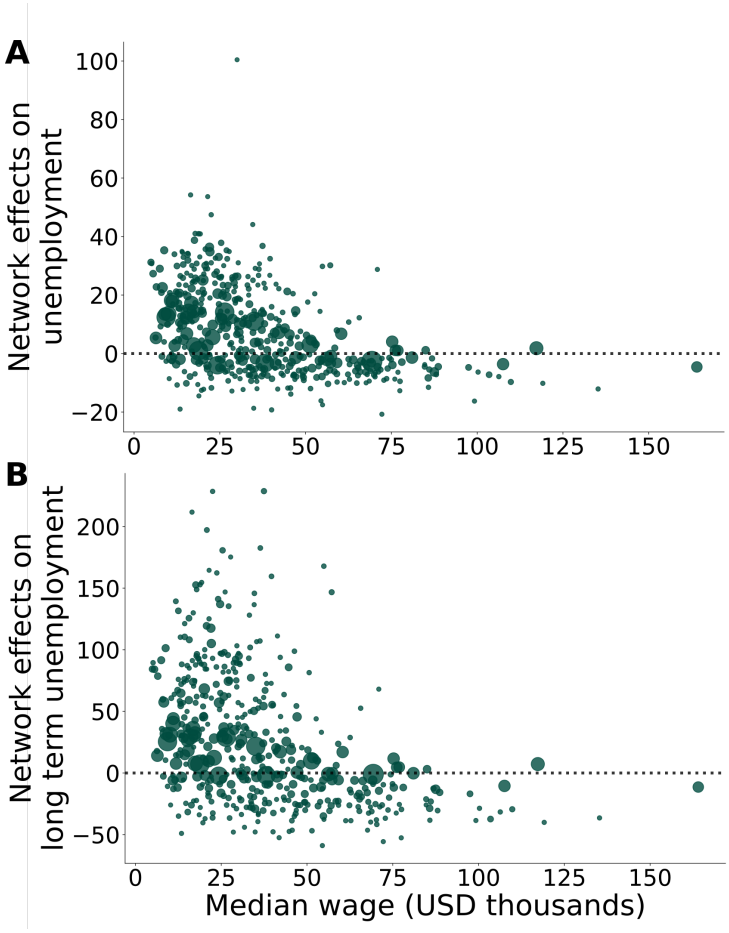


Figure 4.8: **Network effects at the occupation level.** We show the network effects for each occupation i.e. the difference between the green and the black dots in Fig. 4.6. If this difference is positive, network effects are detrimental for the occupation. Panel **(A)** Network effects on percentage change in unemployment rate vs. median wage. Panel **(B)** Network effects on percentage change in long-term unemployment rate vs. median wage.

Brynjolffson et al. shock. We now briefly discuss the results for the Brynjolffson et al. shock at the occupation level. Although there is no noticeable change in the aggregate unemployment rates, the Brynjolfsson et al. shock still affects occupations disproportionately. This effect depends not only on the suitability for machine learning but also on each occupation's network position.

As we observe in Fig. 4.9A and B, the change in the long-term unemployment and unemployment varies substantially for occupations with similar suitability for machine learning. For example, both machinists and avionic technicians have a high 0.70 suitability for machine learning score. Still, long-term unemployment for machinists slightly increases, while the long-term unemployment for avionic technicians *decreases*

by more than 20%.

Unlike the Frey and Osborne shock, in the Brynjolfsson et al. shock the network workers of different occupations face have no significant correlation with median wages (see Fig. 4.9C and D). This results is not surprising given that the suitability for machine learning is not correlated with the occupation's wage.

4.3.3 Network structure

In this section, we further explore how the network structure affects the impact of automation on employment in two ways. First, we run the model on randomised occupational mobility networks (rewiring edges or reshuffling weights) and analyse the results. Second, we model a retraining strategy by adding edges between occupations that share work activities and where one occupation substantially increases its long-term unemployment and the other decreases it. We do this analysis using the Frey and Osborne shock.

Randomising the network structure. We randomise the network structure of the occupational mobility network in two ways: the *weight reshuffling* and the *edge rewiring*. We keep the edges fixed in the weight reshuffling but randomise the number of transitions between occupations and then renormalize to have a column-stochastic adjacency matrix. In the edge rewiring, we randomise the ending point of an edge and preserve the edge's weight and starting point. We do not need to renormalize the adjacency matrix since by keeping the starting point of the edge and its weight fixed, the column-sum of the adjacency matrix remains constant.

These randomisations preserve different information about the network structure. The weight reshuffling preserves the topology, while the edge rewiring preserves the out-strength (i.e. weighted out-degree) but changes the topology. We run each randomisation ten times and analyse how our results change.

At the aggregate level, both randomisations have a lower spike in unemployment and long-term unemployment during the automation shock (see Fig. 4.10). This is expected given that, in the occupational mobility network, occupations with high automation probability tend to be clustered together. Randomising the structure destroys clustering and allows an easier transition into the new steady-state.

As shown in Fig. 4.10, the reshuffling of weights (blue lines) shows a much larger decrease in unemployment and long-term unemployment than the rewiring of edges (yellow lines). This happens because most of the information on the occupational mobility network is in the weights. The occupational mobility network is dense; the

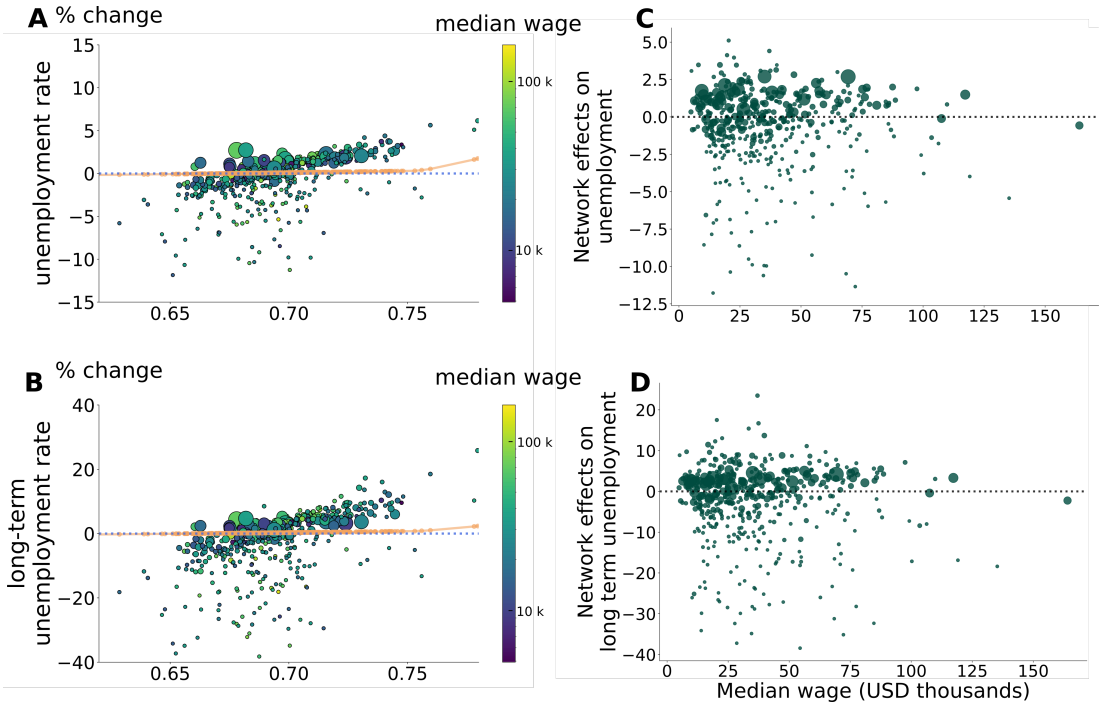


Figure 4.9: **Impact of the Brynjolfsson et al. shock on unemployment and long-term unemployment at the occupation level.** The green dots are for the occupational mobility network and the orange dots are for the complete network. The size of the green dots is proportional to the employment of the occupation they represent. Panel **(A)** shows the percentage change in the unemployment rate vs the automation level for each occupation, while panel **(B)** shows the same thing for the long-term unemployment rate. The scatter in the results demonstrates that, due to network effects, the automation level only partially explains occupational unemployment. The right panels show the network effects i.e. the difference between the green and the black dots in the left panels. If this difference is positive, network effects are detrimental for the occupation. Panel **(C)** Network effects on percentage change in unemployment rate vs. median wage. Panel **(D)** Network effects on percentage change in long-term unemployment rate vs. median wage.

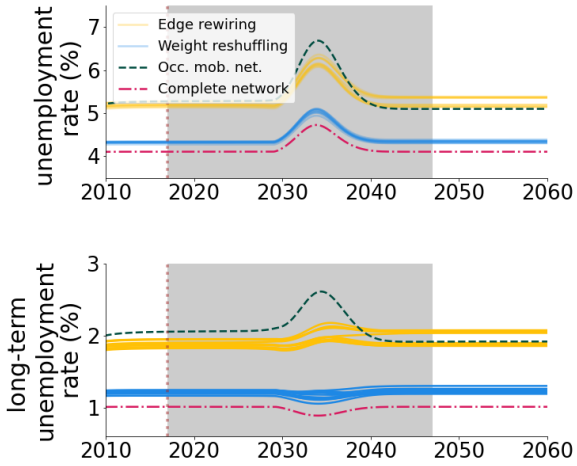


Figure 4.10: **Network structure effect at the aggregate level.** We show the change in unemployment and long-term unemployment due to the automation shock for different networks. The increase in unemployment rates is larger for the occupational mobility network. Both randomisations decrease the automation spike in unemployment rates, the weight reshuffling substantially more.

mean degree of a node is 155. In other words, on average, there has been at least one worker of each occupation transitioning into 155 out of 464 other occupations. The median degree of a node is 134, which is close to the mean. However, the number of transitions encoded in the weights vary substantially. When we randomise the weights, we lose the most significant part of the structure, while rewiring edges does not destroy too much information.

At the disaggregated level, occupations experience a considerably lower increase in unemployment and long-term unemployment percentage change in the randomised networks than in the occupational mobility network. In Fig. 4.11, we compare the percentage changes in the unemployment rates when we run the model using the occupational mobility network and one realisation of the reshuffled version. The black line is the identity, and the green line the linear fit. In all panels, the slope of the green line is noticeably lower than one. In other words, the change in unemployment rates for occupations is substantially less for the randomised versions of the occupational mobility network. The difference is more pronounced for long-term unemployment. This makes sense since the unemployment change is driven to a larger extent by the automation shock, while the network structure has a higher effect on long-term unemployment.

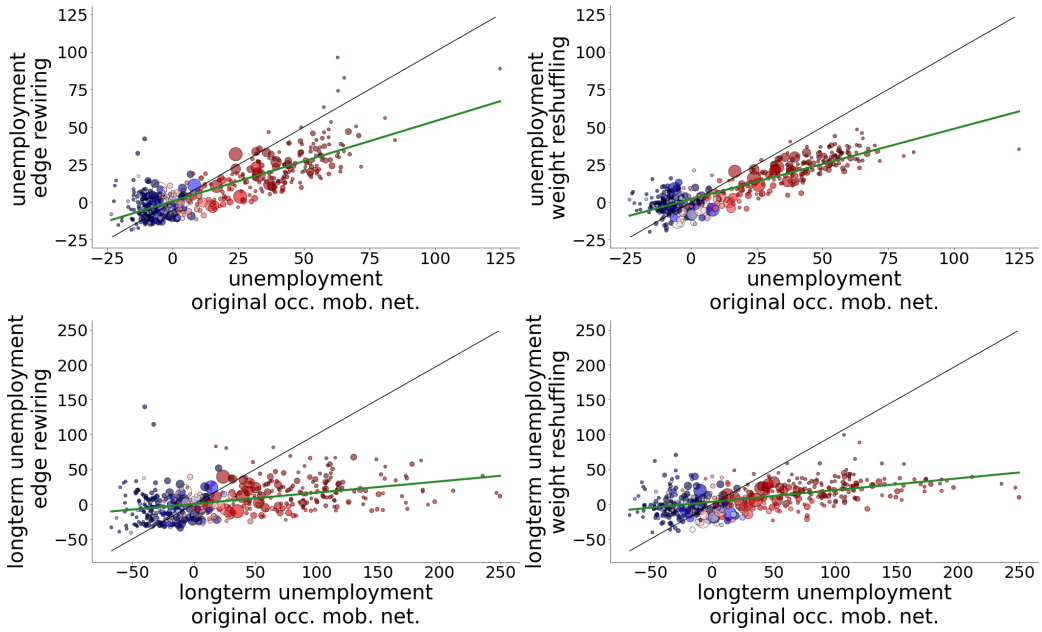


Figure 4.11: **Randomised network structure comparison with original network results**. In the *top* panels, we plot the percentage change in unemployment when running the model using the OMN and the reshuffled version of the OMN. In the *bottom* panels, we plot the percentage change in long-term unemployment comparison. In the *left* panels, we show the egde reshuffling, and in the *right* panels, the weight reshuffling. In both scenarios the changes in the unemployment rate are considerably smaller for the reshuffled versions of the OMN.

Retraining policies. We model retraining schemes by adding links to the occupational mobility network. Previous research (Mealy et al. 2018) has built the Job Space network, where occupations are linked according to the number of work activities they share (weighted by a scarcity factor). This research shows that sharing work activities is the best predictor of empirical occupation transitions. Acknowledging that retraining is most plausible between somewhat similar occupations, we restrict the possible edges that we can add as part of a retraining scheme using the Job Space network. In particular, we consider it viable to add an edge in the occupational mobility network only if the same edge is present in the Job Space with a weight above $\iota_{js} = 0.0013$. [1] We consider two retraining schemes, a targeted and a random one.

The targeted retraining focuses on the subset of occupations that experience a high increase in long-term unemployment (above 50%). We link these occupations to ten occupations that experience lower long-term unemployment. We choose these ten occupations according to the weight of the edge connecting them in the Job Space. For all occupations in the high unemployment subset, we are able to find ten low unemployment occupations that have a link with the high unemployment occupation with weight above ι_{js}. Once we selected the ten low unemployment occupations for each of the high unemployment occupations, we add edges with a weight equal to $\iota_{omn} = 0.006$, which is the mean weight of the edges in the occupational mobility network. Finally, we renormalise the occupational mobility network with the added retraining links so that the adjacency matrix is column-stochastic.

The second retraining scheme we consider is random. We add the same number of links as in the previous exercise, but we select occupation pairs randomly. We keep the restriction of only adding edges with a weight larger than ι_{js} in the Job Space and also add them with a weight of ι_{omn} in the occupational mobility network. As before, we also renormalise so that the matrix is column-stochastic.

Our results show that both the targeted retraining random retraining decrease the spike in unemployment and long-term unemployment caused by automation at the aggregate and occupational level (see Figs. 4.12 and 4.13). Similar to what happened for the occupational mobility network, the post-automation steady-state unemployment rate is lower than the pre-automation unemployment rate for the networks with retraining (see Fig. 4.12). The random retraining decreases the aggregate unemployment rate slightly more than the targeted retraining policy. This is likely because the occupations whose long-term unemployment increased the most tend to have low

[1] This value is roughly half the mean weight of the Job Space network.

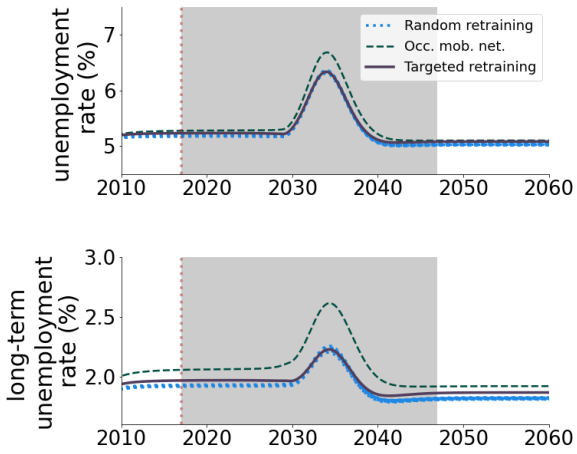


Figure 4.12: **Retraining policies.** The aggregate outcome of retraining policies, for either random retraining or targeted retraining. In blue, we plot 10 realisations of the random retraining strategy, all show very similar outcomes.

employment. Therefore targeted retraining may be targeted to occupations with few workers and thus involve fewer workers than the random retraining overall, which is shown in the aggregate level post-automation steady-state.

It might be surprising that the random retraining works so well at the disaggregated level. This is another result that highlights the importance of the network structure. Even if retraining is not directly targeted towards the most affected occupations, increasing mobility helps the affected occupations through second-order effects by freeing vacancies in adjacent occupations. It is important to bear in mind that we do not consider wage dynamics, which would play an important role for workers being harshly affected.

4.4 Concluding remarks

In this chapter, we applied the model developed in Chapter 2 to analyse the impact of automation on unemployment. We found that employment impacts for workers are likely to depend not only on the automatability of their current occupation but also on the alternative occupations they can transition into. We also found that the distribution of labour demand across occupations in the network can affect the steady-state unemployment. Most importantly, automation can increase unemployment (and even more so long-term unemployment) during the transition period, even if the total number of jobs stays constant, due to the mismatch between unemployed workers and job vacancies.

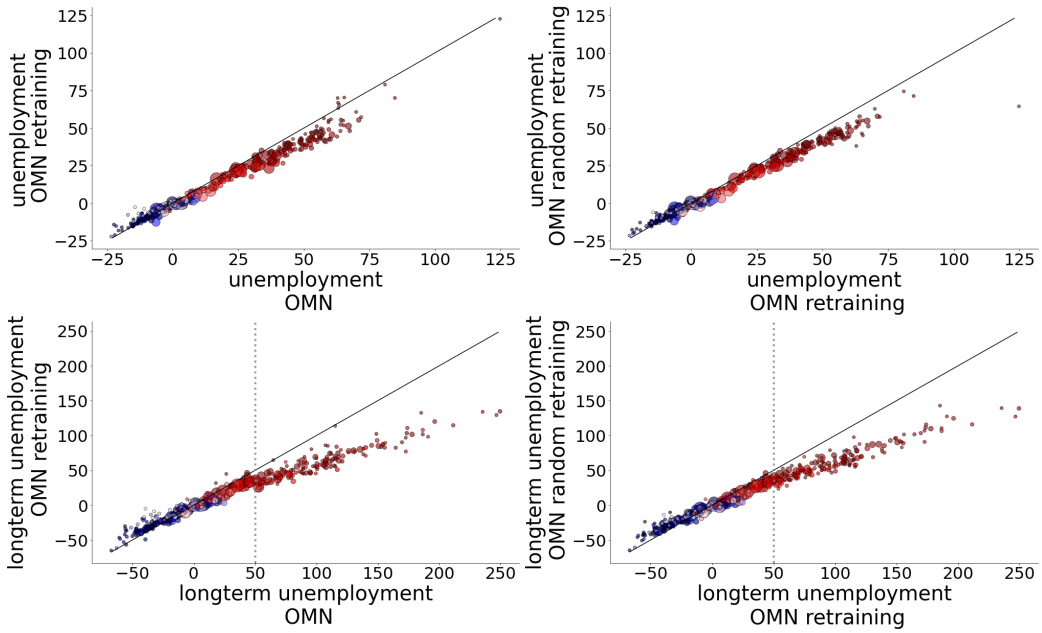


Figure 4.13: **Retraining policies at the occupation level.** In the *left* panel, the targeted retraining towards occupation with long-term unemployment percentage change above 50. In the *right* panel, random retraining. We pick one at random one realisation of the random retraining to show in this plot.

During this transition period, wages will play an important role. For the Frey and Osborne shock, our results suggest that low wage occupations are more likely than high wage occupations to see an increase in their long-term unemployment due to the frictions imposed by the occupational mobility network.

Our work complements previous efforts that have studied automation and job displacement based on the *task approach* (Acemoglu & Autor 2011, Acemoglu & Restrepo 2018, Autor 2013), but provides a networks perspective on job transitions that goes beyond classifying workers into low, middle, and high skill categories. Our work is also closely related to work that has used networks to study the effects of labour market frictions (Axtell et al. 2016, Guerrero & Axtell 2013, Neffke et al. 2018) or the propagation of economic shocks (Diodato & Weterings 2014, Dworkin 2019, Nimczik 2017). However, our study shows that indirect network effects of occupational mobility can be crucial for estimating how automation may increase the unemployment and long-term unemployment of different occupations.

Some studies have focused exclusively on the automatability of occupations when assessing the outlook of workers. Here, we proposed a wider view by considering the automatability of occupations and workers' possibilities for transitioning into

occupations with open vacancies. In some cases, this perspective yields different and seemingly counter-intuitive results, where workers in some occupations at high risk of automation may actually have better employment prospects than workers is seemingly 'safer' occupations.

Our model can be particularly useful in helping policymakers target employment assistance packages and skill development programs to workers who are more likely to face longer periods of unemployment. Our results suggest that there is scope for retraining policies to leverage the occupational mobility network structure to reduce the adverse effects of automation. And while this particular chapter has focused on labour market shocks relating to automation, our model is quite general and, as we show in the next chapter, it can be used to understand other phenomenon affecting the labour market.

Chapter 5

The COVID-19 pandemic and the labour market

5.1 Introduction

The COVID-19 pandemic is having an unprecedented impact on societies around the world. As governments mandate social distancing practices and instruct non-essential businesses to close to slow the spread of the outbreak, there is significant uncertainty about the effect such measures will have on lives and livelihoods. While demand for specific sectors such as grocery stores increased in the early weeks of the pandemic, other sectors such as air transportation and tourism have seen demand for their services evaporate. At the same time, many sectors are experiencing issues on the supply side, as governments curtail the activities of non-essential industries and workers are confined to their homes.

In this chapter we present early predictions[1] of the economic shocks workers of different occupations and industry may experience due to the COVID-19 pandemic, and explore second order effects using our labour market model. We aim to provide analytical clarity about the supply and demand shocks caused by public health measures and changes in preferences caused by avoidance of infection. We estimate (i) supply-side reductions due to the closure of non-essential industries and workers not being able to perform their activities at home, and (ii) demand-side changes due to peoples' immediate response to the pandemic, such as reduced demand for goods or services that are likely to place people at risk of infection (e.g. tourism).

Several researchers have already provided estimates of the supply shock from

[1]The shock predictions presented in this chapter was prepared in March and early April 2020 and released on 16 April (del Rio-Chanona et al. 2020). The shocks presented in this chapter contain only minor changes rather than comprehensive updates.

labour supply (Dingel & Neiman 2020, Hicks et al. 2020, Koren & Pető 2020). Here we improve on these efforts in three ways: (i) we propose a methodology for estimating how much work can be done from home based on work activities, (ii) we identify industries for which working from home is irrelevant because the industries are considered essential, and (iii), we compare our estimated supply shocks to estimates of the demand shock, which in many industries is the more relevant constraint on output.

A number of papers have also emphasized the importance of demand-side factors in the pandemic. For example, Guerrieri et al. (2020) showed that low substitutability in consumption, asymmetric labour supply shocks can lead to reductions in demand that are higher than the initial shock. Recent work has also shown that demand shocks have predominantly affected low-wage workers (Chetty et al. 2020). Furthermore, consumption has decreased more amongst the older and richer population (Eichenbaum et al. 2020).

To see why it is important to compare supply and demand shocks, consider the following thought experiment. Following social-distancing measures, suppose industry i is capable of producing only 70% of its pre-crisis output, e.g. because workers can produce only 70% of the output while working from home. If consumers reduce their demand by 90%, the industry will produce only what will be bought, that is, 10%. If instead consumers reduce their demand by 20%t, the industry will not be able to satisfy demand but will produce everything it can, that is, 70%. In other words, the experienced first-order reduction in output from the immediate shock will be the greater of the supply and the demand shock.

It is important to stress that the shocks that we predict here should not be interpreted as the *overall impact* of the COVID-19 pandemic on the economy. Deriving overall impact estimates involves modelling second-order effects, such as the additional reductions in demand as workers who are stood down or laid off experience a reduction in income and additional reductions in supply as potential shortages propagate through supply chains. Further effects, such as cascading firm defaults, which can trigger bank failures and systemic risk in the financial system, could also arise. Understanding these impacts requires a model of the macro-economy and financial sector. For example, Pichler et al. (2019) present results from such an economic model, but we make our estimates of first-order impacts available separately here, for other researchers or governments to build upon or use in their own models.

This chapter is structured as follows. In Section 5.2, we describe our methodology for estimating supply shocks, which involves developing a new Remote Labour Index (RLI) for occupations and combining it with a list of essential industries. Section 5.3

discusses likely demand shocks based on estimates developed by the U.S. Congressional Budget Office (2006) to predict the potential economic effects of an influenza pandemic. In Section 5.4, we show a comparison of the supply and demand shocks across different industries and occupations and identify the extent to which different activities are likely to be constrained by supply or demand. This section also explores which occupations are more exposed to infection and make comparisons to wage and occupation-specific shocks. In Section 5.5, we use the labour market model we developed in Chapter 2 to estimate the impact of the COVID-19 pandemic on employment. Finally, in Section 5.7, we discuss our findings in light of existing research and outline avenues for future work. We also make all of our data available in a continuously updated online repository.

5.2 Supply shocks

Supply shocks from pandemics are mostly thought of as labour supply shocks. Several pre-COVID-19 studies focused on the direct loss of labour from death and sickness (e.g. McKibbin & Sidorenko (2006), Santos et al. (2013)), although some have also noted the potentially large impact of school closure (Keogh-Brown et al. 2010). McKibbin & Fernando (2020) consider (among other shocks) reduced labour supply due to mortality, morbidity due to infection, and morbidity due to the need to care for affected family members. In countries where social distancing measures are in place, social distancing measures will have a much larger economic effect than the direct effects from mortality and morbidity. This is in part because if social distancing measures work, only a small share of population will be infected and die eventually. Appendix II.4.1 provides more quantitative estimates of the direct mortality and morbidity effects and argues that the they are likely to be at least an order of magnitude smaller than those due to social distancing measures, especially if the pandemic is contained.

For convenience we neglect mortality and morbidity and assume that the supply shocks are determined only by the amount of labour that is withdrawn due to social distancing. We consider two key factors: (i) the extent to which workers in given occupations can perform their requisite activities at home and (ii) the extent to which workers are likely to be unable to come to work due to being in non-essential industries. We quantify these effects on both industries and occupations. Figure 5.1 gives a schematic overview of how we predict industry and occupation specific supply

shocks. We explain this in qualitative terms in the next few pages; for a formal mathematical description see Appendix II.1.1.

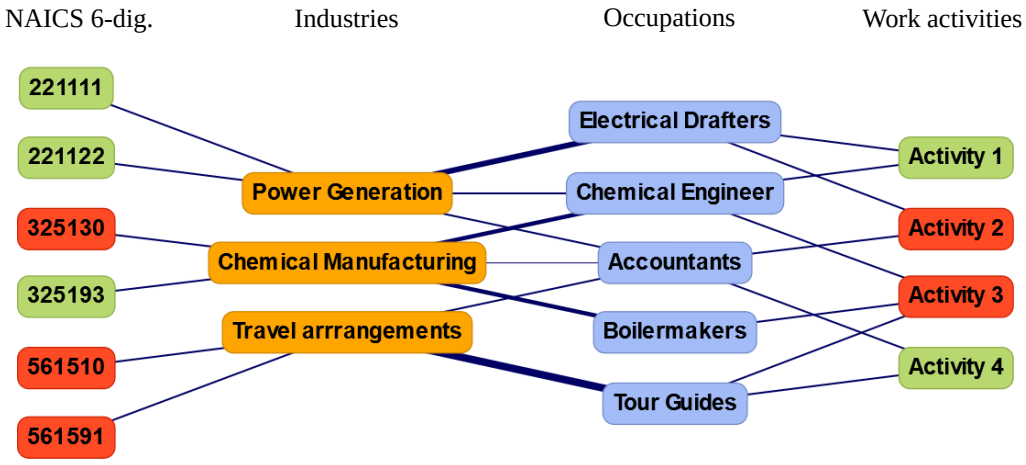


Figure 5.1: **A schematic network representation of supply side shocks.** The nodes to the left represent the list of essential industries at the NAICS 6-digit level. A green node indicates essential, a red node non-essential. The orange nodes (center-left) are more aggregate industry categories (e.g. 4-dig. NAICS or the BLS industry categories) for which further economic data is available. These two sets of nodes are connected through industry concordance tables. The blue nodes (center-right) are different occupations. A weighted link connecting an industry category with an occupation represents the number of people of a given occupation employed in each industry. Nodes on the very right are O*NET work activities. Green work activities mean that they can be performed from home, while red means that they cannot. O*NET provides a mapping of work activities to occupations.

5.2.1 How much work can be performed from home?

Similar to (Dingel & Neiman 2020), we use O*NET data to predict the degree to which workers of different occupations can work from home. However, we to a more granular level than 'work context', and instead draw on O*NET's 'intermediate work activity' data, which provides a list of the activities performed by each occupation based on a list of 332 possible work activities. For example, a Nurse undertakes activities such as "maintain health or medical records", "develop patient or client care or treatment plans" and "operate medical equipment', while a Computer Programmer performs activities such as "resolve computer programs", "program computer systems or production equipment" and "document technical designs, producers or activities". In Fig 5.1 these work activities are illustrated by the rightmost set of nodes.

Which work activities can be performed from home? Four of the authors of our paper del Rio-Chanona et al. (2020) independently assigned a subjective binary rating to each work activity as to whether it could successfully be performed at home. The individual results were in broad agreement. Based on the responses, we assigned an overall consensus rating to each work activity[2]. Ratings for each work activity are available in an online data repository[3]. While O*NET maps each intermediate work activity to 6-digit O*NET occupation codes, employment information from the U.S. Bureau of Labor Statistics (BLS) is available for the 4-digit 2010 Standard Occupation Scheme (SOC) codes, so we mapped O*NET and SOC codes using a crosswalk available from O*NET.[4] Our final sample contains 740 occupations.

[2]An activity was considered to be able to performed at home if three or more respondents rated this as true. We also undertook a robustness analysis where an activity was considered to able to be performed at home based on two or more true ratings. Results remained fairly similar. In post-survey discussion, we agreed that the most contentious point is that some work activities might be done from home or not, depending on the industry in which it is performed.

[3]https://zenodo.org/record/3751068

[4]Available at https://www.onetcenter.org/crosswalks.html.

Occupation	**RLI**
Credit Analysts	1.00
Insurance Underwriters	1.00
Tax Preparers	1.00
Mathematical Technicians	1.00
Political Scientists	1.00
Broadcast News Analysts	1.00
Operations Research Analysts	0.92
Eligibility Interviewers, Government Programs	0.92
Social Scientists and Related Workers, All Other	0.92
Technical Writers	0.91
Market Research Analysts and Marketing Specialists	0.90
Editors	0.90
Business Teachers, Postsecondary	0.89
Management Analysts	0.89
Marketing Managers	0.88
Mathematicians	0.88
Astronomers	0.88
Interpreters and Translators	0.88
Mechanical Drafters	0.86
Forestry and Conservation Science Teachers, Postsecondary	0.86
...	...
Bus and Truck Mechanics and Diesel Engine Specialists	0.00
Rail Car Repairers	0.00
Refractory Materials Repairers, Except Brickmasons	0.00
Musical Instrument Repairers and Tuners	0.00
Wind Turbine Service Technicians	0.00
Locksmiths and Safe Repairers	0.00
Signal and Track Switch Repairers	0.00
Meat, Poultry, and Fish Cutters and Trimmers	0.00
Pourers and Casters, Metal	0.00
Foundry Mold and Coremakers	0.00
Extruding and Forming Machine Setters, Operators, and Tenders, Synthetic and Glass Fibers	0.00
Packaging and Filling Machine Operators and Tenders	0.00
Cleaning, Washing, and Metal Pickling Equipment Operators and Tenders	0.00
Cooling and Freezing Equipment Operators and Tenders	0.00
Paper Goods Machine Setters, Operators, and Tenders	0.00
Tire Builders	0.00
Helpers–Production Workers	0.00
Production Workers, All Other	0.00
Machine Feeders and Offbearers	0.00
Packers and Packagers, Hand	0.00

Table 5.1: **Top and bottom 20 occupations ranked by Remote Labor Index (RLI)**, based on proportion of work activities that can to be performed by home. There are 44 occupations with an RLI of zero; we show only a random sample.

From work activities to occupations. We then created a Remote Labour Index (RLI) for each occupation by calculating the proportion of an occupation's work activities that can be performed at home. An RLI of 1 would indicate that all of the activities associated with an occupation could be undertaken at home, while an RLI of 0 would indicate that none of the occupation's activities could be performed at home.[5] The resulting ranking of each of the 740 occupations can be found in the online repository (see footnote 5). While the results are not perfect[6], most of the rankings make sense. For example, in Table 5.1, we show the top 20 occupations having the highest RLI ranking. Some occupations such as credit analysts, tax preparers and mathematical technician occupations are estimated to be able to perform 100% of their work activities from home. Table 5.1 also shows a sample of the 43 occupations with an RLI ranking of zero, i.e. those for which there are no activities that are able to be performed at home.

To provide a broader perspective of how the RLI differs across occupation categories, Fig. 5.2 shows a series of box-plots indicating the distribution of RLI for each 4-digit occupation in each 2-digit SOC occupation category. We have ordered 2-digit SOC occupations in accordance with their median values. Occupations with the highest RLI relate to education, training and library; computer and mathematical, and business and financial roles, while occupations relating to production; farming, fishing and forestry, and construction and extraction tend to have lower RLI.

From occupations to industries. We next map the RLI to industry categories to quantify industry-specific supply shocks from social distancing measures. We obtain occupational compositions per industry from the BLS, which allows us to match 740 occupations to 277 industries[7].

[5]We omitted ten occupations that had less than five work activities associated with them. These occupations include Insurance Appraisers Auto Damage; Animal Scientists; Court Reporters; Title Examiners, Abstractors, and Searchers; Athletes and Sports Competitors; Shampooers; Models; Fabric Menders, Except Garment; Slaughterers and Meat Packers' and Dredge Operators.

[6]There are a few cases that we believe are misclassified. For example, two occupations with a high RLI that we think cannot be performed remotely are real estate agents (RLI = 0.7) and retail salespersons (RLI = 0.63). However, these are exceptions – in most cases the rankings make sense. The full list can be examined on our online repository at https://zenodo.org/record/3751068. We believe the problems will be fixed when we redo the analysis using fine grained work activities, and we doubt that our results will be qualitatively changed.

[7]We use the May 2018 Occupational Employment Statistics (OES) estimates on the level of 4-digit NAICS (North American Industry Classification System), file *nat4d_M2018_dl*, which is available at https://www.bls.gov/oes/tables.htm under *All Data*. Overall, our merged dataset covers 136.8 out of 144 million employed people (95%) initially reported in the OES.

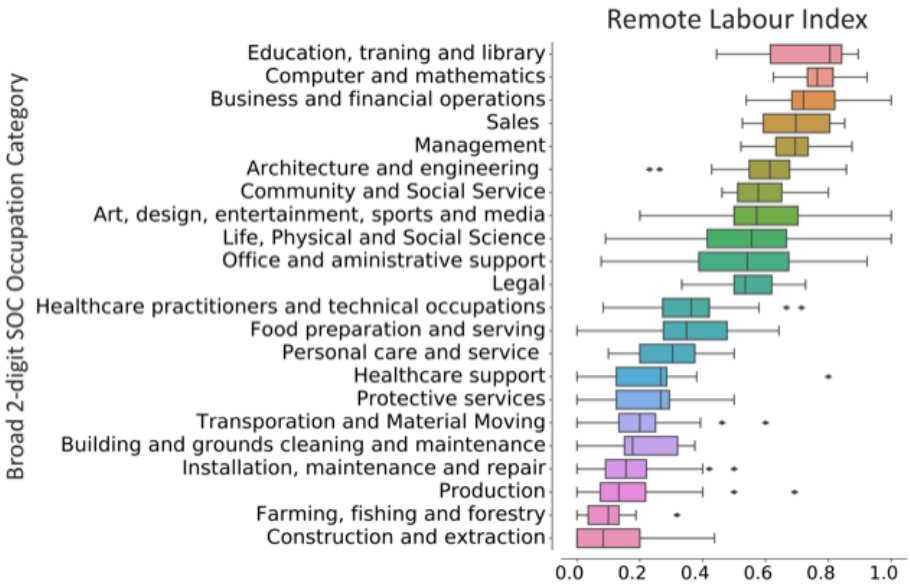


Figure 5.2: **Distribution of Remote Labour Index across occupations**. Boxplots showing the distribution of RLI for each 4-digit occupation in each 2-digit SOC occupation category.

In Fig. 5.3, we show the RLI distribution for each 4-digit occupation category falling within each broad 2-digit NAICS category. Similar to Fig. 5.2, we have ordered the 2-digit NAICS industry categories in accordance with the median values of each underpinning distribution. As there is a greater variety of different types of occupations within these broader industry categories, distributions tend to be much wider. Industries with the highest median RLI values relate to Information, Finance and Insurance, and Professional, Science and Technical Services, while industries with the lowest median RLI relate to Agriculture, Forestry, Fishing and Hunting and Accommodation and Food Services.

In the Appendix, we show industry-specific RLI values for the more detailed 4-digit NAICS industries. To arrive at a single number for each 4-digit industry, we compute the employment-weighted average of occupation-specific RLIs. The resulting industry-specific RLI can be interpreted as a rough estimate of the fraction of jobs which can be performed from home for each industry.

5.2.2 Which industries are "essential"?

Across the world, many governments have mandated that certain industries deemed 'essential' should remain open over the COVID-19 crisis duration. What constitutes an 'essential' industry has been the subject of significant debate, and it is likely that

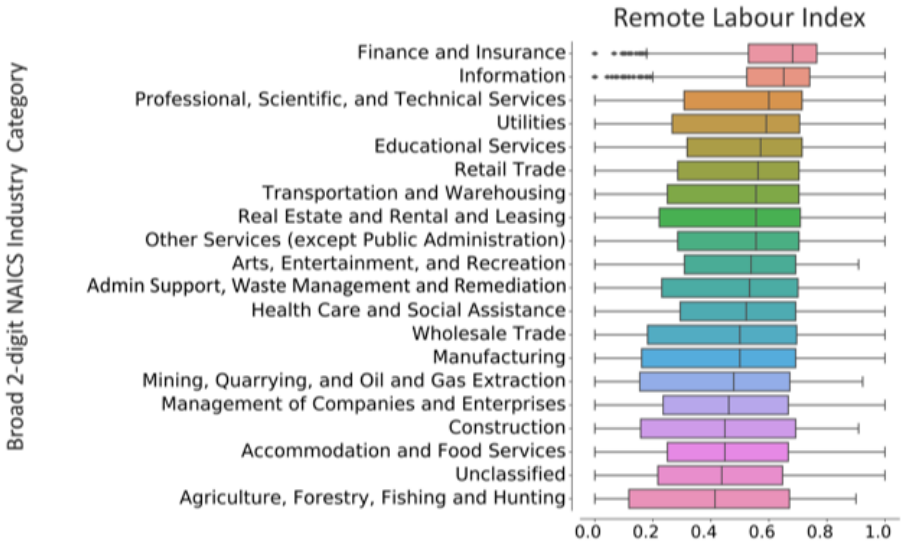


Figure 5.3: **Distribution of Remote Labour Index across industries**. Boxplots showing the distribution of RLI for each 4-digit occupation in each 2-digit NAICS Industry category.

the endorsed set of essential industries will vary across countries. As the U.S. government has not produced a definitive list, here we draw on the list of essential industries developed by Italy and assume it can be applied, at least as an approximation, to other countries such as the U.S. as well. This list has two key advantages. First, as Italy was one of the countries affected earliest and most severely, it was one of the first countries to invest significant effort considering which industries should be deemed essential. Second, Italy's list of essential industries includes NACE industrial classification codes, which can be mapped to the NAICS industry classification we use to classify industrial employment in this chapter.[8]

Table 5.2 shows the total numbers of NAICS essential industries at the 6-digit and 4-digit level. More than 50% of 6-digit NAICS industries are considered essential. At the 6-digit level the industries are either classified as essential, and assigned essential score 1, or non-essential and assigned essential score 0. Unfortunately, it is not possible to translate this directly into a labour force proportion as BLS employment data at detailed occupation and industry levels are only available at the NAICS

[8]Mapping NACE industries to NAICS industries is not straightforward. NACE industry codes at the 4-digit level are internationally defined. However, 6-digit level NACE codes are country specific. Moreover, the list of essential industries developed by Italy involves industries defined by varying levels of aggregation. Most essential industries are defined at the NACE 2-digit and 4-digit level, with a few 6-digit categories thrown in for good measure. As such, much of our industrial mapping methodology involved mapping from one classification to the other by hand. We provide a detailed description of this process in Appendix II.2.1.

4-digit level. To derive an estimate at the 4-digit level, we assume that labour in a NAICS 4-digit code is uniformly distributed over its associated 6-digit codes. We then assign an essential 'share' to each 4-digit NAICS industry based on the proportion of its 6-digit NAICS industries that are considered essential. (The distribution of the essential share over 4-digit NAICS industries is shown in the Appendix II). Based on this analysis, we estimate that about 92 million (or 67%) of U.S. workers are currently employed in essential industries.

Total 6-digit NAICS industries	1057
Number of essential 6-digit NAICS industries	612
Fraction of essential industries at 6-digit NAICS	0.58
Total 4-digit NAICS industries in our sample	277
Average rating of essential industries at 4-digit NAICS	0.56
Fraction of labour force in essential industries	0.67

Table 5.2: **Essential industries.** Essential industries at the 6-digit level and essential 'share' at the 4-digit level. Note that 6-digit NAICS industry classifications are binary (0 or 1) whereas 4-digit NAICS industry classifications can take on any value between 0 and 1.

5.2.3 Supply shock: non-essential industries unable to work from home

Having analysed both the extent to which jobs in each industry are essential and the likelihood that workers in a given occupation can perform their requisite activities at home, we now combine these to consider the overall first-order effect on labour supply in the US. In Fig. 5.4, we plot the Remote Labour Index of each occupation against the fraction of that occupation employed in an essential industry. Each circle in the scatter plot represents an occupation; the circles are sized proportional to current employment and colour coded according to the median wage in each occupation.

Figure 5.4 indicates the vulnerability of occupations due to supply-side shocks. Occupations in the lower left-hand side of the plot (such as Diswashers, Rock Splitters and Logging Equipment Operators) have lower RLI scores (indicating they are less able to work from home) and are less likely to be employed in an essential industry. If we consider only the immediate supply-side effects of social distancing, workers in these occupations are more likely to face reduced work hours or be at risk of losing their job altogether. In contrast, occupations on the upper right-hand side of the plot, such as Credit Analysis, Political Scientists and Operations Research Analysts) have higher RLI scores and are more likely to employed in an essential industry. These

occupations are less economically vulnerable to the supply-side shocks (though, as we discuss in the next section, they could still face employment risks due to first-order demand-side effects). Occupations in the upper-left hand side of the plot (such as Farmworkers, Healthcare Support Workers and Respiratory Therapists) are less likely to be able to perform their job at home, but since they are more likely to be employed in an essential industry their economic vulnerability from supply-side shocks is lower. Interestingly, there are relatively few occupations on the lower-right hand side of the plot. This indicates that occupations that are predominantly employed in non-essential industries tend to be less able to perform their activities at home.

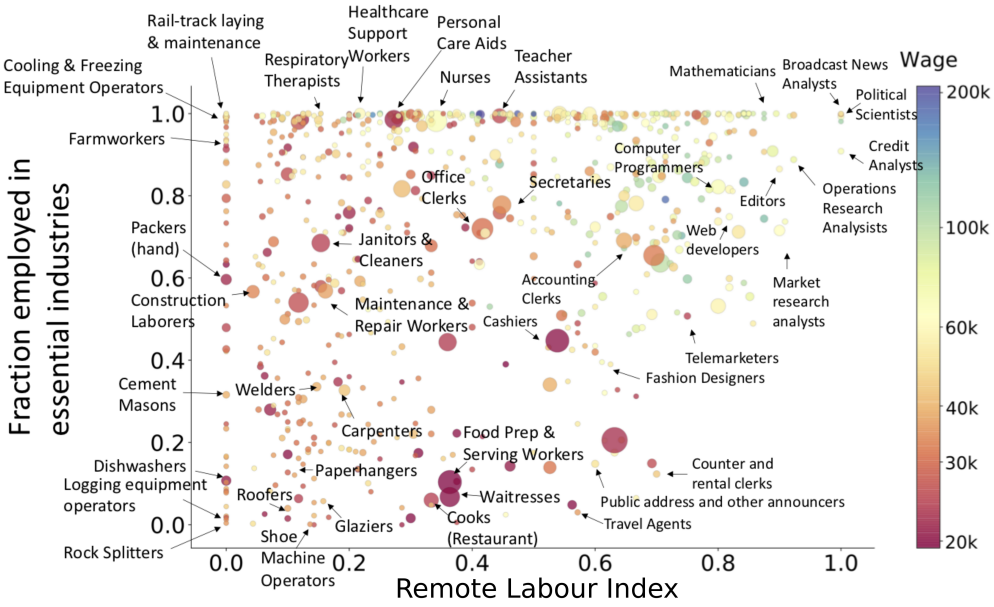


Figure 5.4: **Fraction employed in an essential industry vs Remote Labour Index for each occupation**. Omitting the effect of demand reduction, the occupations in the lower left corner, with a small proportion of workers in essential industries and a low Remote Labour Index, are the most vulnerable to loss of employment due to social distancing.

To help visualize the problem we provide a summary in the form of a Venn diagram in Fig. 5.5. Before the pandemic, 33% of workers were employed in non-essential jobs. 56% of workers cannot do their job remotely. 19% of workers are in the intersection corresponding to non-essential jobs that cannot be performed remotely. In addition, there are 30% of workers in essential industries that can also work from home.[9]

[9]In fact we allow for a continuum between the ability to work from home, and an industry can be partially essential.

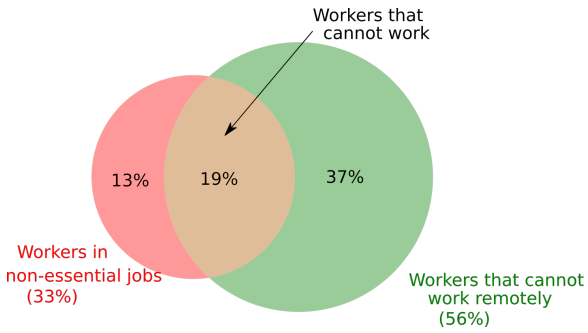


Figure 5.5: **Workers that cannot work**. On the left, the percentage of workers in a non-essential job (33% in total). On the right, the percentage of workers that cannot work remotely (56% in total). The intersection is the set of workers that cannot work, which is 19% of all workers. A remaining 30% of workers are in essential jobs where they can work remotely.

5.3 Demand shock

The pre-COVID-19 literature on epidemics and the discussions of the current crisis make it clear that epidemics strongly influence patterns of consumer spending. Consumers are also likely to seek to reduce their risk of exposure to the virus and decrease demand for products and services that involve close contact with others. In the early days of the outbreak, stockpiling behaviour also drives a direct demand increase in the retail sector (Baker et al. 2020).

Estimates from the CBO. Our estimates of the demand shock are based on expert estimates developed by the U.S. Congressional Budget Office (2006) that attempted to predict the potential impact of an influenza pandemic. Similar to the current COVID-19 pandemic, this analysis assumes that demand is reduced due to the desire to avoid infection. While the analysis is highly relevant to the present COVID-19 situation, it is important to note that the estimates are "extremely rough" and "based loosely on Hong Kong's experience with SARS". The CBO provides two scenarios. We take the severe scenario, which "describes a pandemic that is similar to the 1918-1919 Spanish flu outbreak. It incorporates the assumption that a particularly virulent strain of influenza infects roughly 90 million people in the United States and kills more than 2 million of them".

In this chapter, we simply take the CBO estimates as immediate (first-order) demand-side shocks. The CBO lists demand side estimates for broad industry categories, which we mapped to the 2-digit NAICS codes by hand. Table 5.3 shows the

CBO's estimates of the percent decrease in demand by industry, and Table II.3 in Appendix II shows the full mapping to 2-digit NAICS.

Broad industry name	Severe Scenario Shock
Agriculture	-10
Mining	-10
Utilities	0
Construction	-10
Manufacturing	-10
Wholesale trade	-10
Retail trade	-10
Transportation and warehousing (including air, rail and transit)	-67
Information (Published, broadcast)	0
Finance	0
Professional and business services	0
Education	0
Healthcare	15
Arts and recreation	-80
Accommodation/food service	-80
Other services except government	-5
Government	0

Table 5.3: Demand shock by sector according to the Congressional Budget Office (2006)'s severe scenario.

These estimates, of course, are far from perfect. They are based on expert estimates made more than ten years ago for an hypothetical pandemic scenario. It is not entirely clear if they are for gross output or for final (consumer) demand. However, in Appendix II.5, we describe three other sources of consumption shocks (Keogh-Brown et al. 2010, Muellbauer 2020, OECD 2020) that provide broadly similar estimates by industry or spending category.

Transitory and permanent shocks. An important question is whether demand reductions are just postponed expenses, and if they are permanent (Mann 2020, Keogh-Brown et al. 2010). (Baldwin & Weder di Mauro 2020) also distinguishes between "practical" (the impossibility to shop) and "psychological" demand shocks (the wait-and see attitude adopted by consumers facing strong uncertainty.) We see three possibilities: (i) expenses in a specific good or service are just delayed but will take place later, for instance if households do not go to the restaurant this quarter, but go twice as much as they would normally during the next quarter; (ii) expenses

are not incurred this quarter, but will come back to their normal level after the crisis, meaning that restaurants will have a one-quarter loss of sales; and (iii) expenses decrease to a permanently lower level, as household change their preferences in view of the 'new normal'. Appendix II.5 reproduces the scenario adopted by Keogh-Brown et al. (2010), which distinguishes between delay and permanently lost expenses.

Other components of aggregate demand. We do not include direct shocks to investment, net exports, and net inventories. Investment is typically very pro-cyclical and is likely to be strongly affected, with direct factors including cash-flow reductions and high uncertainty (Boone 2020). The impact on trade is likely to be strong and possibly permanent (Baldwin & Weder di Mauro 2020), but would affect exports and imports in a relatively similar way, so the overall effect on net exports is unclear. Finally, it is likely that due to the disruption of supply chains, inventories will be run down so the change in inventories will be negative (Boone 2020).

5.4 Combining supply and demand shocks

Having described both supply and demand-side shocks, we now compare the two at the industry and occupation level.

5.4.1 Industry-level supply and demand shocks.

Figure 5.6 plots the demand shock against the supply shock for each industry. The radius of the circles is proportional to the gross output of the industry[10].

Essential industries have no supply shock and so lie on the horizontal '0' line. Of these industries, sectors such as Utilities and Government experience no demand shock either, since immediate demand for their output is assumed to remain the same. Following the CBO predictins, health experiences an increase in demand and consequently lies below the identity line. Transport on the other hand experiences a reduction in demand and lies well above the identity line. This reflects the current situation, where trains and buses are running because they are deemed essential, but

[10]Since relevant economic variables such as total output per industry are not extensively available on the NAICS 4-digit level, we need to further aggregate the data. We derive industry-specific total output and value added for the year 2018 from the BLS input-output accounts, allowing us to distinguish 169 private industries for we which we can also match the relevant occupation data. The data can be downloaded from https://www.bls.gov/emp/data/input-output-matrix.htm.

they are mostly empty[11]. Non-essential industries such as Entertainment, Restaurants and Hotels, experience both a demand reduction (due to consumers seeking to avoid infection) and a supply reduction (as many workers are unable to perform their activities at home). Since the demand shock is bigger than the supply shock they lie above the identity line. Other non-essential industries such as manufacturing, mining and retail have supply shocks that are larger than their demand shocks and consequently lie below the identity line.

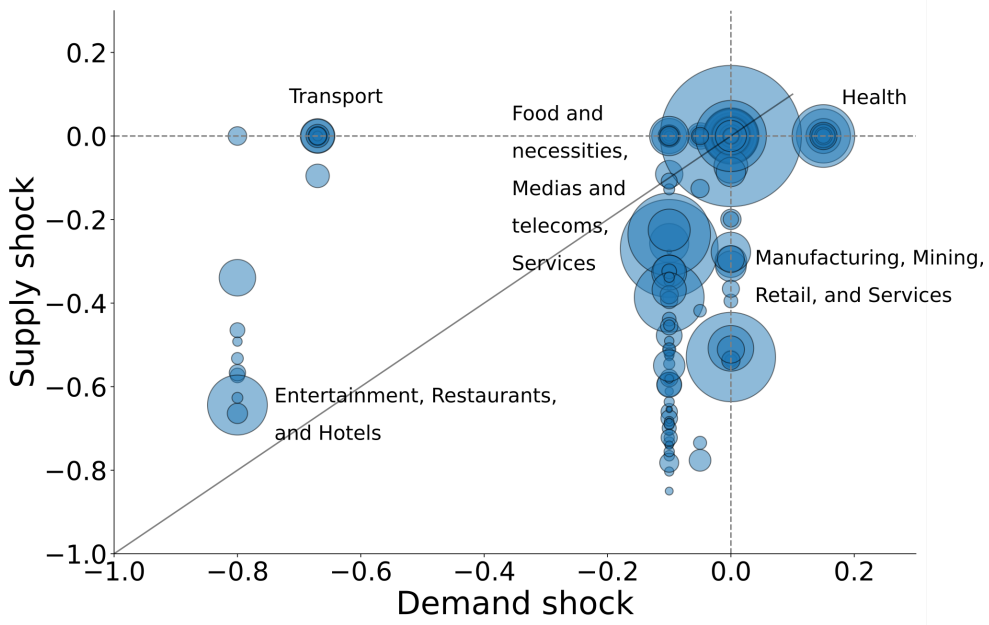


Figure 5.6: **Supply and demand shocks for industries**. Each circle is an industry, with radius proportional to gross output. Many industries experience exactly the same shock, hence the superposition of some of the circles. Labels correspond to broad classifications of industries.

5.4.2 Occupation-level supply and demand shocks.

In Fig. 5.7, we show the supply and demand shocks for occupations rather than industries. For each occupation this comparison indicates whether it faces a risk of unemployment due the lack of demand or a lack of supply in its industry.

Several health-related occupations such as nurses, medical equipment preparers and healthcare social workers are employed in industries experiencing increased demand. Occupations such as airline Pilots, lodging managers and hotel desk clerks

[11]Some transport companies such as Transport for London have reduced their traffic, but this is a second-order impact, not a first-order shock. This decrease in output resulting from reduced demand should be modelled as an endogenous effect, not a first-order shock.

face relatively mild supply shocks and strong demand shocks (as consumers reduce their demand for travel and hotel accommodation) and consequently lie above the identity line. Other occupations such as stonemasons, rock splitters, roofers and floor layers face a much stronger supply shock as it is very difficult for these workers to perform their job at home. Finally occupations such as cooks, dishwashers and waiters suffer both adverse demand shocks (since demand for restaurants is reduced) and supply shocks (since they cannot work from home and tend not to work in essential industries).

For the majority of occupations the supply shock is larger than the demand shock. This is not surprising given that we only consider immediate shocks and no feedback-loops in the economy. We expect that once second-order effects are considered the demand shocks are likely to be much larger.

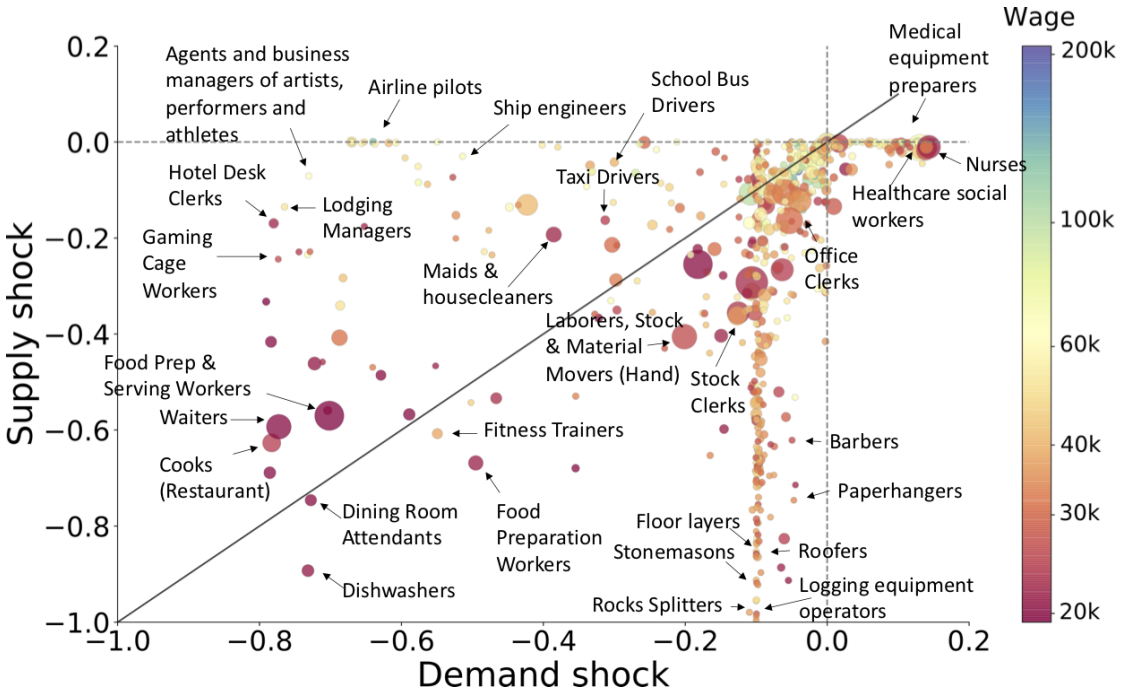


Figure 5.7: **Supply and demand shocks for occupations.** Each circle is an occupation with radius proportional to employment. Circles are colour coded by the log median wage of the occupation. The correlation between wages and demand shocks is 0.26 (p-value $= 2.84 \times 10^{-13}$) and between wages and supply shocks is 0.40 (p-value $= 1.5 \times 10^{-30}$).

5.4.3 Aggregate shocks

We now aggregate shocks to obtain estimates for the whole economy. We assume that, in a given industry, the total shock will be the largest of the supply or demand shocks.

For example, if an industry faces a 30% demand shock and 50% supply shock because 50% of the industry's workforce cannot work, the industry is assumed to experience an overall 50% shock to output. For simplicity, we assume a proportional relationship between output and labour: i.e. when industries are supply constrained, output is reduced by the same fraction as the reduction in labour supply. This assumption also implies that the demand shock that workers of an industry experience equals the industry's output demand shock in percentage terms. For example, if transport faces a 67% demand shock and no supply shock, bus drivers working in this industry will experience an overall 67% employment shock. The shock on occupations depends on the prevalence of each occupation in each industry (see Appendix II.1 for details). We then aggregate shocks in three different ways.

First we estimate the decline in employment by weighting occupation-level shocks by the number of workers in each occupation. Second, we estimate the decline in total wages paid by weighting occupation-level shocks by the share of occupations in the total wage bill. Finally, we estimate the decline in GDP by weighting industry level shocks by the share of industries in GDP[12].

Table 5.4 shows the results. In all cases, by definition, the total shock is larger than both the supply and demand shock, but smaller than the sum. Overall, the supply shock appears to contribute more to the total shock than does the demand shock.

The wage shock is around 16% and is lower than the employment shock (23%). This makes sense, and reflects a fact already well acknowledged in the literature (Office for National Statistics 2020, Adams-Prassl et al. 2020) that occupations that are most affected tend to have lower wages. We discuss this more below.

The shock on value added (20%) is higher than the shock to the wage bill. This also makes sense; because value added is the sum of the wage bill and payment to capital, this suggests that capital intensive industries are hit harder.

For industries and occupations in the health sector, which experience an increase in demand in our predictions, there is no corresponding increase in supply. Table II.1 in Appendix II.1.7 provides the same estimates as Table 5.4, but now assuming that the increased demand for health will be matched by increased supply. This corresponds to a scenario where the healthcare sector would be immediately able to hire as many workers as necessary and pay them at the normal rate. This assumption

[12]Since rents account for an important part of GDP, we make an additional robustness check by considering the Real Estate sector essential. In this scenario the supply and total shocks drop by 3 percentage points.

Aggregate Shock	Employment	Wages	Value Added
Supply	-19	-14	-16
Demand	-13	-8	-7
Total	-26	-16	-20

Table 5.4: **Aggregate shocks to employment, wages and value added**. The size of each shock is shown as a percentage of the pre-pandemic value. Demand shocks include positive values for the health sector. The total shock at the industry level is the minimum of the supply and demand shock, see Appendix II.1. Note that these are only first order shocks (not total impact), and instantaneous values (not annualized).

does not, however, make a significant difference to the aggregate total shock. In other words, the increase in activity in the health sector is unlikely to be large enough to compensate significantly for the losses from other sectors.

5.4.4 Shocks by wage level

To understand how the pandemic has affected workers of different income levels differently, we present results for each wage quartile. The results are in Table 5.4, columns $q_1 \dots q_4$.[13] In Table 5.5, we show employment shocks by wage quartile. This table shows that workers whose wages are in the lowest quartile (lowest 25%) will bear much higher relative losses than workers whose wages are in the highest quartile. Our results confirm the survey evidence reported by the Office for National Statistics (2020) and Adams-Prassl et al. (2020), showing that low-wage workers are more strongly affected by the COVID crisis in terms of lost employment and lost income. Furthermore, Table 5.5 shows how the total loss of wages in the economy in split amongst the different quartiles. Even though the lowest quartile have lower salaries, the shock is so high that they bear the highest share of the total loss.

	q_1	q_2	q_3	q_4	Aggregate
Percentage change in employment	-41	-23	-20	-6	-23
Share of total lost wages (%)	31	24	29	17	-16

Table 5.5: **Total Wages or employment shocks by wage quartile**. We divide workers into wage quartiles based on the average wage of their occupation (q_1 corresponds to the 25% least paid workers). The first row is the number of workers who are vulnerable due to the shock in each quartile divided by the total who are vulnerable. Similarly, the second row is the fraction of whole economy total wages loss that would be lost by vulnerable workers in each quartile. The last column gives the aggregate shocks from Table 4.

[13] As before, Table II.1 in Appendix II.1.7 gives the results assuming positive total shocks for the health sector, but shows that it makes very little difference.

Next we estimate labour shocks at the occupation level. We define the labour shocks as the declines in employment due to the total shocks in the industries associated with each occupation. We use Eq. (II.14) (Appendix II.1.7) to compute the labour shocks, which allows for positive shocks in healthcare workers, to suggest an interpretation in terms of a change in labour demand. Figure 5.8 plots the relationship between labour shocks and median wage. A strong positive correlation (Pearson $\rho = 0.41$, p-value $= 5.8 \times 10^{-31}$) is clearly evident, with almost no high wage occupations facing a serious shock.

We have also coloured occupations by their exposure to disease and infection using an index developed by O*NET[14] (for brevity we refer to this index as "exposure to infection"). As most occupations facing a positive labour shock relate to healthcare[15], it is not surprising to see that they have a much higher risk of being exposed to disease and infection. However, other occupations such as janitors, cleaners, maids and childcare workers also face higher risk of infection.

In the next section, we explore how the shocks in Fig. 5.8 propagate, using the labour model we developed in Chapter 2.

5.5 Occupational mobility network effects of the COVID-19 pandemic

In this section we study the behaviour of our labour market model under the COVID-19 shocks predicted in the previous section. The work presented in this section is not a prediction, since this work was done after the first waves of the pandemic. Furthermore, as we discuss in Section 5.6, the shock prediction for the healthcare sector was not precise, and this error will propagate through the model. Nevertheless, since the distribution of the COVID-19 across occupations is different from the distribution of the automation shocks, we consider it interesting to explore how our model behaves under COVID-19 shocks.

The COVID-19 shocks we derived in the previous section were calculated for the SOC classification system. To translate these shocks into the occupational mobility network, we crosswalk the shocks from SOC to ACS codes. Since the crosswalk is many-to-many, there is some noise in the translation of the shocks. At the ACS classification level, the total employment shock is 20%. Figure 5.9 shows how the

[14] https://www.onetonline.org/find/descriptor/result/4.C.2.c.1.b

[15] Our demand shocks do not have an increase in retail, but in the UK supermarkets have been trying to hire several tens of thousands workers (Source: BBC, 21 March, https://www.bbc.co.uk/news/business-51976075). Baker et al. (2020) documents stock piling behaviour in the US.

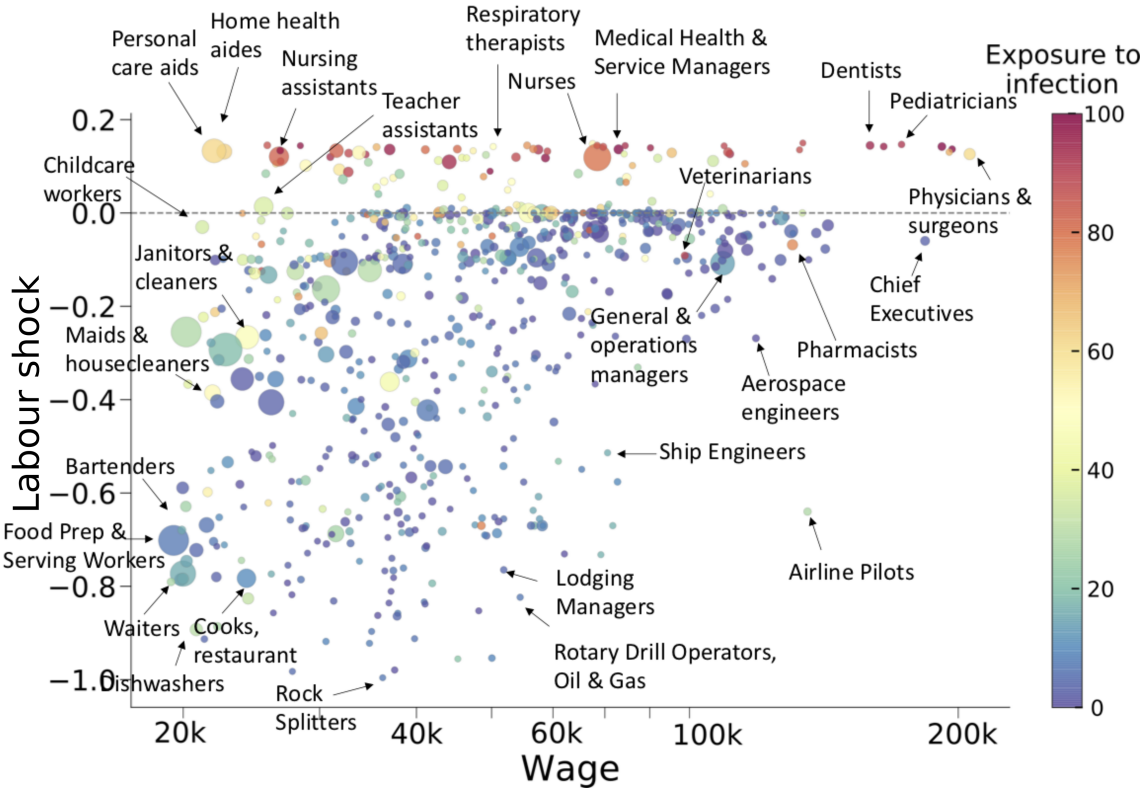


Figure 5.8: **Labour shock vs. median wage for different occupations.** We colour occupations by their exposure to disease and infection. There is a 0.40 correlation between wages and the labour shock (p-value $= 4.4 \times 10^{-40}$). Note the striking lack of high wage occupations with large labour demand shocks.

Remote Labour Index and the essential score are distributed across the occupational mobility network. This figure shows that occupations positioned at the top of the occupational mobility network tend to have a higher Remote Labour Index and essential score. Suggesting that, workers from occupations at the bottom of the network may find that their odds of being unemployed increased due to the pandemic. In the following subsections, we explore this idea further.

Figure 5.9: **Remote Labour Index and Essential Score in the occupational mobility network.** Panel (**A**) corresponds to the distribution of the Remote Labour Index across occupation in the occupational mobility network, where red indicates that the occupation has a high Remote Labour Index and green indicates a low Remote Labour Index. Panel (**B**) the essential score of occupations, where red indicates a high essential score and green a low one. In both panels red occupations are more exposed to job displacement either due to having a low Remote Labour Index or belonging to an essential industry.

5.5.1 COVID-19 as a labour demand shock

We interpret the COVID-19 occupation shocks as a change in the labour demanded from workers of each occupation. Positive shocks imply that there is an increase in labour demand, while negative shocks imply a decrease. According to our predictions, with the exception of some workers in healthcare occupations, most workers would face a decrease in the demand for their labour. Figure—5.10 shows the distribution of the COVID-19 shocks across occupations. Most occupations have mild negative shocks (between zero and -0.2), but several occupations present a considerably large negative shock (-0.2 or stronger), while only a few occupations in healthcare have a positive shock.

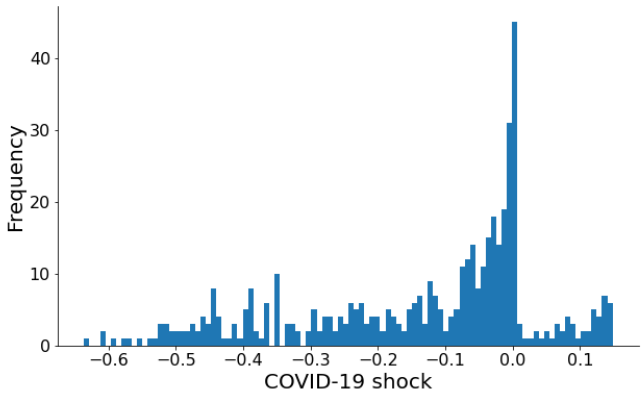


Figure 5.10: **Distribution of the COVID-19 shocks to labour demand.** We plot the distribution across occupations. Some occupations in the healthcare sector experience an increase in demand, while most occupations experience a decrease in demand.

The shocks we computed do not consider that labour contracts and laws are likely to help some workers retain their jobs (Lazear 1990). To take this into account, we diminish the magnitude of the shocks by one-third and interpret these diminished shocks as the percentage change in labour demand for workers of each occupation. The extent to which labour contract and laws helped retain workers is unclear, we consider one-third to be a reasonable assumption but admit that the choice is arbitrary and a more careful analysis would explore this assumption further. In Appendix II, we explore different choices and show that main results are robust.

The COVID-19 shocks have both a labour reallocation effect (e.g. demand for workers in the health care sector increase while demand for workers in the entertainment sector decreases) and an aggregate labor demand effect (the shocks reduce

the aggregate labour demand by 13%). This contrasts with the previous Chapter 4, where we assumed automation shocks only had a labour reallocation effect and did not affect the aggregate labour demand.

Figure 5.11A illustrates the time dynamics we assume for the COVID-19 labour demand shocks. We consider that the shocks happen sharply at the beginning of the year 2020 and do not fade. Arguably, one could expect that labour demand for workers in entertainment and other consumer-facing industries recovers as the pandemic's end nears and we have adjusted towards a "new normal". However, at the time of writing it is unclear when the pandemic will end, and how labour demand may look like in the "new normal". Because of this, we assume the shocks and aggregate decrease in labour demand remains indefinitely, but we measure the effects of the pandemic only during the period 2020–2022 (see grey shading in Fig. 5.11).

5.5.2 Employment outcomes

Aggregate effects. At the aggregate level, the COVID-19 shock sharply increases unemployment and long-term unemployment for both the occupational mobility network and the complete network. Due to the pandemic the unemployment rate increases from 5.3% to 16.5% when we use the occupational mobility network. When we use the complete network the unemployment rate increases from 4.1% to 16.3%. In contrast to the pre-pandemic period, the post-pandemic unemployment rates predicted by the occupational mobility network and the complete network are very similar, with only a 0.2 percentage point difference. This means that the post-pandemic unemployment rate is mostly caused by the 13% reduction in the aggregate labour demand, and to a lesser extent by the labour market frictions imposed by the occupational mobility network.

We observe similar dynamics for the long-term unemployment rate. For both the occupational mobility network and for the complete network the long-term unemployment rate increases sharply, from 2.1% to 12.9% and from 1.0% to 12.6%. Figure 5.11 shows that there is a lag between the time the shock is introduced and the time at which the long-term unemployment starts increasing. This is understandable given the definition of long-term unemployment, which states that an unemployed worker is consider long-term unemployed after 27 weeks, i.e. roughly after half a year of being unemployed.

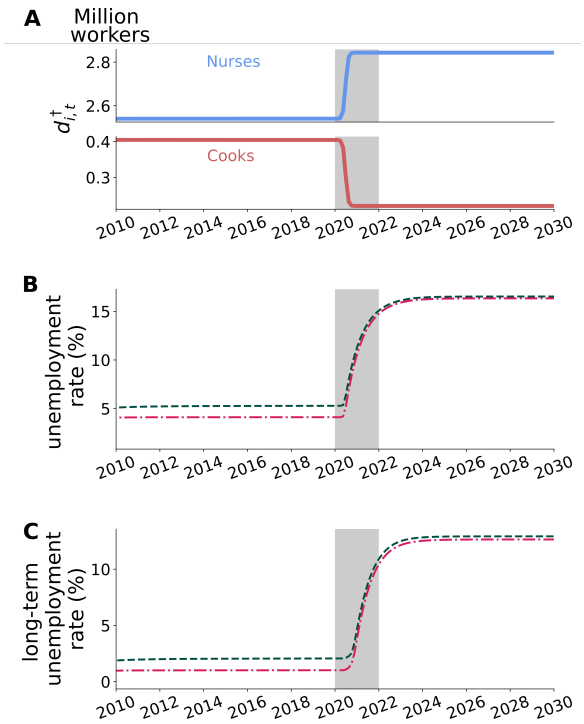


Figure 5.11: **Unemployment and long-term unemployment rate at the aggregate level.** In all panels, the grey area denotes the period during which we measure the percentage change in unemployment and long-term unemployment. Panel (**A**) shows the evolution of the target demand for two occupations. The occupation coloured in blue has a positive COVID-19 shock, while the occupation coloured in red a negative COVID-19 shock. Panels (**B**) and (**C**) show the unemployment and long-term unemployment rates through time. The green line corresponds to the occupational mobility network while the red line to the complete network.

Occupation effects. As in Chapter 4, we measure the effect of the shocks on workers of different occupations using Eqs. (4.6) and (4.7) which give the average unemployment and long-term unemployment rate during period T. We then compare the unemployment rates before the pandemic with the unemployment rate between 2020–2022 as marked by the grey area in Fig. 5.11. In Fig. 5.12, we show the percentage change in the unemployment rates for each occupation, using the occupational mobility network (blue/green dots) and the complete network (orange dots).

Similar to Chapter 4, Fig. 5.12 shows that the network structure plays a crucial role: occupations that experience similar COVID-19 shocks present different changes in the unemployment and long-term unemployment rate. An interesting result is that, when we use the complete network, all occupations with a positive COVID-19 shock have an equal unemployment rate percentage change (i.e. the orange dots in the positive side of the horizontal axis form a straight horizontal line in Fig. 5.12). There are

two reasons behind this result (i) in the complete network all workers have the same probability of becoming unemployed and (ii) workers in occupations with positive COVID-19 have the same probability δ_u of being separated from their job. The latter is not true for workers with a negative COVID-19 shock, since they face an additional occupation-specific probability of being separated $(1-\delta_u)\gamma \max\left\{d^{\dagger}_{i,t} - d_{i,t}, 0\right\}$.

In comparison to the automation shocks, the COVID-19 shock has a much larger effect on the occupation-specific unemployment rates. During the Frey and Osborne automation shock, the maximum percentage increase in occupational unemployment rate was roughly of 125%, while during the COVID-19 shock a few occupations present a percentage increase in their unemployment rate of more than 800%. Although the difference is striking, it is not unexpected given that (i) the COVID-19 shocks is concentrated in one year, while the automation shocks are introduced smoothly during 30 years, with the sharpest decrease spanning 10 years; (ii) during the COVID-19 shock the demand for work does not remain constant, but decreases by 13%.

Finally, as in Chapter 4, we measure the network effects and find a significant negative correlation between the network effects and the median wage of occupations. In other words, we find that low-wage occupations are more likely to face higher unemployment due to labour market frictions imposed by the occupational mobility network than high-wage occupations.

5.6 Validation

The major part of this chapter has focused on presenting predictions made in mid April of the COVID-19 economic shocks. Since then more data has become available, which opens the possibility of validating our predictions. A detailed validation of our predictions is out of the scope this thesis, but we briefly discuss the performance of our shock predictions in this section.

Bartik et al. (2020) presented firm level surveys have asked on the prevalence and productivity of remote work. They compared their results with the work from home index by Dingel & Neiman (2020) and found that this index performs remarkably well. Since our Remote Labour Index is strongly correlated with the work from home index by Dingel & Neiman (2020) and also based on O*NET data, the findings of Bartik et al. (2020) suggest that the Remote Labour Index may also perform well. In the U.S. different states imposed different restrictions during the pandemic. Many states published their guidelines online but few published a list of essential industries using

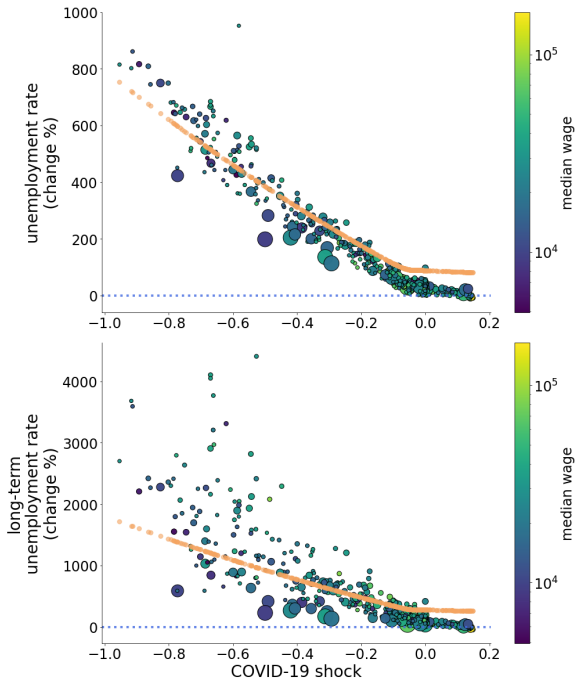


Figure 5.12: **Occupation specific changes in unemployment and long-term unemployment** The green dots are for the occupational mobility network and the orange dots are for the complete network. The size of the green dots is proportional to the employment of the occupation they represent and the colour gradient indicates the median wage of the occupation.

NAICS codes. However, upon inspecting the guidelines for industries in New York[16], we found that the guidelines are close to the Italian list of essential industries. There are a few notable exceptions, among them the fact that New York allowed hotels to open.

To validate the demand shocks, we also review papers that have appeared more recently and contained estimates of consumption changes (Chetty et al. 2020, Carvalho et al. 2020). Taken together, these papers suggest that the shocks from the CBO were broadly accurate. Very large declines in the hospitality, entertainment and transport industries, milder declines in manufacturing, and a more resistant business services sector. The main features that have been missed are the increase in demand, at least early on, in some specific retail categories (groceries), and the decline in health consumption, in sharp contrast with the CBO prediction of a 15% increase. These mistakes may be attributed to the level of aggregation of the CBO shocks. Although there was an increase in demand for the section of the healthcare sector related to

[16]https://www.governor.ny.gov/news/governor-cuomo-issues-guidance-essential-services-under-new-york-state-pause-executive-order

COVID-19 care, the section of healthcare likely decrease their demand due to fear of infection. This obstacle could be overcome in further predictions by using finer-grained shocks for the healthcare sector.

Finally, in a follow-up work, we translated the shocks predicted here into the U.K. economy and predicted the impact of the COVID-19 pandemic using an input-output model Pichler et al. (2020). In this follow-up work we accurately predicted the decrease in GDP for the UK in the first quarter (our prediction was 21.5% decrease and the realised was 22.3%). In comparison, the Bank of England predicted a decrease of roughly 30%.[17]

5.7 Concluding remarks

The first part of this chapter sought to provide quantitative predictions for the U.S. economy of the supply and demand shocks associated with the COVID-19 pandemic. To characterise supply shocks, we developed a Remote Labour Index to estimate the extent to which workers can perform activities associated with their occupation at home and identified which industries are classified as essential vs. non-essential. We also reported plausible estimates of the demand shocks, in an attempt to acknowledge that some industries will have an immediate reduction in output due to a shortfall in demand, rather due to an impossibility to work. We emphasize that these are *predictions*, not measurements: The estimates of the demand shocks were made in 2006, and the RLI and the list of non-essential industries contain no pandemic-specific information, and could have been made at any time. The second part of this chapter used the labour market model introduced in Chapter 2 to estimate how the COVID-19 shocks would propagate through the occupational-mobility.

When considering total shocks at the occupation level, we find that high-wage occupations are relatively immune from both supply and demand-side shocks, while many low-wage occupations are much more economically vulnerable to both. Interestingly, low-wage occupations that are *not* vulnerable to supply and/or demand-side shocks are nonetheless at higher risk of being exposed to COVID-19 (see the colour code in Fig. 5.8). Furthermore, the occupational mobility network effects are more likely to be detrimental (i.e. to increase unemployment) for workers in low-wage occupations. Such findings suggest that the COVID-19 pandemic is likely to exacerbate income inequality in what is already a highly unequal society.

[17] https://www.bankofengland.co.uk/-/media/boe/files/monetary-policy-report/2020/may/monetary-policy-report-may-2020

The work presented in this chapter was the first to compare supply-side shocks with corresponding demand-side shocks at the occupation and industry level. At the time the first version of this chapter was written, the most relevant demand-side estimates available were admittedly 'rough' and only available for aggregate (2-digit) industries. Yet, this suggests that sectors such as transport are more likely to have output constrained by demand-side shocks, while sectors relating to manufacturing, mining and services are more likely to be constrained by supply-side shocks. Entertainment, restaurants and tourism face both very large supply and demand constraints, with demand shocks dominating in our estimates. By quantifying supply and demand shocks by industry, our work speaks to the debate on the possibility of inflation after the crisis. Goodhart & Pradhan (2020) argue that the lockdown causes a massive supply shock that will lead to inflation when demand comes back after the crisis. But as Miles & Scott (2020) note, in many sectors it is not obvious that demand will come back immediately after the crisis, and if a gradual reopening of the economy takes place, it may be that supply and demand rise slowly together. However, this work was the first to raise the fact that because supply and demand shocks are so different by sectors, even a gradual reopening may leave important supply-demand imbalances within industries. Such mismatches could consequently lead to an unusual level of heterogeneity in the inflation for different goods.

For policymakers there are three key implications from this study. First, the magnitude of the shocks being experienced by the U.S. economy is very large, with around a fifth of the economy not functioning. As Table 4 shows, our estimates of the potential impacts are a drop in employment of 23%, a decline in wages of 16%, and loss in value added of 20%. Bearing in mind the caveats about shocks vs. total impacts, the potential impacts are a multiple of what was experienced during the global financial crisis (e.g. where employment dropped 3.28 percentage points)[18] and comparable only to the Great Depression (e.g. where employment dropped 21.7% 1929-32 (Wallis 1989, Table 2)). Second, as the largest shocks are from the supply-side, strategies for returning people to work as quickly as possible without endangering public health must be a priority. Virus mitigation and containment are clearly essential first steps, but strategies such as widespread antibody testing to identify people who are safe to return to work, and rapid testing, tracing, and isolation to minimize future lock-downs, will also be vital until if and when a vaccine

[18]Employment Rate, aged 15-64, all persons for the US (FRED LREM64TTUSM156N) fell from 71.51 in December 2007 to 68.23 in June 2009, the employment peak to trough during the dates of recession as defined by the NBER.

is available. Furthermore, aggressive fiscal and monetary policies to minimize first-order shocks cascading into second-order shocks are essential, in particular policies to keep workers in employment and maintain incomes (e.g. the "paycheck protection" schemes announced by several countries), as well as policies to preserve business and financial solvency. Third, the inequalities highlighted by this study will also require policy responses. Again, higher income knowledge and service workers will likely see relatively little impact, while lower income workers will bear the brunt of the employment, income, and health impacts. In order to ensure that burdens from the crisis are shared as fairly as possible, assistance should be targeted at those most effected, while taxes to support such programs be drawn primarily from those least effected.

We have made a number of strong assumptions and used data from different sources. To recapitulate, we assume that the production function for an industry is linear and that it does not depend on the composition of occupations who are still able to work; we neglect absenteeism due to mortality and morbidity, as well as loss of productivity due to school closures (though we have argued these effects are small – see Appendix II.4.1). We have constructed our Remote Labour Index based on a subjective rating of work activities and we assumed that all work activities are equally important and they are additive. We have also applied a rating of essential industries for Italy to the U.S. Nonetheless, we believe that the analysis here provides a useful starting point for macroeconomic models attempting to measure the impact of the COVID-19 pandemic on the economy.

Part II

The multiplex nature of global financial contagions

Chapter 6

Literature review

6.1 Overview

In this chapter, we review the literature on multilayer networks and their applications to study financial contagion. In the Section 6.2, we focus on multilayer networks and introduce our notation. In Section 6.3, we focus on studies of financial contagion and introduce some terminology.

6.2 Multilayer and multiplex networks

Multilayer networks are an extension of ordinary (i.e. monolayer) networks that allow one to encode multiple types of edges, multiple subsystems, and other complications (Kivelä et al. 2014, Aleta & Moreno 2019, Porter 2018). A monolayer network $G = (\mathcal{V}, E)$ consists of a set $\mathcal{V}$ of nodes and a set $E \subseteq \mathcal{V} \times \mathcal{V}$ of edges that connect the nodes to each other. In contrast to monolayer networks, multilayer networks include "layers" (which we use to encode different types of edges), node-layer tuples (which we call "node-layers" as a shorthand) (i, α) that represent node i on layer α, and both intralayer and interlayer edges. Intralayer edges connect node-layer tuples within a layer, whereas interlayer edges connect node-layer tuples from different layers. In a directed multilayer network, one writes an edge as an ordered pair of node-layers. Following the convention of Kivelä et al. (2014), we use Latin letters to denote nodes and Greek letters to denote layers.

One can mathematically represent multilayer networks with the adjacency tensor $M_{j\beta}^{i\alpha}$, which components are given by

$$M_{j\beta}^{i\alpha} = \begin{cases} \omega_{j\beta}^{i\alpha}, & (i\alpha, j\beta) \in E, \\ 0, & (i\alpha, j\beta) \notin E, \end{cases} \tag{6.1}$$

where $\omega_{j\beta}^{i\alpha}$ is the weight of the edge $((i,\alpha),(j,\beta))$. In directed networks, $M_{j\beta}^{i\alpha}$ to correspond to the edge which starts in i,α and ends in j,β.

Multiplex networks are node-aligned multilayer networks, as interlayer edges can only connect node-layer tuples of the same node index (i.e. ones that correspond to the same entity). In a multiplex network, each layer represents a particular interaction type and can itself be viewed as a monolayer network. In the present study, we work with multiplex networks without interlayer edges, so we are using edge-coloured multigraphs. We encode the information of such a multiplex network by a set $\mathcal{A} = \left\{A^{[1]},\ldots,A^{[L]}\right\}$ of adjacency matrices, where L denotes the number of layers.

Several methods and algorithms have been designed to study the structure of multilayer networks (Kivelä et al. 2014). For example, community detection algorithms have been extended to multiplex and multilayer networks (Mucha et al. 2010, Jeub et al. 2017). Centrality diagnostic, which measure the relative importance of nodes (or edges or other structures) in a network (Newman 2018*b*), have also been extended to multilayer network.

6.2.1 Centrality measures

Several multilayer centrality measures have been developed for multilayer networks (Kivelä et al. 2014, Bravo & Óskarsdóttir 2020). We give a brief explanation of some of these measurements in this subsection.

Multilayer PageRank and Eigenvector centrality. In monoplex networks, several centrality measures can be computed by calculating an eigenvector of a matrix. For example, Eigenvector centrality and PageRank centrality. De Domenico et al. (2013) generalized Eigenvector and PageRank centrality via *eigentensors*. A rank-2 eigentensor $\Pi_{i\alpha}$ of rank-4 tensor $T_{j\beta}^{i\alpha}$ satisfies the following tensor equation

$$\sum_{i,\alpha} T_{j\beta}^{i\alpha}\Pi_{i\alpha} = \lambda\Pi_{j\beta}. \tag{6.2}$$

The multilayer eigenvector centrality of node layer (i,α) is given by (i,α) component of the eigentensor of the adjacency tensor $M_{j\beta}^{i\alpha}$ of a network (De Domenico et al. 2013). Similarly, the PageRank centrality of node layer (i,α) is given by the eigentensor of the teleporting random walk matrix

$$R_{j\beta}^{i\alpha} = \kappa T_{j\beta}^{i\alpha} + \frac{1-\kappa}{NL}u_{j\beta}^{i\alpha}, \tag{6.3}$$

where $T_{j\beta}^{i\alpha}$ is the transition probability tensor used in the benchmark random walk on networks. In other words, the transition probability tensor which models a random walker that chooses uniformly at random to which neighbor to jump to. The second part of the equation accounts for the teletransportation parameter.

One disadvantage of these centrality measures is that they require the user to specify interlayer edges. However, some multiplex networks lack interlayer edges, or the weights may be unknown. In these cases, *MultiRank* can be a useful centrality measure.

MultiRank. MultiRank is a centrality measure proposed by Rahmede et al. (2018) which assigns a MultiRank centrality to each node i as a whole and an MultiRank to each later α, rather than assigning a MultiRank to each node-layer (i, α). MultiRank assigns a large centrality to a node when central nodes point to it in layers with high influence (Rahmede et al. 2018). One can determine the the MultiRank (i.e. "influence") of a layer in one of two ways of a layer in one of two ways, depending on a parameter $\delta \in \{-1, 1\}$. If $\delta = 1$, layers tend to be more influential if very central nodes in it are active. By contrast, if $\delta = -1$, layers tend to be more influential if there are few nodes with nonzero in-degree and if the nodes that do have in-edges have large centralities.

To compute node i's MultiRank x_i and layer α's influence $z^{[\alpha]}$ one solves the following equation recursively until convergence

$$\begin{aligned} z^{[\alpha]} &= \frac{1}{N}(W^{[\alpha]})^{\phi}\left(\sum_{i=1}^{n} B_{\alpha i}^{\text{in}}(x_i)^{\delta\gamma}\right)^{\delta}, \\ x_i &= \kappa \sum_{j=1}^{n} \frac{G_{ji}}{k_j} x_j + \chi v_i \,, \end{aligned} \tag{6.4}$$

where $W^{[\alpha]} = \sum_{i,j} A_{ij}^{[\alpha]}$, the adjacency matrix G is $G = \sum_{\alpha=1}^{L} z^{[\alpha]} A_{ij}^{[\alpha]}$, and the normalized in-strength $B_{\alpha i}^{\text{in}}$ of node i in layer α is

$$B_{\alpha i}^{\text{in}} = \frac{\sum_j A_{ji}^{[\alpha]}}{W^{[\alpha]}} \,. \tag{6.5}$$

Furthermore,

$$
\begin{aligned}
k_j &= \max\left\{1, \sum_i G_{ji}\right\} , \\
v_j &= \theta\left(\sum_j (G_{ij} + G_{ji})\right) , \\
\chi &= \frac{1}{\sum_i v_i} \sum_j \left(\left[1 - \kappa\theta\left(\sum_i G_{ji}\right)\right] x_j\right) ,
\end{aligned} \tag{6.6}
$$

where θ is the Heaviside function.

The parameter $\gamma > 0$ influences the contribution of small-centrality nodes by either enhancing ($\gamma > 1$) or suppressing ($\gamma < 1$) it, and $\phi \in \{0, 1\}$ determines whether one considers (if $\phi = 1$) or does not consider (if $\phi = 0$) the total edge weight in layer α (Rahmede et al. 2018). In the same way than in the PageRank algorithm, κ denotes the probability that a random walker moves to an adjacent node and $1 - \kappa$ is the teleportation probability, but now we use the adjacency matrix G_{ij}.

6.3 Financial contagion on networks

Many researchers have used network analysis to analyse the interconnected nature of financial systems (Markose 2012, Battiston et al. 2016, Anand et al. 2013, Wiersema et al. 2019), and several studies have used a network-based approach to estimate the systemic importance of financial entities. In this section we review some of these studies.

6.3.1 Financial monoplex networks

Multiple methods have been proposed to identify important nodes in monoplex networks. Perhaps most prominently, a variety of studies have developed centrality measures that are tailored to examining the importance of financial entities. For example, "impact centrality" recursively takes into account that a node is more systemically important if it impacts many systemically important nodes (Kaushik & Battiston 2013); "default cascade impact" (Battiston, Gatti, Gallegati, Greenwald & Stiglitz 2012) estimates the impact of a defaulting node; and "DebtRank" (Battiston, Puliga, Kaushik, Tasca & Caldarelli 2012) attempts to examine how "financial distress" propagates through a network, where financial distress refers to the inability or difficulty of a financial institution (or set of institutions) to pay its debt (their

debts, in the case of multiple institutions). Recent studies have used DebtRank to analyse financial networks (Li et al. 2018, Silva et al. 2018, Diem et al. 2019, Pichler et al. 2018, Poledna et al. 2018) and other studies have focused on developing different versions of DebtRank (e.g. by considering nonlinear forms of distress propagation) (Bardoscia et al. 2015, 2016). Below we explain the original DebtRank in more detail.

DebtRank. DebtRank, a centrality measure whose name is motivated by PageRank Gleich (2015), is an algorithm for monolayer networks that estimates the economic loss that a node (or a set of nodes) would cause if it becomes financially distressed (or they become distressed, in the case of multiple nodes) (Bardoscia et al. 2015).

DebtRank supposes that a network's adjacency-matrix elements A_{ij} encode the monetary amount that an institution i has invested in institution j, and it considers a capital buffer vector $\mathbf{E}$ whose entries encode the capital buffer of each node i.

Given an adjacency matrix A and an equity vector $\mathbf{E}$, one defines a relative-economic-value vector $\mathbf{v}$ with components

$$v_j = \frac{\sum_j A_{ij}}{\sum_{l,j} A_{lj}} \tag{6.7}$$

and an impact matrix M with components

$$M_{ij} = \min\left\{1, \psi \frac{A_{ij}}{E_j}\right\} , \tag{6.8}$$

where $\psi \in [0,1]$ is an impact scaling factor. The impact matrix, as it name suggests, determines the impact that node j has on node i if the former becomes distressed.

The functional form of Eq. (6.8) incorporates the assumption that the maximum impact is 1 and that the impact increases linearly with the assets that i has invested in j (until one reaches the maximum impact of 1).

DebtRank follows an iterative procedure to compute two state variables, which we denote by $h_i(t)$ and $r_i(t)$. The variable $h_i(t) \in [0,1]$ indicates the financial distress of node i at time t, and the variable $r_i(t) \in \{U, D, I\}$ indicates whether node i node is financially undistressed, distressed, or inactive.

To compute the DebtRank of the nodes in a set $\mathcal{U} \subseteq \mathcal{V}$, one initialiaes the state variables by letting $h_i(0) = 0$ for all $i \notin \mathcal{U}$ and $h_i(0) = 1$ for all $i \in \mathcal{U}$. Similarly, one initialises $r_i(0) = U$ for all $i \notin \mathcal{U}$ and $r_i(0) = D$ for all $i \in \mathcal{U}$. To determine a node's state at a discrete time $t \geq 1$, one then performs the following iteration:

$$h_i(t) = \min\left\{1,\ h_i(t-1) + \sum_{j \in D_f(t-1)} M_{ij}\right\} , \tag{6.9}$$

where $D_f(t-1)$ is the set of distressed nodes at the previous time (i.e. with $r_i(t-1) = D$). One then updates the state variable $r_i(t)$ using the following rule:

$$r_i(t) = \begin{cases} D\,, & \text{if} \quad h_i(t) > 0 \text{ and } r_i(t-1) \neq I \\ I\,, & \text{if} \quad r_i(t-1) = D \\ r_i(t-1)\,, & \text{otherwise}\,. \end{cases} \tag{6.10}$$

One iterates the dynamics until some steady-state time τ at which all nodes are either undistressed or inactive. Finally,

$$\mathrm{DR}(\mathcal{U}) = \sum_j h_j(\tau) v_j - \sum_j h_j(0) v_j\,. \tag{6.11}$$

gives the DebtRank of the nodes in the set $\mathcal{U} \in \mathcal{V}$.

Threshold models. Other studies of financial networks have used threshold models to study interconnected financial systems (Gai & Kapadia 2010, Battiston, Gatti, Gallegati, Greenwald & Stiglitz 2012). Threshold models are a popular type of spreading process (Porter & Gleeson 2016) that can exhibit cascading behaviour and have played an important role in the study of financial contagions (Caccioli et al. 2018). In a threshold model, each node has a one of several discrete states (in practice, two states is the most common choice), and node states update according to some rule, such as a deterministic one that is based on the states of a node's neighbors (Gleeson 2013). Typically, when modelling a financial contagion with a threshold model, nodes represent financial entities, directed edges encode lending–borrowing relationships, and the states of the nodes correspond to a solvent or defaulted status (Caccioli et al. 2018). Nodes update their status according to a threshold rule: when a node loses more assets than it has equity, the node changes status from solvent to defaulted. One reason that threshold models have been adopted widely for modelling financial contagions is that, in large networks (specifically, when the number N of nodes becomes infinite), a vanishingly small fraction of initially defaulted financial institutions can cause the default of a significant fraction (i.e. one that is nonzero in the $N \to \infty$ limit) of all nodes in a network (Gai & Kapadia 2010).[1] This idea is useful for modelling "systemic risk", which describes the risk that a significant fraction of a financial system defaults as a consequence of the failure of one (or a small group) of financial institutions (Poledna et al. 2015).

[1] See Porter & Gleeson (2016) for a brief discussion of various approaches for defining the notion of a "cascade".

Other methods. Notions of systemic risk have also been developed in models of interconnected financial institutions that take the form of coupled stochastic differential equations (Eisenberg & Noe 2019, Feinstein & Sojmark 2019). Other insightful approaches for studying financial contagion networks include using branching processes to examine financial contagions from common asset holdings (i.e. overlapping portfolios) (Caccioli et al. 2014) and drawing from ideas from network ecology and disease spread to explore complexity and stability in financial networks (Haldane & May 2011). In the present thesis, we focus on centrality analysis and the dynamics of threshold models.

6.3.2 Financial multilayer networks

Studies like the ones above have provided interesting insights into the dynamics of financial contagions (Caccioli et al. 2018), but most of these investigations have examined monolayer networks, such as by considering lending and borrowing of a single type of asset or an aggregate of several types of assets. Multilayer networks can model interconnected structures more effectively than monolayer networks (Kivelä et al. 2014, Aleta & Moreno 2019), and they can also have significant effects on dynamical processes on them (de Domenico et al. 2016), so it is important to generalize threshold models and network measures of systemic importance (such as DebtRank) to multilayer networks. Multilayer networks include layers that can encode different subsystems and/or different types of interactions between nodes. One particular type of multilayer network is a multiplex network, in which interlayer edges can occur only between node-layers (i.e. an instantiation of a node on a specific layer) that correspond to the same entity (Kivelä et al. 2014, Bianconi 2018). Several studies have used multiplex networks to study financial contagions and economic systems (Aoki et al. 2010, Poledna et al. 2015, Brummitt & Kobayashi 2015, Bargigli et al. 2015, Greenwood et al. 2015, Starnini et al. 2019). Existing research on multiplex financial networks has studied microlevel and country-specific data (e.g. analysing a network of banks at a national level) (Diem et al. 2019, Poledna et al. 2015), examined DebtRank for different layers of a multiplex network (Poledna et al. 2015), and applied threshold models but with a focus on unweighted and overly simplistic synthetic networks (e.g. by using layers that consist of Erdős–Rényi networks) (Brummitt et al. 2012, Brummitt & Kobayashi 2015).

Most studies of threshold models on financial networks have focused on monolayer networks (Gai & Kapadia 2010, Battiston, Gatti, Gallegati, Greenwald & Stiglitz 2012), although there have been some extensions of threshold models to multiplex

networks. For example, Brummitt et al. (2012) examined threshold dynamics on unweighted multiplex networks in which a node activates if, in any layer, the fraction of active neighbors exceeds a given threshold. Lee et al. (2014) extended this model by including some nodes that activate only if the fraction of active neighbors in each layer is larger than a specified threshold. Yağan & Gligor (2012) examined a multiplex threshold model with a single update rule that incorporates all edge types. In applications to financial networks, Brummitt & Kobayashi (2015) considered a multiplex threshold model with two layers in which different layers correspond to different debt seniorities (i.e. to the chronological order of debt repayments when there are sales or defaults).

These studies have illustrated that multilayer financial networks can be more vulnerable than monolayer networks to global cascades in some dynamical processes. Additionally, using empirical data, Poledna et al. (2015) and Leon et al. (2014) reported that systemic risk in the Mexican and Columbian banking systems, respectively, are underestimated significantly when one examines monolayer networks that encode only one type of financial asset, rather than a multiplex network that encodes multiple types of financial assets.

Chapter 7

Multiplex financial contagion

7.1 Introduction

The 2008 global financial crisis illustrated interconnectedness in the financial stability of countries. In September 2008, Lehman Brothers Holdings Inc., one of the largest investment banks in the United States, went bankrupt. By October 2008, HBOS plc and Lloyds Banking Group were bailed out in the United Kingdom (Fernando et al. 2012, Tebogo 2012). Financial distress spread to other countries, and the crisis became global (De Haas & Van Horen 2012). Financial interconnectedness continues to be pervasive. Since the middle of the 1990s, there has been a rapid expansion of cross-border financial positions (i.e. in the balances of the recorded assets, liabilities, and equity), with the stocks of foreign assets and foreign liabilities increasing from about 75–77 % of the world's gross domestic product (GDP) in 1995 to about 172–189 % of the world's GDP by 2016 (Lane & Milesi-Ferretti 2017).

This significant interconnectedness and the 2008 global financial crisis both highlight the importance of identifying "systemically important" financial institutions and countries (i.e. financial institutions and countries whose financial distress can trigger distress in other institutions). In the present chapter, we conduct a multiplex network analysis to help improve understanding of financial contagions and identify systemically important countries.

We use a multilayer network analysis to examine how financial distress propagates between countries. In particular, we simulate financial shocks and their propagation through a network. We start with a single financially distressed country and examine the ensuing distress of other countries. Our primary aim is to examine the so-called "global financial system" — an interconnected network of lending and borrowing of financial assets between countries — and perform multilayer network analysis to

identify systemically important countries. We study multiplex networks, so interlayer edges can occur only between node-layers that correspond to the same country. To examine important countries, we employ the MultiRank centrality measure (Rahmede et al. 2018), and we extend multilayer threshold models (Brummitt et al. 2012, Brummitt & Kobayashi 2015, Korniyenko et al. 2018)) for the propagation of financial contagions by incorporating a mechanism for shock propagation between different layers.

We go further than previous research by generalizing studies of multiplex threshold models that examined almost immediate contagion spreading between layers (Korniyenko et al. 2018, Brummitt & Kobayashi 2015). In our generalization, we include a threshold rule for the interlayer spreading of a contagion. When this threshold is 0, our model reduces to a situation with almost immediate spreading of contagions between layers. Another novelty of our model is that we incorporate a threshold rule for intralayer contagions that depends on the "capital adequacy" of a country's financial system; this gives a rough measure of a country's ability to withstand shocks. Capital buffers are important, as recent work (Farmer et al. 2020) has illustrated that countries with adequate capital buffers can use them to absorb shocks and reduce systemic risk during financial distress.

By considering different layers and allowing spreading between them, we are able to examine financial contagions that take the heterogeneity of financial exposures into consideration. Specifically, we study contagion dynamics on a multiplex network whose layers (which each represent one type of asset), encode cross-country assets in equity, debt, and banks' loans and deposits.

This chapter is structured as follows. We start by explaining how we build a multiplex network in which nodes are countries, edges are links between country assets, and layers correspond to different asset types. Similar to Korniyenko et al. (2018), we build multiplex and monolayer networks using data from the International Monetary Fund (IMF) and the Bank of International Settlements (BIS).[1] We also explain MultiRank (Rahmede et al. 2018), an algorithm for identifying important nodes in a multilayer network, that we use for an initial exploratory analysis and the subsequent multilayer threshold model that we develop. We then discuss our results

[1]Our multiplex network has three layers. We use a single equity layer, which encodes both cross-border direct and cross-border portfolio investments in equity. Similarly, we construct the debt layer from both cross-border direct and cross-border portfolio investments in debt assets. In Korniyenko et al. (2018), our multiplex network had five layers, as the equity layer and debt layers were each split into two layers.

from the MultiRank algorithm and our threshold model, and we complete this chapter with a short concluding discussion.

7.1.1 A global financial multiplex network

We construct a multiplex network of financial assets between countries. In this "global financial multiplex network", each node represents a country and each layer represents an asset type. An edge from node i to node j in layer α indicates that country i has an asset of type α in country j. The edges are directed, and we weight them by their asset value (in United States dollars [USD]). To construct a global financial multiplex network (one for each year from 2008 to 2018), we use data from the IMF and BIS. Specifically, we use annual data from the IMF Coordinated Direct Investments Survey (CDIS) (International Monetary Fund 2018*a*) and the IMF Coordinated Portfolio Investment Survey (CPIS) (International Monetary Fund 2018*b*) on countries' cross-border equity and debt holdings. We use quarterly data from the BIS International Banking Statistics on countries' cross-border bank loans and deposits (Bank for International Settlements 2018).

Using these data sets, for each year, we construct a directed and weighted multiplex network of 3 layers and 131 countries. The layers correspond to different types of financial assets: (i) cross-border portfolio and direct investments in equity; (ii) cross-border portfolio and direct investments in debt assets; and (iii) cross-border bank loans and deposits. Our use of different types of financial assets is motivated by empirical research. Prior research has established that different types of financial assets have different volatility, so the loss of some types of asset can cause more financial distress than the loss of other types (Becker et al. 2009, Claessens et al. 1995). A country's response to financial distress depends on the type of financial assets that are impacted. For example, in response to distress in banking assets, it may be sufficient to direct a policy action exclusively at banks. However, one would respond differently to distress in other asset types. For example, such a response may include a fiscal package that compensates for distress by providing government transfers, credit, or tax reliefs to support the economy.

We also use data from the IMF's Financial Soundness Indicators (FSIs) database (International Monetary Fund 2018*c*) to obtain the capital adequacy ratio (CAR) of the financial system. The CARs for some countries are missing in the FSIs data, so we retain only countries where this data is available for our network analysis. This yields the 131 countries in our networks. See Fig. 7.1 for a schematic representation

of a global financial multiplex network. We encode the information in a multiplex network with layer adjacency matrices

$$A_{ij}^{[\alpha]} = \begin{cases} \omega_{ij}\,, & \text{if country } i \text{ has an asset of type } j \\ 0\,, & \text{otherwise}\,, \end{cases} \tag{7.1}$$

where ω_{ij} is the value in USD of the assets that country i has invested in country j. These multiplex networks do not have interlayer edges, so they are edge-coloured multigraphs.

Given a global financial multiplex network, one can compute important financial quantities. For example, the amount of assets of type α that country i holds in other countries is

$$a_{i\alpha} = \sum_j A_{ij}^{[\alpha]}\,, \tag{7.2}$$

and the total amount of external assets that country i holds is

$$a_i = \sum_\alpha a_{i\alpha}\,. \tag{7.3}$$

In addition to global financial multiplex networks, we construct associated monolayer networks by aggregating the layers of a global financial multiplex network to obtain an adjacency matrix as the sum of the adjacency matrices of the three layers. In other words, we aggregate all assets equally without distinction. This yields a single adjacency matrix A with elements A_{ij}, which encode the assets in USD that country i has invested in country j for each i and j. We refer to this network as the "global financial aggregate network". We also examine each layer as a separate monolayer network. We refer to these monolayer networks from layers 1, 2, and 3 as the "equity", "debt", and "banking" networks, respectively.

7.2 Identifying systemically important countries

In this section, we discuss our methods for identifying systemic important countries, which can cause a significant fraction of the countries to become financially distressed. As an initial approach, we use MultiRank (Rahmede et al. 2018), a centrality measure for multiplex networks. We then study a multiplex threshold model, which allows a contagion to spread between different layers.

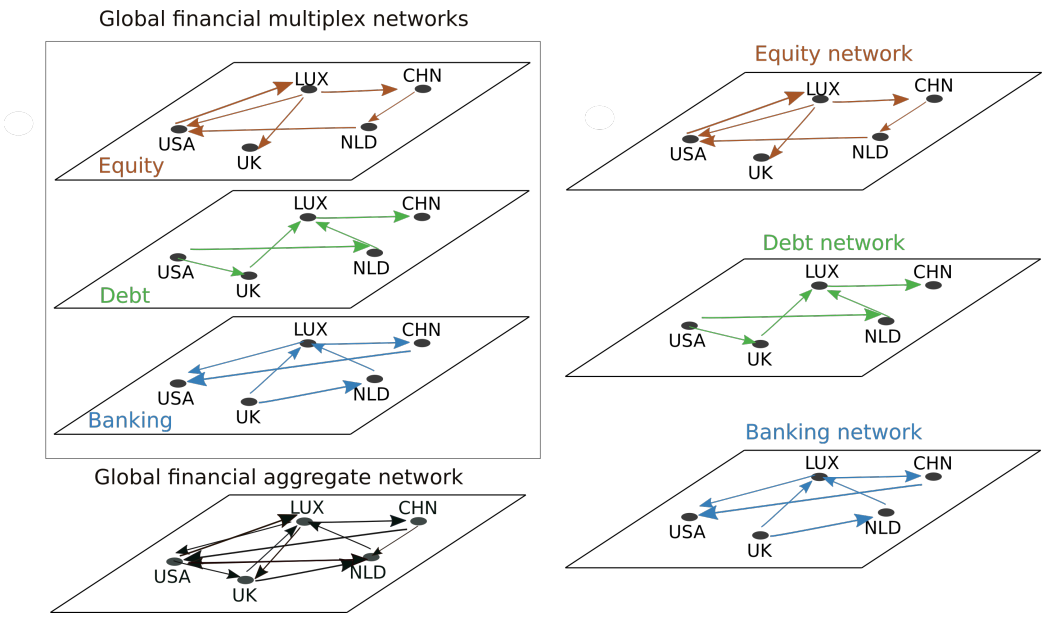


Figure 7.1: **Global financial networks**. Schematics of the networks that we construct using IMF data. In all networks, nodes represent countries and edges encode cross-country financial assets. The multiplex network has three layers, which correspond to three types of assets: equity (layer 1), debt (layer 2), and banking loans and deposits (layer 3). The multiplex network has no interlayer edges, so it is an edge-coloured multigraph.

7.2.1 PageRank and MultiRank

Centrality diagnostics measure the relative importances of nodes (or edges or other structures) in a network (Newman 2018*b*), and it is thus natural to use centrality measures to try to identify systemically important nodes.

To identify important countries in a monolayer network, we calculate PageRank centrality, a measure of a node's importance based on a random walk with teleportation (Newman 2018*b*, Gleich 2015). In our calculations, we use the teleportation-parameter value $1-\kappa = 0.15$, where κ is the probability that a random walker moves to an adjacent node and $1-\kappa = 0.15$ is the probability that a random walker teleports to any node (which we choose uniformly at random) in the network (Gleich 2015). The choice of $1-\kappa = 0.15$ is the most common choice in PageRank, although it is an arbitrary value.

To examine important countries and important asset types in our global financial multiplex networks, we calculate MultiRank (Rahmede et al. 2018). We use MultiRank both because it is easy to implement and because it allows us to rank both nodes and layers. In Chapter 6 we introduced the equations to compute MultiRank. Here we simplify these equations for multiplex networks where each node has an out-degree of at least 1 in at least one layer.

When all nodes have an out-degree of at least 1 in at least one layer, the system of equations 6.6 can be simplified as follows

$$\chi = \frac{1}{n}\sum_j (1-\kappa)x_j$$
$$v_j = 1\,,$$

and

$$k_j = \sum_i G_{ji}\,, \tag{7.4}$$

is the out-degree of node j in G. Since $\sum_j x_j = 1$, we obtain that

$$z^{[\alpha]} = \frac{1}{N}(W^{[\alpha]})^{\phi}\left(\sum_{i=1}^{n} B^{in}_{\alpha i}(x_i)^{\delta\gamma}\right)^{\delta},$$
$$x_i = \kappa\sum_{j=1}^{n}\frac{G_{ji}}{k_j}x_j + \frac{1-\kappa}{n}. \tag{7.5}$$

Together Eqs. (7.4) and (7.5) are the simplified version of systems of equations 6.4 and 6.6 in Chapter 6 that define the MultiRank algorithm.

In this chapter, suppose that edge weights are important (because they encode money), so we use $\phi = 1$. For simplicity, we choose $\gamma = 1$, which implies that nodes contribute to layer influence (i.e. layer MultiRank) in a linear fashion. We use $\delta = 1$, because we consider a type of asset to be more important when important countries use them. As in the PageRank algorithm, we use $1 - \kappa = 0.15$.

MultiRank has been used to study various multiplex networks (Bonaccorsi et al. 2019, van Lidth de Jeude 2019), and it is useful for exploration of important countries in a multiplex financial network. However, to properly study a global financial contagion, it is important to use tools that are designed specifically for that topic.

7.2.2 Systemic importance of nodes

As discussed in Chapter 6, DebtRank was specifcally designed to study financial contagion in monoplex networks. To study financial contagion in multiplex networks it may seem natural to generalize DebtRank. Unfortunately, DebtRank has an issue that makes it inadequate for our problem.[2] As we explain below, DebtRank is not additive, in the sense that the DebtRank of the set $\{i, j\}$ of two nodes (i and j) can be smaller than the DebtRank of the individual nodes i or j.

The DebtRank algorithm assumes that a node, once distressed, becomes inactive (which implies that it does not lose more assets and that it has no further effect on other nodes) in the next time step. This implies that, once a node becomes inactive from a particular neighbor being distressed, it becomes immune to any financial distress that could have been caused by a different neighboring node later becoming distressed. For example, even if all neighbors of node i are inactive by some time τ_0 (because they were distressed in some previous time step), the estimate of the financial losses of node i may consider only a subset of node i's neighbors.

This feature makes it possible for DR with one initially distressed node to be larger than DR with two or more initially distressed nodes. In particular, it is possible that $\mathrm{DR}(\{i\}) > \mathrm{DR}(\{i, j\})$. For example, consider the weighted and directed network in Fig. 7.2, where f and g are the weights of the edges. With $f = 1$ and $g = 5$, because nodes become inactive one time step after being distressed, we obtain

$$\mathrm{DR}(\{1\}) \approx 0.72 > \mathrm{DR}(\{1, 5\}) \approx 0.61\,. \tag{7.6}$$

[2] Additionally, in the context of DebtRank, shocks tend to propagate only a small distance in densely connected networks. Therefore, DebtRank also has other features that one may wish to modify when studying a problem like ours.

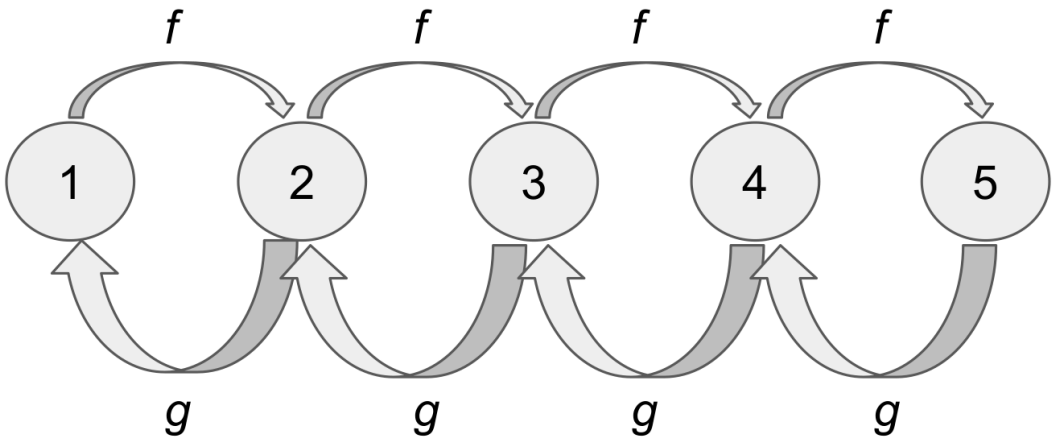


Figure 7.2: **Example network for illustrating DebtRank.** A weighted and directed network with five nodes. The quantities f and g denote the weights of the edges. We use arrow widths to highlight our assumption that $g > f$.

This lack of additivity makes it difficult to generalize DebtRank to multiplex networks. In particular, a straightforward extension of DebtRank to multiplex networks can yield a higher DebtRank for node-layer (i, α) than for the set $\{(i, \alpha)\}_{\alpha=1}^{L}$ of node-layers. In such a generalization, the financial distress of a country in one layer causes a larger monetary loss than the loss from a financial distress of that country in all layers. From an economic perspective, this seems unreasonable.

Given the above issue, we propose a different approach: we develop a threshold model of financial contagions for multiplex networks that can be informative about the systemic importance of different countries in the global financial network. We model the propagation of a financial contagion and identify the countries that cause a sufficiently large fraction of the system to become distressed. In particular, we examine the fraction of countries whose sum of assets is larger than 90% of the total value of all assets in a network.

A multiplex threshold model of financial contagions. We develop a discrete-time, deterministic system of threshold dynamics; and we simulate it on multiplex networks that are weighted and directed. Like the nodes in the Watts threshold model (WTM) of a social contagion (Watts 2002) and the Gai–Kapadia model of a financial contagion (Gai & Kapadia 2010), node-layers in our networks have binary states:

node-layers (i.e. country-assets) can be either undistressed or distressed.[3] The state of node-layer (i, α) at time t is

$$s_{i\alpha}(t) = \begin{cases} 0\,, & \text{if } (i,\alpha) \text{ is undistressed at time } t, \\ 1\,, & \text{if } (i,\alpha) \text{ is distressed at time } t\,. \end{cases} \tag{7.7}$$

Similar to the Gai–Kapadia model (Gai & Kapadia 2010), where nodes have capital buffers, we suppose that each country has some assets that it holds as capital buffers in each layer. The purpose of these buffers is to absorb some amount of loss and avoid distress. Inspired by how national regulators track the CARs of banks according to their risky assets, we suppose that the amount of assets that are held by country i as capital buffers in layer α is proportional to the assets $a_{i\alpha}$ that it holds in that layer.

We model the spread of a contagion as follows. If a country loses more assets in a layer than its layer buffer, the country becomes distressed and loses all assets. Additionally, if the sum of lost assets of a country over all layers exceeds a fraction c_i of its total assets a_i, then the country becomes distressed in all layers. This last contagion mechanism is based on risk propagation within countries: when investors perceive that the risk of financial distress of a country increases, the value of that country's assets decreases.

With these ideas in mind, our spreading rule for a financial contagion is

$$s_{i\alpha}(t+1) = \begin{cases} 1\,, & \text{if } \frac{\text{loss}_{i\alpha}(t)}{a_{i\alpha}} > b_{i\alpha}, \\ 1\,, & \text{if } \frac{\text{loss}_{i}(t)}{a_{i}} > c_{i}, \\ 0\,, & \text{otherwise}\,, \end{cases} \tag{7.8}$$

where $s_{i\alpha}(t)$ denotes the state of country i in layer α at time t, the quantity $b_{i\alpha}$ is the fraction of assets that are held by country i in layer α as cash reserves,

$$\text{loss}_{i\alpha}(t) = \sum_j A_{ij}^{[\alpha]} s_{j\alpha}(t) \tag{7.9}$$

is the monetary amount of lost assets by the end of time t (i.e. including both at and before time t) for country i in layer α, and

$$\text{loss}_{i}(t) = \sum_\beta \text{loss}_{i\beta}(t) \tag{7.10}$$

[3] Several studies of financial networks of banks and firms that used threshold models have employed the terms "solvent" and "defaulted" instead of "undistressed" and "distressed" (Lee et al. 2014, Brummitt & Kobayashi 2015, Gai & Kapadia 2010). However, because we are studying financial networks of countries, rather than banks or firms, we prefer to use the terms "undistressed" and "distressed".

is the monetary amount of lost assets by the end of time t of country i.

We refer to each $b_{i\alpha}$ (which encode capital buffers) as a "horizontal threshold" and each c_i (which encode intra-country risk-propagation buffers) as a "vertical threshold". To the best of our knowledge, our incorporation of vertical thresholds is a novel feature of our multiplex threshold model.

We expect that the idea of having both vertical and horizontal thresholds will also be useful for studying other types of contagions, such as the spread of information and ideas in multichannel communication networks.

For our monolayer financial networks, we run a similar contagion model, but we now neglect the interlayer (i.e. intracountry) contagions. Specifically, we use the update rule

$$s_i(t+1) = \begin{cases} 1\,, & \text{if } \frac{\text{loss}_i(t)}{a_i} > b_i, \\ 0\,, & \text{otherwise}\,, \end{cases} \tag{7.11}$$

where $s_i(t)$ denotes the state of country i at time t, the quantity a_i denotes the assets of country i, and b_i is a threshold that represents the capital buffer of country i. The total value of the financial assets that are lost by country i in a monolayer network is

$$\text{loss}_i(t) = \sum_j A_{ij} s_i(t)\,. \tag{7.12}$$

Initial conditions and parameters of the models. To calibrate the horizontal thresholds (i.e. capital buffers), we use data on CAR, which measures the ratio of a bank's capital to its risk-weighted assets[4].

The Basel Committee on Banking Supervision (BCBS) recommends a minimum amount of CARs that financial institutions must maintain to have sufficient buffers to provide a cushion against losses due to financial distress (Basel Committee on Banking Supervision 2019).

Although CARs are only for banks, we use them as a threshold in all layers as a proxy for the heterogeneity in the robustness of different countries towards shocks. In other words, we assume that $b_{i\alpha} = b_i = \text{CAR}_i$ for all α. For the vertical thresholds, we suppose that the countries are homogeneous (i.e. that $c_i = c$ for all i. As a reference value, we take $c_i = 1/L$, where L is the number of layers in each global financial multiplex network.

[4]Regulators calculate risk-weighted assets by considering a bank's assets, evaluating the risk levels of those assets, and then assigning a weight. The risk-weighted asset value of an asset is the value of the asset multiplied by a weight that is assigned to it by a regulator. When there are several assets, one calculates a weighted sum of those assets.

In Section 7.3, we also examine our threshold model using different values of the vertical threshold c.

We run our threshold model on the global financial multiplex networks and on the monolayer networks (i.e. the aggregated, equity, debt, and banking networks). All of these networks are weighted and directed. We initialise our model with a single country i^* that is distressed in all layers, and we measure the subsequent number of countries that become distressed. Our model has a deterministic update rule, so each individual simulation has a deterministic outcome. Once a node-layer is distressed (and, in a monolayer network, once a node is distressed), it remains distressed for the rest of a simulation. A simulation ends when no node-layers change their state anymore, and we use τ to denote the time step at which a simulation ends. We consider a country to be distressed when it is distressed in all layers. The extent $R(i^*)$ of a contagion from a seed country i^* is the percent of assets that are collectively held by the distressed countries at the end of the simulation. That is,

$$R(i^*) = 100 \times \frac{\sum_i a_i \prod_\alpha s_{i\alpha}(\tau)}{\sum_i a_i} \,. \tag{7.13}$$

We say that country i^* is *systemically important* when $R(i^*) > 90\,\%$. Analogously, for the monolayer networks, the extent of a contagion is

$$R(i^*) = 100 \times \frac{\sum_i a_i s_i(\tau)}{\sum_i a_i} \,. \tag{7.14}$$

7.3 Results

We examine the temporal evolution of the global financial multiplex networks in two ways. We first use MultiRank to examine how layers and countries change in rank over time. We compare the MultiRank values of countries in the multiplex global financial network with the PageRank values of countries in the aggregate global financial networks. We then study global contagions using our threshold model and examine the number of systemically important countries. We do this for the multiplex global financial network and for the equity, debt, banking, and aggregate global financial networks.

In Fig. 7.3, we show a visualization of the global financial multiplex network for 2018. We label and highlight the countries on which we focus our discussion. Plotting this multiplex network illustrates the heterogeneity across layers. For example, the degree of the United Kingdom is considerably larger in the banking layer than in

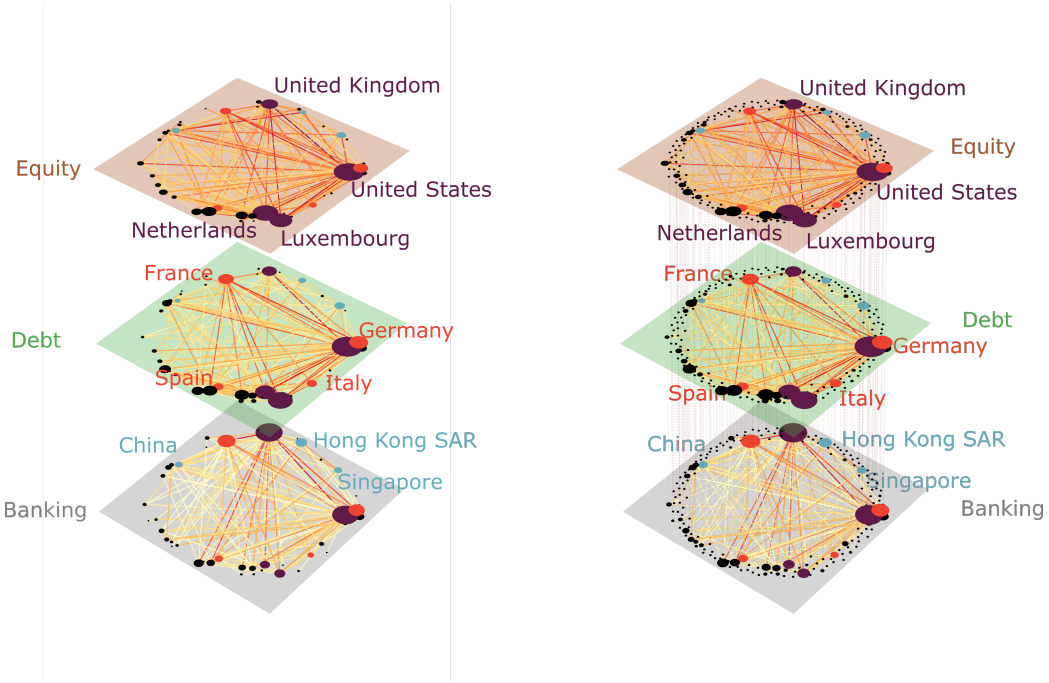


Figure 7.3: **The global financial multiplex network for 2018.** The size of a node is proportional to its degree. For visualization purposes, we show only edges whose weight is ten or more times larger than the mean weight of the edges in the layer. We colour the edges based on the natural logarithm of their weight; darker colours indicate larger weights. In the *left* panel, we show the edge-coloured multigraph on which we run MultiRank. In the *right* panel, we also include interlayer edges that encode the vertical contagion mechanism in our threshold model. All layers have the same layout; in other words, the position of a country is the same in each layer. We label the countries on which we focus our discussion. To avoid overcrowding the figure with labels and because countries have the same position in all layers, we only label each country in one layer. [We produced the network plots using PYMNET (Kivelä 2014) and additional labels added with INKSPACE.]

the equity layer. By contrast, the opposite is the case for The Netherlands and Luxembourg.

7.3.1 MultiRank evolution

Our centrality analysis yields different results on the multiplex financial network than it does for any of the four monolayer networks (the equity, debt, banking, and aggregate networks). For brevity, we focus on discussing the differences between the multiplex global financial network and the aggregated monolayer network. Computing MultiRank and PageRank for the multiplex and aggregate networks, respectively, produces results that correlate strongly with each other, with a Kendall rank corre-

lation coefficient of above 0.90 for each year. Table 7.1 shows there is also a slight downward trend in the correlation over the years. Both centrality measures illustrate that (i) the United States, the United Kingdom, Luxembourg, and the Netherlands are top ranked for most years; (ii) several major European countries decrease in centrality over time; and (iii) several major Asian countries increase in centrality over time. Notably, MultiRank better highlights the emerging role of Asian countries than PageRank does for the aggregated network. In particular, according to MultiRank, Hong Kong SAR, China, and Singapore rank higher than Italy and Spain by 2018 (see Fig. 7.4).

Year	Kendall coefficient	p-value
2008	0.97	1.66×10^{-61}
2009	0.96	3.52×10^{-60}
2010	0.96	4.32×10^{-60}
2011	0.96	2.00×10^{-60}
2012	0.95	9.69×10^{-59}
2013	0.94	1.70×10^{-58}
2014	0.94	7.71×10^{-58}
2015	0.93	1.19×10^{-56}
2016	0.94	1.44×10^{-57}
2017	0.93	3.21×10^{-56}
2018	0.93	2.27×10^{-56}

Table 7.1: **MultiRank versus PageRank for the global financial networks for each year**. We show the values of the Kendall rank correlation coefficient and the p-value between MultiRank and PageRank for each year from 2008 through 2018.

To better understand the emerging role of Asian countries as systemically important countries we examine changes over time in the different layers of the multiplex global financial network. Since the 2008 global financial crisis, which started in the United States banking sector, many countries have made an effort to deleverage banking assets, especially in Europe[5]. Another trend that we observe (see the right panel of Fig. 7.5) is that the fraction of assets in the equity layer grows over time, whereas those in the other two layers decrease over time (excluding the year 2008).[6] Our calculations of MultiRank also illustrate an increasing centrality of the equity layer.

[5]The persistent aggregate decline in cross-border banking claims of European banks was discussed in Lund et al. (2013), Forbes (2014), Forbes et al. (2017), Lane & Milesi-Ferretti (2017), Mallaby (2016).

[6]Some countries only started systematically reporting CDIS data to the IMF from 2009 onwards. This explains the missing equity data for some countries in 2008. The issue of "missing" data is visible in Fig. 7.5, where we observe a very small value of total equity assets in 2008. Consequently, we exclude 2008 from our discussion of the importance of the layers.

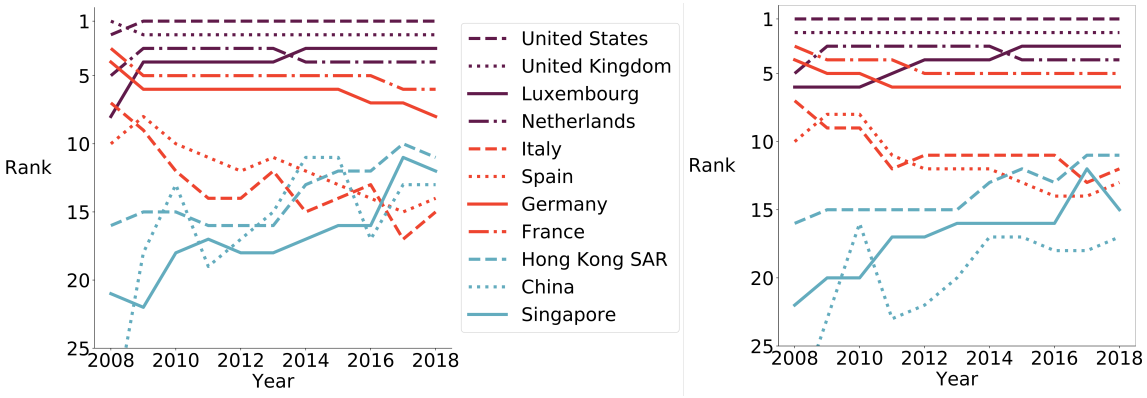


Figure 7.4: **Results of computing MultiRank on the multiplex global financial network versus computing PageRank on the aggregated financial network**. In the *left* panel, the temporal evolution of MultiRank, and in the *right*, PageRank for several countries. We show the dominant financial countries in purple, mainland European countries in orange, and Asian countries in blue.

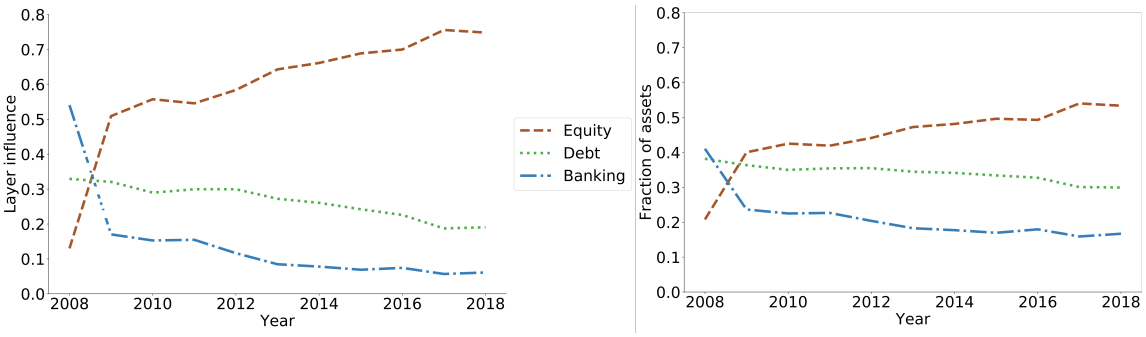


Figure 7.5: **Layer influence (i.e. MultiRank) $z^{[\alpha]}$ and assets in layers.** In the *left* panel, the temporal evolution of the layers with respect to the influence of layers. In the *right* panel, with respect to the fraction of assets in each layer.

(See the left panel of Fig. 7.5.) As the banking layer decreases in importance, some Asian countries (which are more central in the equity and debt layers) increase in centrality. We interpret the MultiRank and the fraction of assets of a layer α as measures of the relative importance of a layer with respect to other layers. In this context, we find that MultiRank amplifies the growth of the relative importance of the equity layer over time: although the fraction of assets in the equity layer increases only from about 0.4 in 2009 to about 0.53 in 2019, the influence of the equity layer increases from about 0.51 in 2009 to about 0.75 in 2019.

7.3.2 Threshold contagion models

Number of systemically important countries. For each year of our data, we run our threshold model (with update rule (7.8)) for the multiplex global financial network and (with update rule (7.11)) for all monolayer networks (namely, the equity, debt, banking, and aggregate networks). For the multiplex network, we consider several different values of the (homogeneous) vertical threshold. As a reference choice, we take $c = 0.33 \approx 1/3 = 1/L$, but we also consider vertical thresholds of 0.5, 0.66, and 0.80. As we show in Fig. 7.6, for $c = 0.33$, the number of systemically important countries in our simulations on the multiplex network (dashed curves) is more than double the number that we obtain in our simulations in each type of monolayer network (solid curves). Furthermore, for vertical thresholds of 0.33, 0.50, and 0.66, the multiplex network has a larger number of systemically important countries than the monolayer networks; for a vertical threshold of 0.80, the number of systemically important countries in the multiplex network is similar to that in the monolayer network. We also observe that the number of systemically important countries tends to increase over time both in the multiplex network and (to a lesser extent) in the equity network. It is not clear if there is a concretely identifiable trend in the other types of monolayer networks.

Effects of capital buffers. A financial entity can reduce its risk of distress by a precautionary building of its capital buffers, which it can use at the time of financial distress.

Because our horizontal thresholds encode capital buffers, we examine how capital buffers can reduce global systemic risk by considering different sizes of these thresholds. We consider seven scenarios. In the first three scenarios, the horizontal thresholds are $0.67 \approx 2/3$, 0.8 and 0.9 times the CAR (i.e. 33%, 20%, and 10% below the CAR). In the fourth scenario, the horizontal thresholds are equal to the CAR. In the last three scenarios, the horizontal thresholds are 1.10, 1.20 and 1.33 times the CAR (i.e. 10%, 20%, and 33% above the CAR). We count the number of systemically important countries in each scenario and for each year (see Fig. 7.7). We find that the horizontal threshold strongly influences the number of systemically important countries in both the multiplex and the monolayer networks. For example, as we show in Fig. 7.7, decreasing the horizontal threshold (i.e. the capital buffer) by approximately one third can double the number of systemically important countries for both the aggregate and the multiplex networks. We also observe a decrease in the number of systemically important countries in the year 2016.

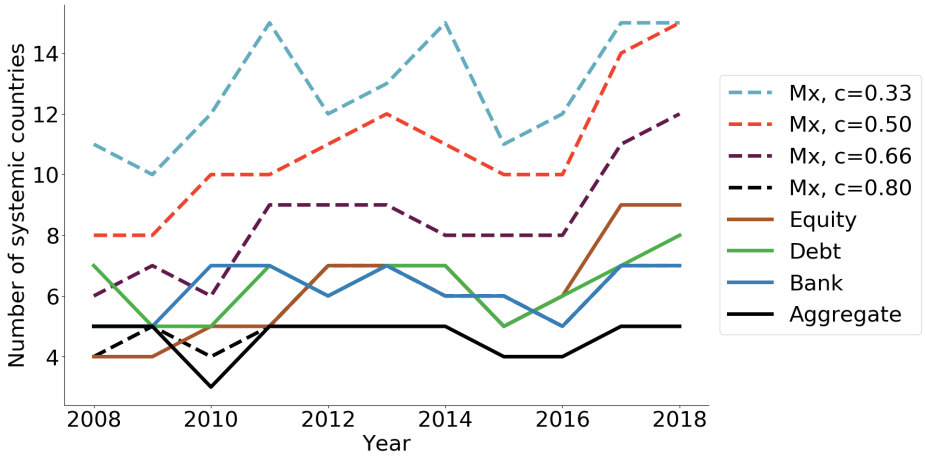


Figure 7.6: **The number of systemically important countries (which we abbreviate as "systemic countries" in the axis label)** over time in the multiplex global financial network and in its associated monolayer networks. We show simulations on the multiplex global financial network (which we abbreviate by "Mx" in the legend) using dashed curves and simulations on the monolayer networks using solid curves. For the multiplex network, blue, orange, purple, and black correspond to vertical thresholds of 0.33, 0.50, 0.66, and 0.80, respectively. For the aggregated network, brown, green, blue, and black correspond to the equity, debt, banking, and aggregated networks, respectively.

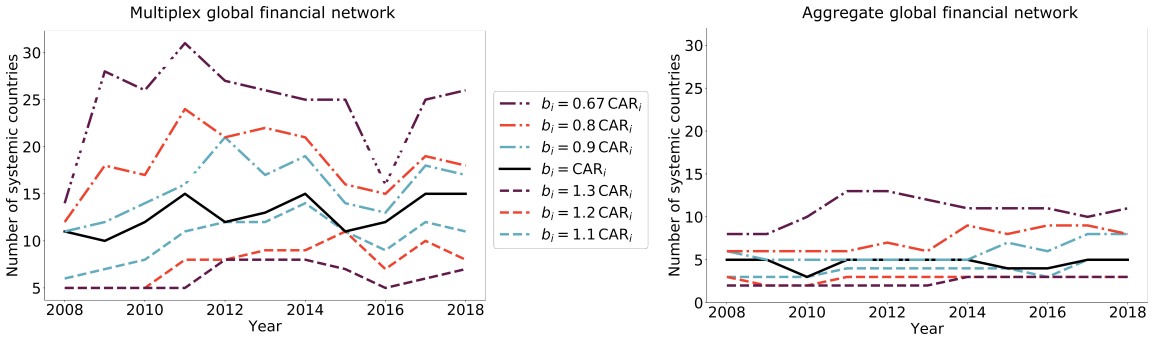


Figure 7.7: **Systemically important countries and horizontal thresholds.** We show the number of systemically important (i.e. "systemic") countries over time for different values of the horizontal threshold (i.e. capital buffer) for the multiplex global financial network (in the *left* panel) and aggregate network (in the *right* panel). The solid curve is for the baseline values $b_i = \mathrm{CAR}_i$, the dot–dashed curves are for values that below the baseline, and the dashed curve are for values above the baseline. The colour of a curve indicates a specific value of the threshold.

We observe which countries are systemically important in each of the seven scenarios (see Fig. 7.8). To illustrate our results, we examine countries that are systemically important in four or more years for the baseline horizontal threshold $b_i = \mathrm{CAR}_i$ (i.e. the fourth scenario, which is a baseline scenario) and check whether they are systemically important or not in the other scenarios. In Fig. 7.8, we illustrate that some countries are not systemically important in the baseline scenario, but become systemically important in scenarios in which the horizontal thresholds are below the baseline. We also observe that some countries are systemically important when the horizontal thresholds equal the CAR, but they are not systematic in scenarios in which the horizontal thresholds are above the baseline.

For example, for the year 2018, an increase of the CAR by 10% reduces the number of systemically important countries from 15 to 12, and a 20 % increase further decreases the number of systemically important countries to 8. However, there are several countries that remain systemically important for most years even when we consider a CAR that is 33 % larger than the baseline. These countries are the United States, the United Kingdom, Luxembourg, and The Netherlands. We also find that China, Hong Kong SAR, and Singapore are systemically important in more scenarios in more recent years.

Figure 7.8 is helpful for comparing our insights from our MultiRank and threshold-model analyses. Both MultiRank and our threshold model suggest that (i) the United States, the United Kingdom, Luxembourg, and The Netherlands are the most systemically important countries and that (ii) several Asian countries have increased in systemic importance over the years. However, although China and Hong Kong SAR are systemically important in more scenarios than Germany from 2014 to 2018, Germany's MultiRank is larger than the MultiRanks of both China and Hong Kong SAR. Similarly, Italy has a larger MultiRank than China and Singapore from 2008 to 2016, and it has a larger MultiRank than Hong Kong SAR from 2008 to 2013. However, China, Singapore, and Hong Kong SAR are systemically important in the same or more scenarios than Italy from 2010 to 2018 (with the exception of Singapore in 2013 and 2016).

7.4 Discussion

Our modelling and computations provide new insights into how different mechanisms of the spreading of financial distress can affect global financial stability. Specifically,

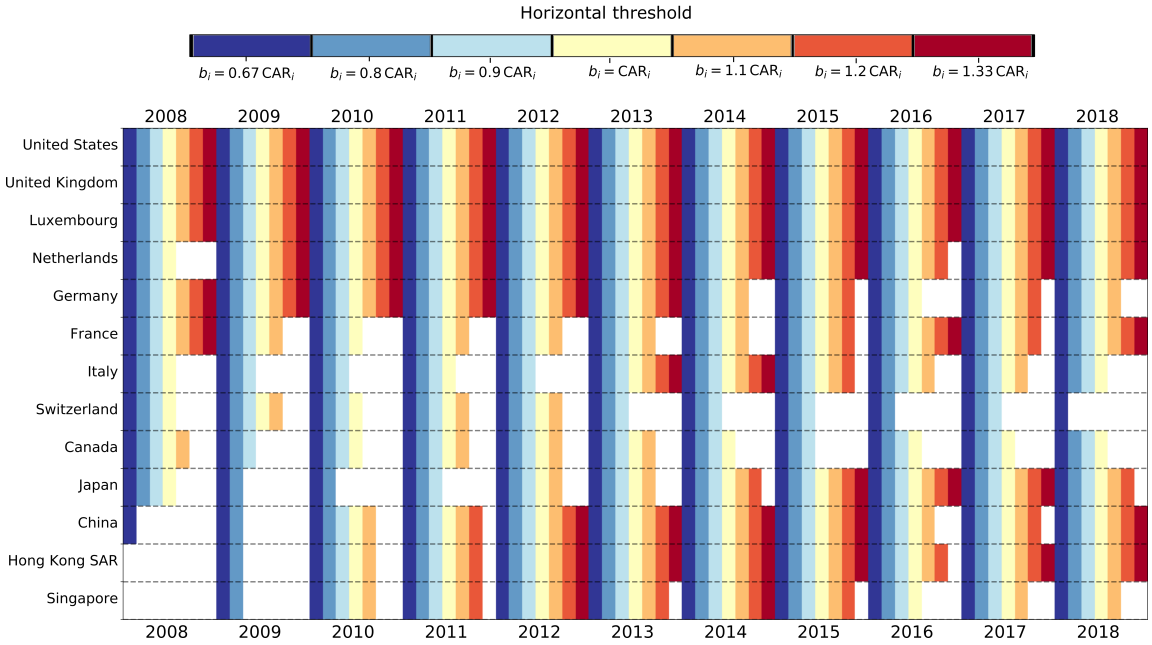


Figure 7.8: **Buffer size and systemic importance of countries in our multi-layer threshold model.** The coloured entries signify that a country is systemically important for a given year and horizontal threshold, and a white entry signifies that a country is not systemically important. The blue entries correspond to a country being systemically important for a horizontal threshold that is smaller than the baseline value $b_i = \mathrm{CAR}_i$, and the red entries indicate systemically important countries for a horizontal threshold that is larger than the baseline. For this plot, we select all countries that are systemically important in at least four years for the baseline horizontal threshold.

using a multilayer threshold model that incorporates both "horizontal" (between different countries) and "vertical" (between different assets for a given country) channels of contagion, we show that a financial contagion can be more extreme for multiplex network structures than for corresponding monolayer networks. In other words, we obtain a larger number of systemically important countries when we use the multilayer threshold model. We also illustrated that several Asian countries increase in their centrality rank over time in the global financial networks, whereas several European countries decrease in their centrality rank over time. We also observed that increasing the horizontal thresholds (which we interpreted as capital buffers) decreases the number of systemically important countries.

We obtained a few similar insights in our MultiRank and threshold-model analyses. Both approaches identified the same top-four systemically important countries and pointed to a rise in systemic importance of some Asian countries. However, we also found that having a larger MultiRank value does not imply that a country is systemically important in all scenarios. When comparing our results for multiplex and monolayer networks, we find that MultiRank and PageRank correlate strongly with each other. By contrast, we observe important differences between the results of the multiplex and monolayer threshold models. Our threshold model suggests that the multiplex structure of the global financial network has a larger number of systemically important countries than its aggregate-network counterpart. These results motivate the development of multilayer centrality measures that are based on contagion dynamics to provide insight about systemic risk.

It is common to use threshold models to model systemic risk (Caccioli et al. 2018). In our multiplex threshold model, we observed for all years (2008–2018) that we examined that there are many more systemically important countries than in the aggregate threshold model. Therefore, our results stress the need for further research on multilayer networks and financial contagions to better understand systemic risk. For example, it would be beneficial to conduct a detailed study of how initial conditions and parameter values affect qualitative results and to compare the observations of such a study with previous financial crises. It is also essential to develop models that go beyond threshold models, as we discuss below and has been suggested by Petrone & Latora (2018).

In the present work, we incorporated the fact that there are multiple types of assets by examining multiplex financial networks. In future work, it will be useful to allow the horizontal and vertical thresholds to be heterogeneous across asset types in a way that incorporates the different characteristics of different assets. For instance,

one can relax the assumption of a country-wide CAR, a measure that is relevant traditionally for banks, and incorporate heterogeneous buffers (such as the amount of corporate cash holdings). One can also take into account additional considerations, such as the level of a country's reserves or its access to emergency borrowing facilities, which may provide protection against losses from financial distress.

One can also incorporate previous approaches to modelling financial distress propagation within a country (Gray 2013) by considering coupling between the public[7], private non-bank, and banking sectors. These three sectors have different types of asset exposures, so they may in turn have different risk of financial distress. Financial distress can originate in any of these sectors, and a model can examine how financial distress can propagate between these sectors within a country and across countries. For instance, a crisis that originates in a banking sector can weaken the private non-banking sector and the public sector, which in turn may exacerbate the banking distress itself (Farhi & Tirole 2018). An analogous situation can occur if a crisis originates in the public sector.

7.5 Concluding remarks

We studied financial contagions in multiplex networks of financial assets. In the years following the 2008 global financial crisis, some countries reduced their banking assets and increased their investments in equity. We calculated centrality measures — MultiRank for the multiplex global financial network and PageRank for the monolayer networks — and found that there was a decreasing trend in the ranking of the systemic importance of some major European countries after the 2008 global financial crisis, whereas there was an increasing trend in the ranking of the systemic importance of some major Asian countries. Building on prior work (see, e.g. Caccioli et al. (2018), Gai & Kapadia (2010)) on threshold models of financial contagions, we developed a multiplex threshold model to examine financial contagions in multiplex global financial networks. Our model includes both "horizontal" (i.e. intralayer) and "vertical" (i.e. interlayer) mechanisms, with associated thresholds for the propagation of financial distress. We examined a variety of parameter values, and we observed for several values of the vertical threshold that financial shocks in the multiplex networks yield a larger number of financially distressed countries than in the monolayer networks. Similarly, for several values of horizontal thresholds, we observed that financial shocks

[7]The public sector is the part of the economy that consists of public services and public enterprises.

in the multiplex networks yield a larger number of systemically important countries than in the aggregate network.

Our study highlights the importance of considering the multiplex nature of financial contagions, including by examining both "horizontal" (i.e. intralayer) and "vertical" (i.e. interlayer) spreading mechanisms, and we expect that incorporating multiple contagion channels will also be insightful for future studies of social contagions.

Conclusion

Many economists have underscored the importance of developing non-equilibrium models to guide policy (Arrow 1986, Farmer & Foley 2009, Arthur 2021). These models have can to enable policy labs, where we run controlled computer experiments to test economic policies (Arthur 2021). However, up until recently, the lack of data and computational resources made it difficult to develop non-equilibrium models.

In this thesis, we presented two data-driven models that have the potential to be used to address policy-related questions. The first model focused on quantifying the impact automation and the COVID-19 pandemic have on employment. The second is a model of global financial contagion and represents a first step towards introducing multilayer network models which policy institutions, such as the IMF, could use to identify systemically important countries.

In our labour market model, the agents are workers who react to the changing nature of the economy, in this case, labour demand shocks. When the demand for labour remains constant, our model reaches a steady-state, where the system fluctuates around a fixed-point. In contrast to many equilibrium models, a steady-state behaviour is not imposed top-down, but emerges naturally from the agents' behaviour. The emergence of a steady-state is an important result, but the main advantage of our model lies in studying the economy during transition periods.

We explored the macroeconomic behaviour of our model during business cycles. We showed that our model reproduces the Beveridge curve, and most importantly, the counter-wise cycling direction. We also modelled the economy in the presence of structural changes induced by automation and by the COVID-19 pandemic. We found that the network of occupational mobility plays an essential role in determining workers' employment prospects. Furthermore, we showed that we can use fine-grained data to predict economic shocks, such as the COVID-19 pandemic, and then model how the economy behaves during a transition period. Both the shock prediction and the model exploration can be made in close to real-time and therefore help guide policy in times of crisis. Uncovering the mechanisms behind the Beveridge curve's

behaviour can help design policies aimed at accelerating the recovery process, while understanding how the economy behaves under structural changes can help workers better prepare for the changing demands of the labour market.

In the second part of this thesis, we developed a multiplex threshold model to identify systemically important countries. We here found that ignoring the different contagion channels due to investment types may lead to underestimating the number of countries that can trigger financial distress in many other countries.

The results from both parts of the thesis are promising, but further work is needed to obtain more reliable quantitative results. In the labour market model, we could add a more realistic behaviour of workers by considering multiple job applications and job-to-job transitions (i.e. allowing employed workers to search for jobs). A key improvement would be to incorporate wage dynamics, where workers prefer applying for jobs in high-wage occupations and where employers may increase wages to attract more job applications when labour supply is scarce. After incorporating wage dynamics and more realistic search behaviour, we could refine our calibration method, and test our model via out of sample predictions. In particular, one could calibrate the model using the job transitions data before the pandemic and validate the model by predicting job transitions during the pandemic. Our model could easily be coupled with production models (e.g. input-output models) via the labour demand. This coupling could help us better understand the economy on a larger scale.

To improve the multiplex network model of financial contagion, one could go beyond threshold models and develop multiplex network models based on equations that describe the dynamics of financial assets. We could also explore the behaviour of threshold models further using mathematical tools. In particular, the message passage approach has been used extensively to better understand propagation mechanisms in tree-like networks.

At the time of writing, it is the start of the year 2021, the pandemic is ongoing, and the economy has not yet converged to an equilibrium. Many workers in the retail and entertainment sectors are unemployed, while pharmaceutical companies cannot meet vaccine demand. There have been several COVID-19 economic studies of excellent quality, but governments are still looking for policies to stop the epidemic spread while minimising economic losses. In the past, when computer power was still limited and data were scarce, using analytical techniques and making strong assumptions on agent rationality seemed to represent the best way to understand the economy. Nevertheless, the recent increase of data availability and computer power allows us to model the economy as a complex system, one which is not in equilibrium and formed

by heterogeneous agents who are not perfectly rational. This thesis contributes to the development of non-equilibrium models that can answering economic policy questions in times of rapid change.

Appendix

I.1 Occupational mobility and automation

I.1.1 Self-loop heterogeneity

Due to a lack of quality data on workers that switch employers but stay in the same occupation, we assumed that the self-loops weight is the same for all occupations, $r_i = r \quad \forall i$. In this section, we explore a way of calibrating heterogeneous self-loops, that is, allowing each occupation i to have a self-loop with weight r_i.

While same-occupation job switches are scarcely recorded in the census, we can use some census information to define heterogenous self-loops after making some assumptions. IPUMS Current Population Survey has an 'EMPSAME' variable indicating whether a person's employer is still the same. However, this data has significant quality issues. Specifically, IPUMS reports that "there are a large number of people in all samples who appear to be in-universe for EMPSAME, that is, they are currently employed and were employed in the previous month, but still have NIU (Not in Universe) values for EMPSAME".

When we use these data to estimate the self-loops, we find unrealistically low values. For example, 41 occupations would have a zero self-loop, meaning that workers in these occupations always change occupation when they change employer. Furthermore, the highest self-loop weight we observe is 0.19, corresponding to lawyers and judges. This would imply that four-fifths of lawyers and judges change occupation when they change employer. However, even though the self-loop weights are unrealistically low, we find that the heterogeneity between occupations is reasonable. Licensed occupations such as lawyers and judges, nurses, and surgeons have high self-loops compared to the rest of the occupations. In the next paragraph, we explain how, with some assumptions, we can use this heterogeneity in the data of workers remaining or not in the same occupation to set a relative likelihood of workers to remain in their occupation.

First, we assume that occupations have positive self-loop weight i.e., $r_i > 0 \quad \forall i$. We assign a 0.0007 self-loop weight to the 41 occupations which previously had $r_i = 0$. We choose 0.0007 since it is the minimum positive weight of the original distribution of self-loops. Second, we define linearly re-scaled r'_i weights, which we obtain using the functional form $r'_i = mr_i + b$, where m and b are parameters and r_i is the self-loop weight of occupation i for the original distribution. We search for parameter values of m and b that satisfy the conditions that i) the maximum self-loop weight is 0.95 and ii) that $\frac{1}{n}\sum_{i=1}^{n} r_i = r$, where r is the homogenous self-loop weight we calibrated as explained in Chapter 3. In other words, we add the condition that the mean of the distribution of the re-scaled self-loops should equal the weight of the homogenous self-loop we previously calibrated to match the annual occupational mobility rate (see Fig. I.1). The parameter values we find are $m = 0.5$ and $b = 0.5$. We used the re-scaled self-loop weight r'_i to run the model with heterogenous self-loops.

Our results vary little when we introduce heterogeneous self-loops. As we show in Fig. I.2 the effect is mild even at the aggregate level. Few occupations deviate from the identity, and none differ by a large amount.

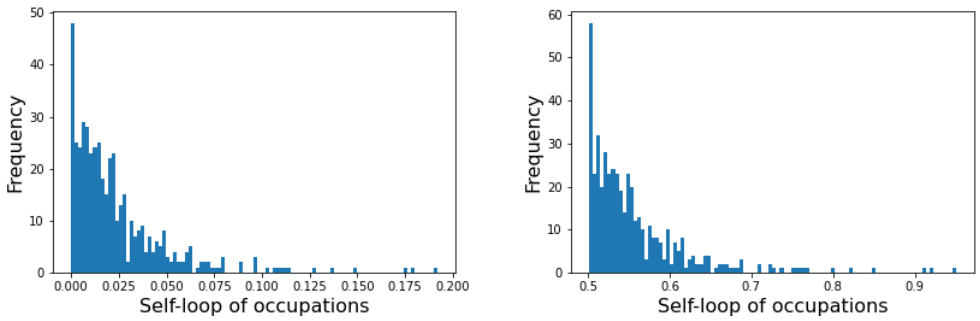


Figure I.1: **Occupation's self-loop distribution.** In the *left* panel, self-loops of occupations before setting minimum value and linear re-scaling. In the *right* panel, self-loop of occupations after setting minimum value and re-scaling.

I.1.2 Results for different values of γ

Here, we check the robustness of our results with respect to different values of the γ parameter. As explained in the main text, γ represents the rate at which the market adjust towards the target demand. In our model δ_u and δ_v represent the probability that workers (vacancies) and separated (open) due to random events; and γ is a adjustment rate. Therefore, we expect that $\gamma \geq \delta_u$ and $\gamma \geq \delta_v$. In the main text

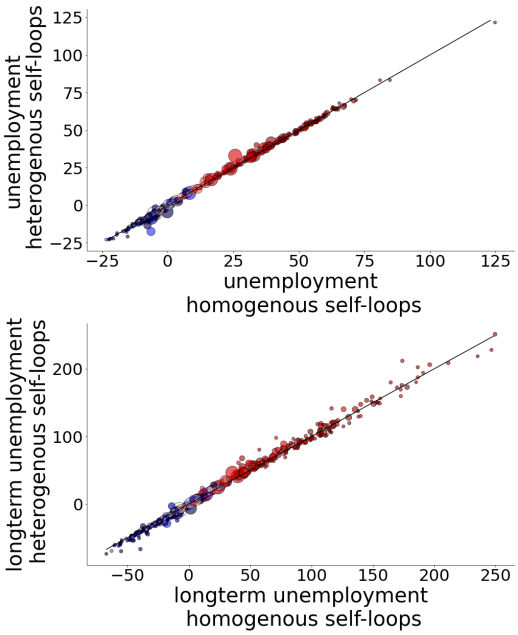


Figure I.2: **Heterogenous self-loops effect** We compare the model results between using homogenous and heterogenous self-loops, for both percentage change in unemployment and long-term unemployment change

we use $\gamma = 10\delta_u$ as a reference point. In this section we explore how our results change for different values. In particular we test for $\gamma = 5\delta_u$ and $\gamma = 20\delta_u$. We chose these ranges since there is little change in the results for larger values of γ and for lower values, we obtain unreasonably high values of the unemployment rate at the aggregate level (more than 15%).

In Fig. I.3, we plot the percentage change in the unemployment rate using $\gamma = 10\delta_u$ (our benchmark) vs the percentage change in unemployment rate when $\gamma = 5\delta_u$ and $\gamma = 20\delta_u$ respectively. Our results show that the changes are very similar although as γ increases so does the increase in unemployment and long-term unemployment for occupations that are likely to be automated. These results are not surprising, since the larger γ is the faster the adjustment towards the target demand and thus sharper the shock.

I.1.3 Measuring the impact of automation

In the main text, we define the occupation-specific average unemployment and average long-term unemployment as follows:

$$u_{i,\text{average}}(T) = \frac{100}{T} \frac{\sum_{t \in T} u_{i,t}}{\sum_{t \in T} (u_{i,t} + e_{i,t})},$$

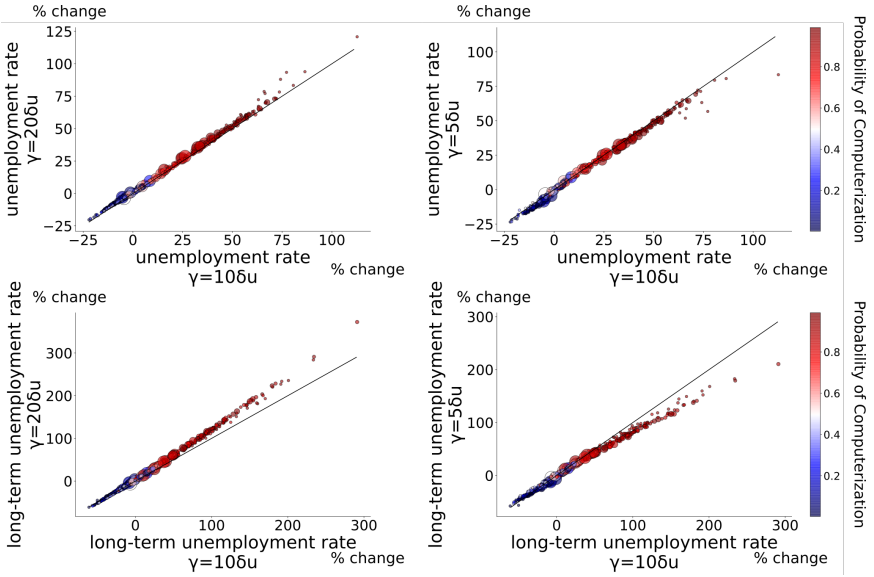


Figure I.3: **Change in unemployment and long-term unemployment with different values of γ.** The *top* panels show the change in unemployment rates vs the automation probability. The *bottom* panels show the change when $\gamma = 5\delta_u$ and $\gamma = 20\delta_u$ on the y axis and on the x-axis when $\gamma = 10\delta_u$.

and

$$u^{(\geq\tau)}_{i,\text{average}}(T) = \frac{100}{T}\frac{\sum_{t\in T} u^{(\geq\tau)}_{i,t}}{\sum_{t\in T}(u_{i,t}+e_{i,t})}.$$

However, there are other measurements of unemployment and long-term unemployment one can compute. For example, we can compute the unemployment rate at each time step of the automation period and then take the average. Thus, we define the alternative unemployment rate and the alternative long-term unemployment rate by

$$u_{i,\text{alternative}}(T) = \sum_{t\in T}\frac{u_{i,t}}{(u_{i,t}+e_{i,t})},$$

and

$$u^{(\geq\tau)}_{i,\text{alternative}}(T) = \sum_{t\in T}\frac{u^{(\geq\tau)}_{i,t}}{(u_{i,t}+e_{i,t})}.$$

In Fig. I.3, we compare the change in the average unemployment and long-term unemployment rate with the change in the alternative unemployment and long-term unemployment rate. On the top and right we show the change in the average unemployment rate in green and the change in the alternative unemployment rate in cyan. On the bottom and right we do the same for the long-term unemployment rate. Both these plots show that there is a strong overlap between the average change and the alternative change.

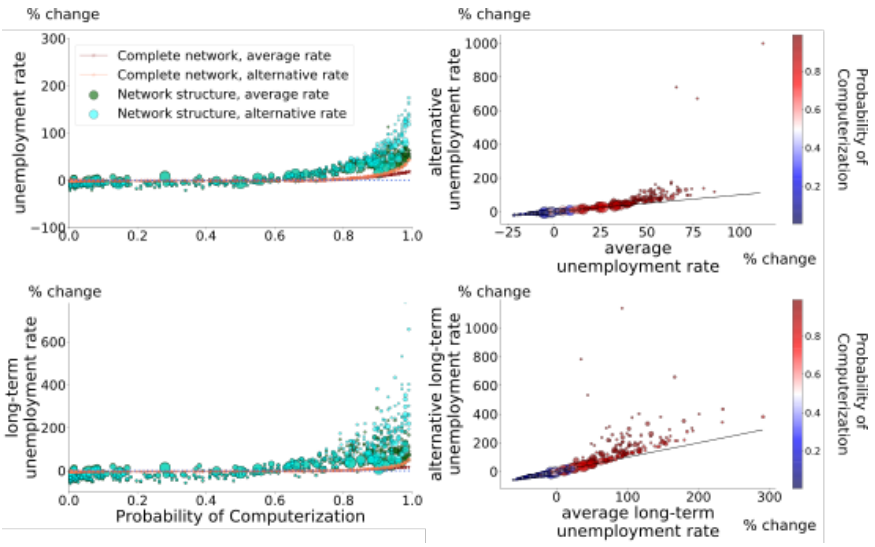


Figure I.4: **Average and alternative unemployment and long-term unemployment rates.** In the *left* panel, the percentage change in the average unemployment and long term unemployment rates in green and the percentage change in the alternative unemployment and long term unemployment rates in cyan. In the *right* panel, the percentage change of the average unemployment and long-term unemployment rate vs the percentage change of the alternative unemployment and long-term unemployment rate

For a better visualization we plot the change in the average unemployment rate vs the alternative unemployment rate on the top left panel. On the bottom left we plot the change in the average long-term unemployment rate vs the alternative long-term unemployment rate. We observe that almost all occupations lie close to the identity line with the exception of occupations that have low employment (small circles) and are highly likely to be automated (red color). The reason behind this discrepancy is that occupations that are highly likely to be automated and have low employment both increase the number of unemployed workers and also decrease the share of employment (due to the structural change). Thus, the ratio between the two, which is considered by the alternative unemployment rates, the increases considerably. Contrarily, when we measure the average unemployment rate, the initial share of employment prevents the sharp increase. However, both measurements exhibit the network effects – occupations with similar automation probabilities have different percentage change in their unemployment and long-term unemployment rates.

I.1.4 Automation shocks that change the aggregate demand

We assumed that the aggregate demand remains constant after the shock. In this section, we break this assumption and we run our model for an increasing and decreasing aggregate demand of 5%. We study the Frey and Osborne automation shock. In Fig. I.5, we plot the change in the period unemployment and long-term unemployment rate when the demand changes for each occupations vs the change in the period unemployment and long-term unemployment rate when the demand remains constant. As expected when the aggregate demand increases the percentage change in unemployment and long-term unemployment is lower and the points lie below the identity line. When the aggregate demand decreases, the percentage change in unemployment and long-term unemployment is higher and the points lie above the identity line. While there is a strong correlation between the changes in the unemployment rates when the demand changes and when the demand remains constant, occupations with low automation probabilities (blue dots) lie further away from the identity line. This result means that the structural part of the automation shock mostly affects the occupations with high estimates of automation. When we include a change in the aggregate labour demand then occupations with low automation estimates are also affected considerably.

I.1.5 Automation time and adoption rate

In this section we discuss how our results change when we assume a different duration of the automation shock. We assume that automation happens within 20 and 40 years, instead of 30. We measure the change in unemployment during the whole automation period and during the *steep* transition period. We define the steep transition period as the middle part of the automation period when of the sigmoid is steepest. In Fig. I.6 we highlight the whole automation period with a grey area coloring and the steep automation period by a coral shadowing.

As expected, the shorter the automation period is, the larger the increase in the aggregate unemployment and long-term unemployment rates (see top panels of Fig. I.6). On the bottom panels of Fig. I.6, we plot the percentage change in the unemployment and long-term unemployment rates of each occupation during the whole automation period vs the percentage change of the unemployment rates of each occupations during the steep automation period. There is a strong correlation between the change in unemployment during the whole transition period and during the steep transition period. However, during the steep automation period the percentage change in

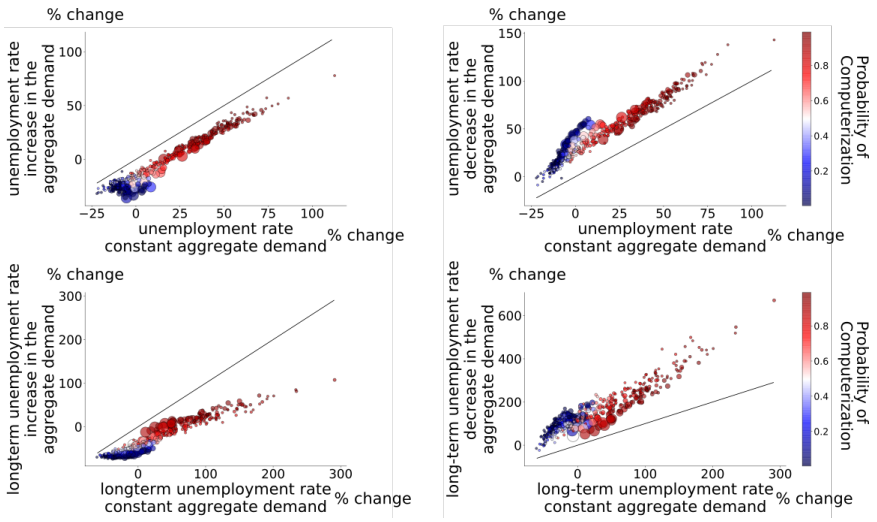


Figure I.5: **Frey and Osborne shock with different post-automation target demand scenarios**. In each panel we plot on the x-axis the percentage change in the period unemployment rates when the aggregate demand does not change and on the y-axis the percentage change in the period unemployment rate when the aggregate demand does change. In the *left* panels we assume the aggregate demand increases by 5%, and in the *right* we assume it decreases by 5%.

the unemployment rate is more extreme than during the whole automation period. Namely, occupations with high automation level have a higher percentage change in the unemployment and long-term unemployment rate during the steep automation period than during the whole automation period. Likewise, occupations with low automation level tend to decrease their unemployment and long-term unemployment rate more during the steep automation period than during the whole automation period

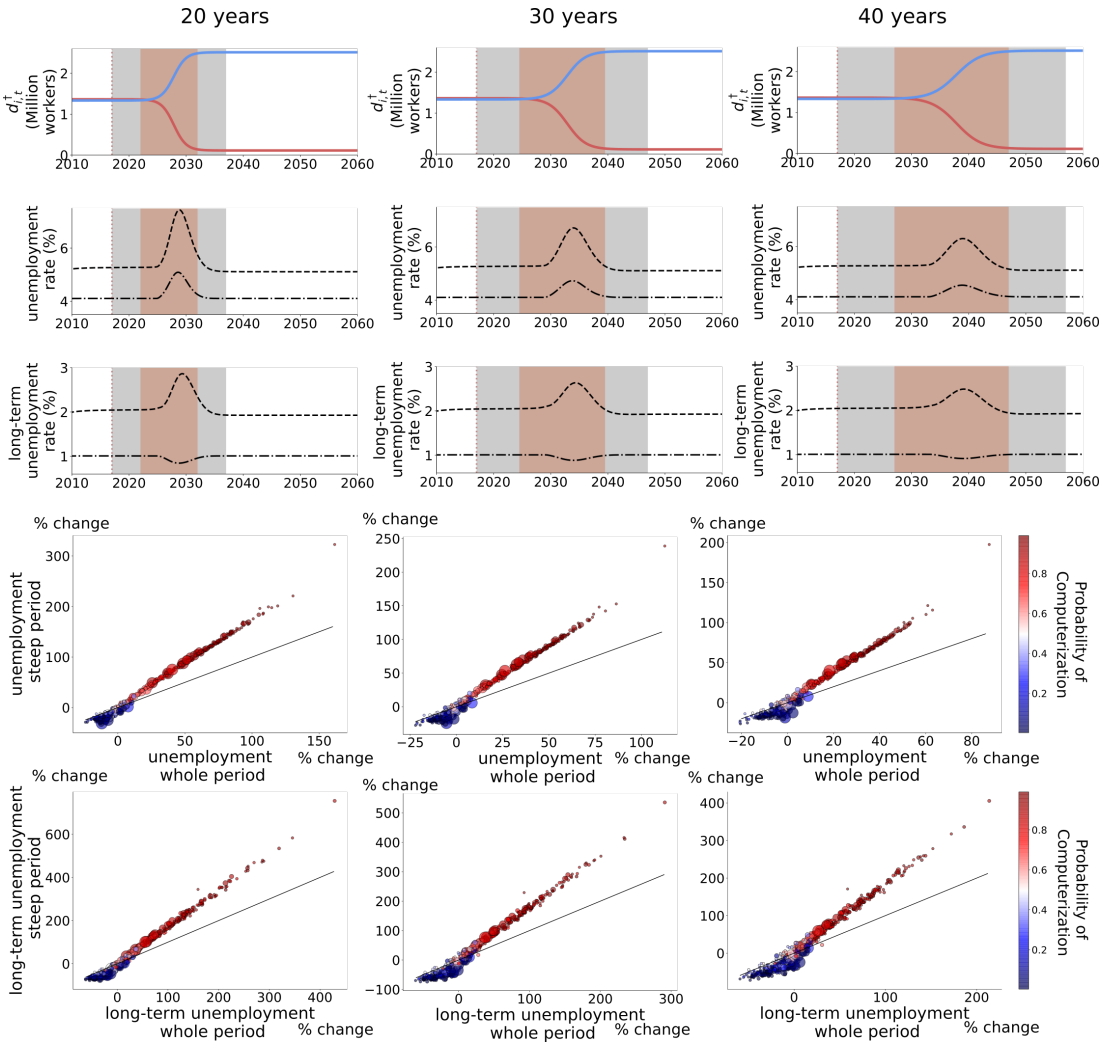


Figure I.6: **Shock duration and measuring windows effect on the measurements on unemployment rates.** In the *top.* panelf, for different duration of the automation shock (20, 30 and 40 years) we show target demand of two occupations with similar demand level (Childcare workers and officer clerks), the unemployment, and long-term unemployment rates. The grey area denotes the whole period of automation, meaning that the target demand has reached the automation level within a 1×10^{-4} tolerance. The coral area denotes the sharp transition period which is middle steepest part of the Sigmoid shock. In the *bottom* panels, for each occupation we plot the percentage change during the whole transition period vs the percentage change during the sharp transition period. Occupations are colored by their automation probability.

Appendix

II.1 Derivation of total shocks

II.1.1 Derivation of supply shocks

As discussed in the main text, we estimate the supply shock by computing an estimate of the share of work that will not be performed, which we compute by estimating the share of work that is not in an essential industry and that cannot be performed from home. We had to use several concordance tables, and make a number of assumptions, which we describe in details here.

Figure II.1 illustrates our method. There are four sets of nodes which are connected by three bipartite networks. The first set of nodes are the 6-digit NAICS industries which are classified to be essential or non-essential. This information is encoded in the K-dimensional column vector u which element $u_k = 1$ if NAICS 6-digit industry k is essential and 0 otherwise. Second, there are N different industry categories on which our economic analysis is based. The 6-digit NAICS codes are connected to these industries by the incidence matrix (concordance table) S. The third set of nodes are the J occupations obtained from the BLS and O*NET data. The weighted incidence matrix M couples industries with occupations where the element M_{nj} denotes the number of people in occupation j being employed in industry n. Fourth, we also have a list of I work activities. Each activity was rated whether it can be performed from home. If activity i can be done from home, the i^{th} element of the vector v is equal 1, and otherwise it is equal to 0. The incidence matrix T denotes whether an occupation is associated with any given work activity, i.e $T_{ji} = 1$ if activity i is relevant for occupation j.

The analysis presented here is based on $I = 332$ unique work activities, $J = 740$ occupations and $K = 1,057$ 6-digit NAICS industries. When relating to industry-specific results we use the BLS industry categories of the input-output accounts, leaving us with $N = 169$ industries for which we have reliable data on value added,

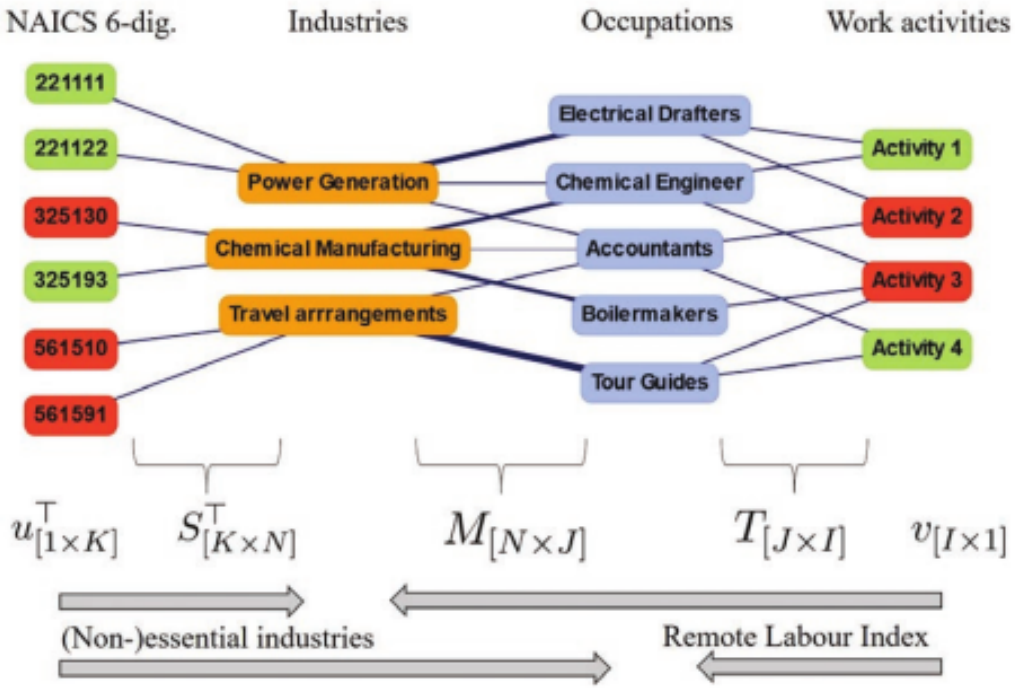


Figure II.1: **Schematic network representation.** The same schematic network representation of supply side shocks as in the main text, but now also including mathematical notation. The K-dimensional vector u below the NAICS 6-dig. (left nodes) encodes essential industries with binary elements. This set of nodes is connected to relevant industry categories by concordance tables (incidence matrix S). Matrix M connects the N industry categories with J occupations where an element represents the corresponding employment number. The ability to perform work activities (right nodes) from home is represented in vector v, also by binary elements. We use occupation-activity mappings provided from O*NET, represented as incidence matrix T. The grey arrows show the direction of shocks to industries and employment. The shock originating from the list of essential industries is mapped directly onto the broader industry categories, before it can be computed for occupations. Conversely, the Remote-Labour-Index is first mapped onto occupations and then projected onto industries.

total output and other key statistics. Employment, occupation and wage statistics are available on a more fine-grained 4-digit NAICS level. We therefore use these $N = 277$ industries for deriving labour-specific results.

II.1.2 Industry-specific shocks

We can use this simple framework for deriving the supply shocks to industries.

(Non-)essential industries. To estimate the extent to which an industry category is affected by a shutdown of non-essential economic activities, we measure the fraction of its 6-digit NAICS sub-industries which are classified as non-essential. In mathematical terms, the essential-score for every industry is therefore a weighted sum which can be written compactly in matrix notation as

$$e = \tilde{S}u, \tag{II.1}$$

where $\tilde{S}$ is the row-normalized version of matrix S with elements $\tilde{S}_{nk} = S_{nk}/\sum_h S_{nh}$.

Note that this assumes that the fine-grained NAICS codes contribute uniformly to the more aggregate industry categories. Although this assumption might be violated in several cases, in absence of further information, we use this assumption throughout the text.

Industry Remote-Labour-Index. We can similarly estimate the extent to which the production of occupations or industries can take place by working from home. Since work activities are linked to occupations, but not directly to industries, we need to take two weighted averages to obtain the industry-specific RLI.

For each occupation we first measure the fraction of work activities that can be done from home. We interpret this as the share of work of an occupation that can be performed from home, or 'occupation-level RLI'. This interpretation makes two assumptions: (i) that every work activity contributes equally to an occupation, which is our best guess since we do not have better data, and (ii) that if $z\%$ of activities cannot be done from home, the other $1 - z\%$ of activities can still be carried out and and are as productive as before.

For each industry i we then take a weighted average of the occupation-level RLIs, where the weights are the shares of workers employed in each occupation and in industry i. Let $\tilde{T}$ denote the row-normalized version of matrix T, i.e $\tilde{T}_{ji} = T_{ji}/\sum_h T_{jh}$ and similarly let the element of matrix $\tilde{M}$ be $\tilde{M}_{nj} = M_{nj}/\sum_h M_{nh}$. Then the industry-specific Remote-Labour-Index is given by the vector

$$r = \tilde{M}\tilde{T}v. \tag{II.2}$$

We interpret the Remote-Labour-Index for an industry, r_n, as the fraction of work in an industry n that can be performed from home. As for assumption (ii) above for the occupation-level RLI, this assumes that if $z\%$ of the work of occupations cannot be done, the other $1 - z\%$ of work can still be carried out.

Immediate industry supply shock. To derive industry supply shocks from the scores above, we need to take into account that industries might be exposed to both effects at the same time, but with different magnitudes. For example, consider the illustrative case of *Chemical Manufacturing* in Figure II.1. Half of the industry is non-essential (red node '325130') and could therefore be directly affected by an economic shutdown. But different occupations can be found in this industry that are affected heterogeneously. In this simple example, *Chemical Manufacturing* draws heavily on Boilermakers who have only work activities that cannot be done from home. On the other hand, this industry also has a tiny share of accountants and a larger share of Chemical Engineers who are able to do half of their work activities from home.

As stated above, the essential score e_n and the RLI r_n can be interpreted as shares of industry-specific work which can be performed, either thanks to being essential or thanks to being adequately done from home. To compute the share of industry-specific work that can performed due to either effect, we interpret shares as probabilities and assume independence,

$$\text{ISS}_n = -\left(1 - e_n\right)\left(1 - r_n\right), \tag{II.3}$$

where ISS stands for 'Industry Supply Shock'. We have multiplied the probability by minus one to obtain negative shocks. Although independence is a strong assumption, we have no reason to believe that the work that can be done from home is more or less likely to be judged essential. The empirical correlation coefficient of e and r is 0.03 and is far from being significant (p-value of 0.7). Under the linear assumptions about combining labour we are making here this indicates independence.

When applying these Industry Supply Shocks to value added, we make the implicit assumption that a $z\%$ decrease in labour will cause a $z\%$ decrease in value added.

II.1.3 Occupation-specific shocks.

We now describe how we compute shocks for specific occupations, rather than specific industries.

Occupations in (non-)essential industries. Occupations are mapped to industries through the weighted incidence matrix M, where an element denotes the number of jobs per occupation and industry. The column-normalized matrix M^* with elements $M^*_{nj} = M_{nj} / \sum_h M_{hj}$ denotes the share of an occupation carried out in a particular industry[1]. The essential-score for occupations is taken as weighted average

[1]Note that we column-normalize M to map from industries to occupations and row-normalize when mapping from occupations to industries.

of the essential score for industries (computed in Eq. II.1),

$$x = M^{*\top} \tilde{S} u = M^{*\top} e. \tag{II.4}$$

Occupation Remote Labour Index. As already indicated in the derivation of the industry-specific RLI, r, in Eq. (II.2), the occupation-specific RLI, y, is a weighted average of all the corresponding work activities that can be done from home. Formally, the occupation-based RLI is given by

$$y = \tilde{T} v. \tag{II.5}$$

Total supply-driven occupation shock. Following the same procedure as in Eq. II.3, we can get the total immediate shock on occupations from the economy's supply side[2]. The combined immediate shock to occupations is then given as

$$\mathrm{OSS}_j = -\left(1 - x_j\right)\left(1 - y_j\right). \tag{II.6}$$

Here, the correlation between RLI and the essential-score is larger, $\rho(x, y) = 0.30$ (p-value $= 8.3 \times 10^{-17}$), and significant which can also be seen from Figure 5.4. It should therefore be noted that the labour-specific results are expected to be more sensitive with respect to the independence assumption, as it is the case for industry-related results.

II.1.4 Derivation of demand shocks

Since we have demand shocks only on the 2-digit NAICS level, disaggregating them into the more fine-grained relevant industry categories is straightforward when assuming that the demand shock holds equally for all sub-industries. We let the industry demand shock in percentages for industry n be $-\mathrm{IDS}_n$.

To map the demand shocks onto occupations, we can invoke the same matrix algebra as above. The occupation-specific shock originating from the economy's demand side is then given by the projection

$$\mathrm{ODS} = M^{*\top} \mathrm{IDS}. \tag{II.7}$$

[2]To be clear, this is a product market *supply-side* shock, but this translates into a reduction in labour *demand* in each occupations.

II.1.5 Total immediate (first-order) shocks

We now combine supply- and demand-driven shocks to total immediate shocks for occupations and industries.

Let us turn to industries first. As discussed in more depth in the main text, the shock experienced is likely to be the worse of the two (supply and demand) shocks. Since we have expressed shocks as negative if they lead to decrease in output, in more mathematical terms, the *industry total shock* then is

$$\text{ITS}_n = \min(\text{ISS}_n, \text{IDS}_n). \tag{II.8}$$

and the *occupation total shock* is

$$\text{OTS}_j = \min(\text{OSS}_j, \text{ODS}_j). \tag{II.9}$$

Note that under these assumptions, the health sector will not experience a positive shock. Our estimates of the shocks suggest that there is an increase in demand, but we have not described positive supply shocks for the health sector. As a result, industries and occupations in the health sector will have a non-positive total shock.

Of course, the actual production of health care is likely to go up. But this is the result of an adjustment to an increase demand, rather than a first-order shock, and is therefore not modelled here. Nevertheless, see Appendix II.1.7 for an alternative.

II.1.6 Aggregate total shocks

To provide economy-wide estimate of the shocks, we aggregate industry- or occupation-level shocks. We do this using different sets of weights.

First of all, consider the interpretation that our shocks at the occupation-level represent the share of work that will not be performed. If we assume that if $z\%$ of the work cannot be done, $z\%$ of the workers will become unemployed, we can weight the occupation shocks by the share of employment in each occupation. Using the vector L to denote the share of employed workers that are employed by occupation j, we have

$$\text{Employment total shock} = \text{OTS}^T L \tag{II.10}$$

The Employment supply (demand) shock is computed similarly but using OSS (ODS) instead of OTS.

Instead of computing how many workers may lose their job, we can compute by how much paid wages will decrease. For each occupation, we compute the total wage

bill by multiplying the number of workers by the average wage. We then create a vector w where w_j is the share of occupation j in the total wage bill. Then,

$$\text{Wage total shock} = \text{OTS}^T w, \tag{II.11}$$

and similarly for the OSS and ODS. Note that we omit three occupations for which we do not have wages (but had employment).

Finally, to get an estimate of the loss of GDP, we can aggregate shocks by industry, weighting by the share of an industry in GDP. Denoting by Y the vector where Y_n is the VA of industry n divided by GDP[3],

$$\text{Value added total shock} = \text{ITS}^T Y, \tag{II.12}$$

and similarly for the industry supply and demand shocks (ISS and IDS). Note that we could have used shares of gross output and compute a shock to gross output rather to GDP.

II.1.7 Aggregate total shocks with growth of the Health sector

Here we make a different assumption about how to construct the total shock for occupations and industries. For industries, we assume that if they experience a positive demand shock, the industries are able to increase their supply to meet the new demand. Instead of Eq. (II.8) we use

$$\text{ITS}_n^h = \begin{cases} \text{ITS}_n, & \text{if } \text{IDS}_n \leq 0 \\ \text{IDS}_n, & \text{if } \text{IDS}_n > 0. \end{cases} \tag{II.13}$$

Since occupations are employed by different industries, the total shock to an occupation can be influenced by positive demand shocks from the healthcare sector and negative demand shocks from non-essential industries. In Eq. (II.9) we consider that occupations only experience the negative shocks. An alternative is to consider both the negative shock caused by non-essential industries and the positive shock caused by the health industries. This gives

$$\text{OTS}_j^h = \begin{cases} \text{OTS}_j, & \text{if } \text{ODS}_j \leq 0 \\ \text{ODS}_j + \text{OSS}_j, & \text{if } \text{ODS}_j > 0. \end{cases} \tag{II.14}$$

[3]Our estimate of GDP is the sum of VA of industries in our sample.

In Section 5.4.4, specifically Fig. 5.8 we use this convention for the y-axis, the labour Shock. Using Eq. (II.14) allows us to observe how health related occupations experience a positive shock.

In Table II.1, we show the aggregate total shocks when using Eqs. (II.13) and (II.14). There is very little difference with the results in the main text. The Health sector and its increase in demand are not large enough to make a big difference to aggregate results.

	Employment					Wages	Value Added
Shock	Aggregate	q_1	q_2	q_3	q_4	Aggregate	Aggregate
Total	-22	-41	-22	-20	-4	-15	-21

Table II.1: **Main results allowing for growth in the Health sector**. The results are the same as those presented in Table 5.4, but assuming that in industries, when demand is positive, the total shock is equal to the demand shock

II.2 Data

In this section we give more details about how we constructed all our variables. We stress that our goal was to produce useful results quickly and transparently, and make them available so that anyone can update and use them. We intend to improve these estimates ourselves in the future, as more information becomes available on the ability to work from home, which industries are essential, and how consumers react to the crisis by shifting their spending patterns.

II.2.1 Italian list of essential industries

The Italian list of essential industries[4] is based on the Statistical Classification of Economic Activities in the European Community, commonly referred to as NACE. The list of essential industries are listed with NACE 2-digit, 4-digit and 6-digit codes. We automatically map industries listed at the 2 or 4-digit NACE level to NAICS 6-digit industries using the crosswalk made available by the European Commission[5]. The 6-digit NACE level classification is country dependent and thus there is no official crosswalk to NAICS codes. We map the 6-digit industries by hand.

[4] Available at http://www.governo.it/sites/new.governo.it/files/dpcm_20200322.pdf, 22 March.

[5] https://ec.europa.eu/eurostat/ramon/relations/index.cfm?TargetUrl=LST_REL&StrLanguageCode=EN&IntCurrentPage=11

In a second step, we looked at the list of resulting list of industries and their essential score and discovered a few implausible cases, resulting from the complex mapping between the various classification systems at different levels. For instance, because Transports are essential, "Scenic and sightseeing transportation, other" was considered essential. In contrast, "Death care services" was classified as non essential. Three of us, as well as two independent colleagues with knowledge of the current situation in Italy, evaluated the list and we proceeded to editing the 4-digit NAICS Essential Scores as follows: From Non-essential to Essential: Grocery stores, Health and personal care stores, Gasoline stations, Death care services. From Essential (sometimes only partly) to non essential: Scenic and sightseeing transportation, Software publishers, Motion picture and video industries, Sound recording industries, and Other amusement and recreation industries. Finally, Federal, State and Local Government were not classified, and we classified as Essential.

II.2.2 Data for occupations

O*NET has work activities data for 775 occupations, out of which 765 occupations have more than five work activities. We compute the remote labour index for the 765 occupations with more than five work activities. From the May 2018 Occupational Employment Statistics (OES) estimates on the level of 4-digit NAICS (North American Industry Classification System), file *nat4d_M2018_dl*, which is available at https://www.bls.gov/oes/tables.htm under *All Data*, we find data for the number of employed workers of 807 occupations in 277 industries. These data covers 144M workers[6]. From the sample of 765 occupations with RLI, and from the sample of 807 occupations with employment data from the BLS, we are able to match 740 occupations, which cover 136.8M workers. Therefore, our final sample has 740 occupations and 136.8M workers.[7]

With the occupation-industry employment data and the essential score of each industry, we estimate the share of essential jobs within each occupation. Additionally, we have wage information for most occupations (i.e. we have median and mean wage data for 737 occupations). We computed all correlations for median wage considering all occupations we had median wage data for. For the 3 occupations for which median wage data was missing, the color coding of occupations in Figures 5.4, 5.7, and the

[6](the US economy had 156M workers mid-2018, see https://fred.stlouisfed.org/series/CE16OV)

[7]Note that the BLS employment data we use here does not include self-employed workers (which currently accounts for about 16 million people).

x-axis in Figure 5.8 corresponds to the average (across all occupations) of the median wage. We used the mean wages and the employment of occupations to define the wage quartiles of our sample. We excluded the 3 occupations for which we did not have mean wage from these calculations.

Finally, we use the O*NET data on exposure to disease and infection of occupations for the color coding in Fig. 5.8. We explain these data further in Appendix II.3. In the following Figs II.2 – II.4 we show the distribution of the remote labour index, exposure to disease and infection, supply, demand and overall shocks across occupations.

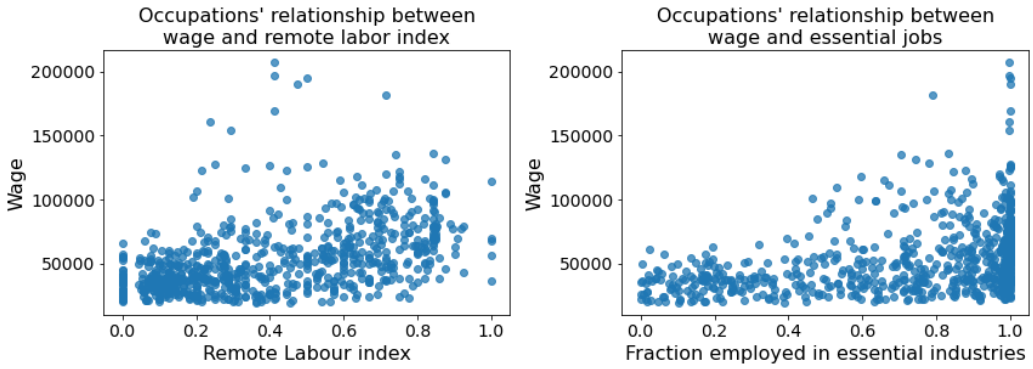


Figure II.2: **Distribution of remote labor index and essential score.** In the *left* panel, the relationship between remote labour index and median wage. The Pearson correlation is 0.42 (p-value $= 7.3 \times 10^{-33}$). In the *right* panel, the relationship between fraction of workers in essential industries and wage. The Pearson correlation is 0.35 (p-value $= 1.12 \times 10^{-22}$).

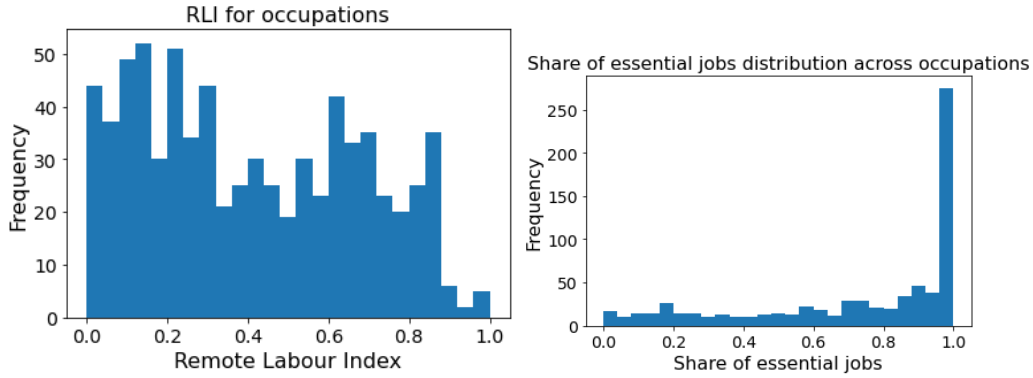


Figure II.3: **Distribution of remote labor index and essential score.** In the *left* panel, the distribution of the remote labour index for the 740 occupations. In the *right* panel, the distribution of the share of essential jobs within each of the 740 occupations.

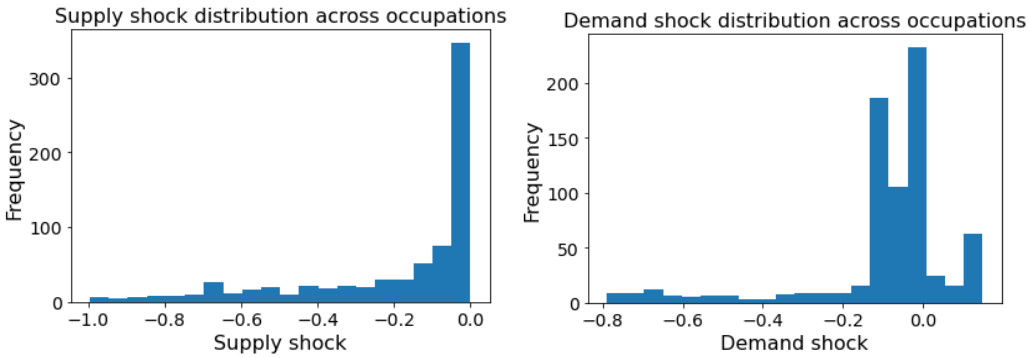


Figure II.4: **Distribution of shocks.** In the *left* panel, the supply shock distribution across occupation. In the *right* panel, the demand shock distribution across occupations.

II.2.3 Data for industries

Matching all data to BLS I-O industries. A key motivation of our work is to provide relevant economic data which can be used by other researchers and policymakers to model the economic impact of the COVID-19 pandemic. We therefore bring the discussed supply and demand shock data into a format that matches directly to US input-output data.

We use the BLS 2018 input-output account, which allows us to discern 179 private sectors. Moreover, there are the additional industries *Private Households*, NAICS 814 and *Postal Service*, NAICS 491. The data also contains 19 different industries relating to governmental activities. Since these industries are not classified with NAICS codes, we aggregate all governmental industries into a single node *Government*, which can be interpreted as the NAICS 2-digit industry 92. This leaves us with 182 industry categories which are a mixture of 2- to 6-digit NAICS industries.

We are able to match occupational data to 170 out of the 182 industry categories, accounting for 97% of total value added. For this subset we compute industry-specific RLIs, essential scores and supply shocks as spelled out in Appendix II.1.1, as well as employment-weighted infection exposures.

Since we have demand shocks only on the 2-digit NAICS level, disaggregating them into the more fine-grained BLS input-output data is straightforward when assuming that the demand shock holds equally for all sub-industries.

In the online data repository we also report total wages and total employment per industry. We use the same OES estimates as for the occupational data, but match every industry category according to the corresponding NAICS 2- to 6-digit digit levels.

Figures II.5 – II.7 show distributions of supply and demand shock-related variables on the industry level. Table II.2 summarizes a few key statistics for these industries, when further aggregated to 72 industry categories.

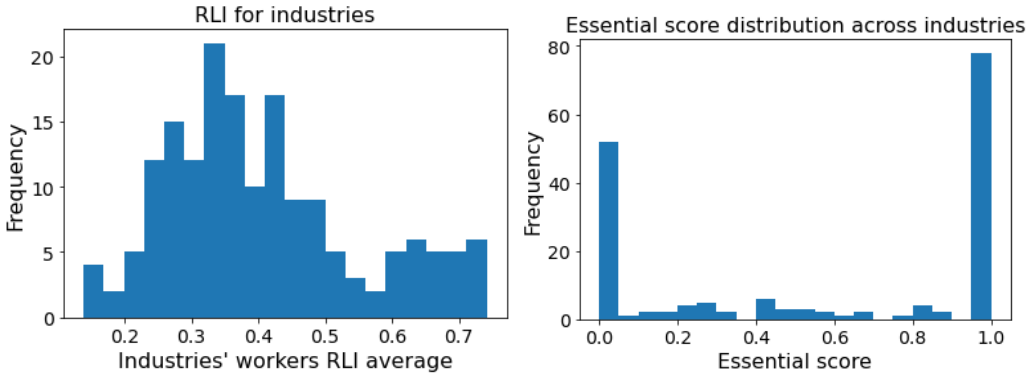


Figure II.5: **Distribution of remote labor index and essential score.** In the *left* panel, the distribution of the remote labour index, aggregated to 169 industries. In the *right* panel, the Fractions of essential sub-industries per industry category.

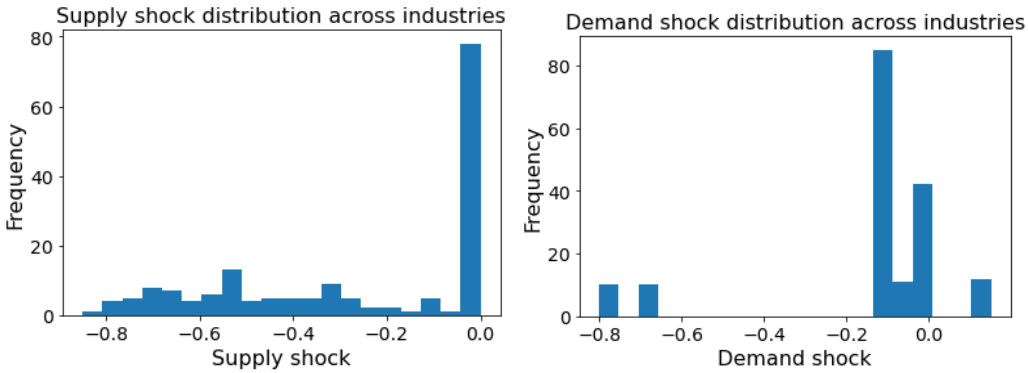


Figure II.6: **Distribution of shocks.** In the *left* panel, the supply shock distribution across industries. In the *right* panel, the demand shock distribution across industries.

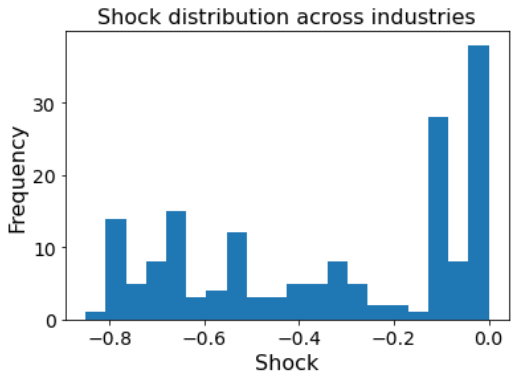


Figure II.7: **Shock distribution across industries.**

NAICS	Title	Outp.	Empl.	Demand	Supply	RLI	Essent.	Expos.
111	Crop Production	209	NA	-10	NA	NA	NA	NA
112	Animal Production and Aquaculture	191	NA	-10	NA	NA	NA	NA
113	Forestry and Logging	19	NA	-10	NA	NA	NA	NA
114	Fishing, Hunting and Trapping	10	NA	-10	NA	NA	NA	NA
115	Support Activities for Agriculture and Forestry	27	378	-10	0	14	100	5
211	Oil and Gas Extraction	332	141	-10	0	47	100	7
212	Mining (except Oil and Gas)	97	190	-10	-54	26	27	8
213	Support Activities for Mining	84	321	-10	-72	28	0	8
221	Utilities	498	554	0	0	42	100	9
23	Construction	1636	7166	-10	-24	31	66	8
311	Food Manufacturing	803	1598	-10	0	21	100	10
312	Beverage and Tobacco Product Manufacturing	192	271	-10	-3	33	96	7
313-4	Wholesale Trade	54	226	-10	-51	26	31	5
315-6	Management of Companies and Enterprises	29	140	-10	-68	25	9	4
321	Wood Product Manufacturing	118	402	-10	-62	26	16	6
322	Paper Manufacturing	189	362	-10	-8	24	89	6
323	Printing and Related Support Activities	80	435	-10	0	38	100	4
324	Petroleum and Coal Products Manufacturing	618	112	-10	-26	36	60	7
325	Chemical Manufacturing	856	828	-10	-2	38	96	9
326	Plastics and Rubber Products Manufacturing	237	722	-10	-8	28	89	7
327	Nonmetallic Mineral Product Manufacturing	140	NA	-10	NA	NA	NA	NA
331	Primary Metal Manufacturing	239	374	-10	-73	27	0	7
332	Fabricated Metal Product Manufacturing	378	1446	-10	-59	33	12	6
333	Machinery Manufacturing	386	1094	-10	-49	42	16	5
334	Computer and Electronic Product Manufacturing	369	1042	-10	-38	58	9	4
335	Electrical Equipment, Appliance, and Component Manufacturing	132	392	-10	-31	45	45	6
336	Transportation Equipment Manufacturing	1087	1671	-10	-58	37	9	5
337	Furniture and Related Product Manufacturing	77	394	-10	-47	35	28	4
339	Miscellaneous Manufacturing	173	601	-10	-16	40	74	11
42	Construction	1980	5798	-10	-27	50	46	7
441	Motor Vehicle and Parts Dealers	334	2006	-10	-23	43	60	11
442-4, 446-8, 451, 453-4	Wholesale Trade	1052	7731	-10	-39	53	17	19
445	Food and Beverage Stores	244	3083	-10	-33	43	43	15
452	General Merchandise Stores	240	3183	-10	-49	51	0	17
481	Air Transportation	210	499	-67	0	29	100	29
482	Rail Transportation	77	233	-67	0	33	100	11
483	Water Transportation	48	64	-67	0	35	100	10
484	Truck Transportation	346	1477	-67	-68	32	0	8
485	Transit and Ground Passenger Transportation	74	495	-67	0	27	100	43
486	Pipeline Transportation	49	49	-67	0	37	100	8
487-8	Management of Companies and Enterprises	146	732	-67	-10	37	85	8

491	Postal Service	58	634	-67	0	35	100	10
492	Couriers and Messengers	94	704	-67	0	37	100	15
493	Warehousing and Storage	141	1146	-67	0	25	100	6
511	Publishing Industries (except Internet)	388	726	0	-16	70	46	4
512	Motion Picture and Sound Recording Industries	155	428	0	-51	49	0	9
515	Broadcasting (except Internet)	196	270	0	0	65	100	6
517	Telecommunications	695	NA	0	NA	NA	NA	NA
518	Data Processing, Hosting, and Related Services	207	319	0	0	70	100	4
519	Other Information Services	192	296	0	-7	71	75	5
521-2	Construction	939	2643	0	0	74	100	11
523, 525	Wholesale Trade	782	945	0	-18	74	32	5
524	Insurance Carriers and Related Activities	1231	2330	0	0	71	100	8
531	Real Estate	1842	1619	0	-53	47	0	19
532	Rental and Leasing Services	163	556	0	-54	46	0	12
533	Lessors of Nonfinancial Intangible Assets (except Copyrighted Works)	182	22	0	-30	70	0	7
541	Professional, Scientific, and Technical Services	2372	9118	0	-2	64	94	9
55	Management of Companies and Enterprises	561	2373	0	0	66	100	8
561	Administrative and Support Services	971	8838	0	-37	35	44	17
562	Waste Management and Remediation Services	109	427	0	0	30	100	22
611	Educational Services	366	13146	0	0	54	100	29
621	Ambulatory Health Care Services	1120	7399	15	0	37	100	59
622	Hospitals	933	6050	15	0	36	100	65
623	Nursing and Residential Care Facilities	262	3343	15	0	28	100	61
624	Social Assistance	222	3829	15	0	40	100	47
711	Performing Arts, Spectator Sports, and Related Industries	181	505	-80	-51	44	8	13
712	Museums, Historical Sites, and Similar Institutions	20	167	-80	-49	51	0	16
713	Amusement, Gambling, and Recreation Industries	158	1751	-80	-65	35	0	21
721	Accommodation	282	2070	-80	-34	33	50	26
722	Food Services and Drinking Places	832	11802	-80	-64	36	0	13
811	Repair and Maintenance	235	1317	-5	-3	29	96	9
812	Personal and Laundry Services	211	1490	-5	-52	28	28	31
813	Religious, Grantmaking, Civic, Professional, and Similar Organizations	260	1372	-5	0	52	100	19
814	Private Households	20	NA	-5	NA	NA	NA	NA
92	All Public Sector (custom)	3889	9663	0	0	44	100	21

Table II.2: Key statistics for different 2- and 3-digit NAICS industries. Column 'Outp.' refers to total output of the industry in current billion USD (2018). 'Emp.' is total employment in thousands. 'Demand' is the immediate severe demand shock in % obtained from the CBO. 'Supply' is the immediate supply shock in % derived from the Remote Labor Index and the list of essential industries. 'RLI' is the industry-specific Remote Labor Index in %. 'Essent.' is the share of sub-industries being classified as essential in %. 'Expos.' denotes the industry-aggregated infection exposure index from O*NET which ranges from 0 to 100 and is explained in Appendix II.3. We make more disaggregated data with further details available in the corresponding data publication.

II.3 Occupations most at risk of contracting SARS-CoV-2

O*NET make available online work context data for occupations that describe the physical and social factors that influence the nature of work. The "Exposed to disease

and infection" work context[8], which we refer to as 'exposure to infection' for short, describes the frequency with which a worker in a given occupation is exposed to disease or infection. It ranges from 0 to 100, where 0 means "never" and 100 "everyday"; an exposed to infection rating of 50 means an exposure of "once a month or more but not every week" and 75 means "Once a week or more but not every day". We have exposure to infection data for 644 of the 740 occupations in our sample. For those occupations for which we did not have the exposure to infection, when colored them as if they had zero exposure to infection.

As we see in Fig. II.8 there is a U shaped relationship between wages and exposure to infection. There is a correlation of 0.36 (p-value $= 1.7 \times 10^{-21}$) between wages and exposure to infection, but this is misleading[9]. Though many high wage occupations are highly exposed to infection (high paid doctors), there are also many low wage occupations with high probability of infection.

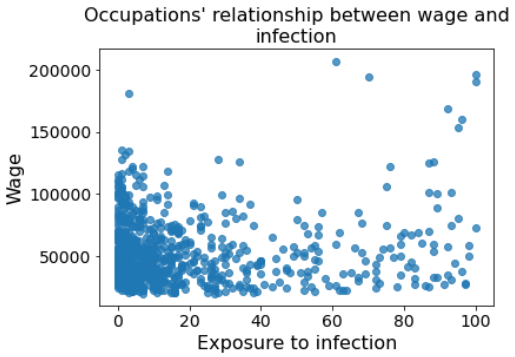


Figure II.8: **Probability of infection and wage.** Relationship between wage and probability of infection in linear and log scale. The Pearson correlation is 0.36 (p-value $= 1.7 \times 10^{-21}$). However we this correlation is mostly driven by high salaries in the health sector, but there are many low-wage occupations with a significant exposure to infection.

[8] https://www.onetonline.org/find/descriptor/result/4.C.2.c.1.b?s=2

[9] For example, Adams-Prassl et al. (2020), using survey evidence for ∼4000 US individuals, found that workers without paid sick leave are more likely to go to work in close proximity to others, which may have suggested a negative correlation between wages and exposures. Note however that our correlation is based on occupations, not individuals, and that wages are not necessarily an excellent predictor of having paid sick leave or not.

II.4 Discussion of labour supply shocks which we do not include

II.4.1 Labour supply shocks from mortality and morbidity

Typical estimates. McKibbin & Fernando (2020) consider attack rates (share of population who become sick) in the range 1-30% and case-fatality rates (share of those infected who die) in the range 2-3%. From attack rates and case fatality rates, they compute mortality rates. They also assume that sick people stay out of work for 14 days. A third effect they assume is that workers would be care givers to family members.

For their severe scenario of an influenza pandemic, Congressional Budget Office (2006) assumed that 30 percent of the workers in each sector (except for Farms, which is 10%) would become ill and would lose 3 weeks or work, at best, or die (2.5% case fatality rate).

Best guess for current effect of COVID-19. In the case of COVID-19, estimating a labour supply shock is made difficult by several uncertainties. First of all, at the time of writing there are very large uncertainties on the ascertainment rate (the share of infected people who are registered as confirmed cases), making it difficult to know the actual death rate.

We report the result from a recent and careful study by Verity et al. (2020), who estimated an infection fatality ratio of 0.145% (0.08-0.32) for people younger than 60, and 3.28% (1.62-6.18) for people aged 60 or more. The age bracket 60-69, which in many countries will still be part of the labour force, was reported as 1.93% (1.11–3.89).

Taking the infection fatality ratio for granted, the next question is the attack rate. In Verity et al. (2020), the infection fatality ratios are roughly one fourth of the case fatality ratios, suggesting that 3/4 of the cases are undetected. For the sake of the argument, consider Italy, a country that has been strongly affected and appears to have reached a peak (at least of a first wave). There are at the time of writing 132,547 cases in Italy[10] In 2018 the population of Italy[11] was 60,431,283. If we assume that Italy is at the peak today and the curve is symmetric, the total number of cases will be double the current number, that is 265,094, which is 0.44% of the population. If we assume that the true number of cases if 4 times higher, the attack rate is, roughly

[10] https://coronavirus.jhu.edu/map.html
[11] https://data.worldbank.org/indicator/SP.POP.TOTL?locations=IT

speaking, 1.76%. These numbers are more than an order of magnitude smaller than the number who cannot work due to social distancing.

Thus, while it is clear that the virus is causing deep pain and suffering throughout Italy, the actual decrease in labour supply, which is massive, is unlikely to be mostly caused by people being sick, and is much more a result of social distancing measures.

Uncontrolled epidemic. Now, it may be informative to consider the case of an uncontrolled epidemic. If we assume that the uncontrolled epidemic has an attack rate of 80% (a number quoted in Verity et al. (2020)), an infection fatality ratio for people in the labour force of 1% (an arbitrary number between 0.145% for people younger than 60, and 1.93% for the 60-69 age bracket) implies an 0.8% permanent decrease of the labour force. If we assume that those who do not die are out of work for 3 weeks, on an annual basis of 48 worked weeks, we have (3/48)*(0.80-0.01)=4.94% decrease of the labour supply.

Overall, this exercise suggests that left uncontrolled, the epidemic can have a serious effect on labour supply. However, in the current context, the effect on the economy is vastly more a result of social distancing than direct sickness and death.

II.4.2 Labour supply shocks from school closure

School closures are a major disruption to the functioning of the economy as parents can no longer count on the school system to care for their children during the day.

Chen et al. (2011) surveyed households following a school closure in Taiwan during the H1N1 outbreak, and found that 27% reported workplace absenteism. Lempel et al. (2009) attempted to estimate the cost of school closure in the US in case of an influenza pandemic. They note that 23% of all civilian workers live in households with a child under 16 and no stay-at-home adults. Their baseline scenario assumes that around half of these workers will miss some work leading to a loss 10% of all labour hours in the civilian U.S. economy, for as long as the school closure lasts.

Some of these effects would already be accounted for in our shocks. For instance, some workers are made redundant because of a supply or demand shock, so while they have to stay at home to care for their children, this is as much a result of labour and supply shock as a result of school closure. For those working from home, we might expect a decline in productivity. Finally, for those in essential industries, it is likely that schools are not close. For instance, in the UK, schools are opened for children of essential workers. Our list of essential industries from Italy includes Education.

Overall, school closure indeed have large effects, but in the current context these may already be accounted for by supply and demand shocks, or non-existent because schools are not fully closed. The loss of productivity from parents working from home remains an open question.

II.5 Additional estimates of demand or consumption shocks

In this appendix we provide additional data on the demand shock. Table II.3 shows our crosswalk between the Industry classification of the Congressional Budget Office (2006) and NAICS 2-digit industry codes, and, in addition to the "severe" shocks used here, shows the CBO's "mild" shocks. We have created this concordance table ourselves, by reading the titles of the categories and making a judgement. Whenever NAICS was more detailed, we reported the CBO's numbers in each more fine-grained NAICS.

We also provide two sources of consumption shocks (in principle, these estimates are meant to reflect actual decreases in consumption rather than shifts of the demand curve). Table II.4 shows the consumption shocks used by Keogh-Brown et al. (2010) to model the impact of potential severe influenza outbreak in the UK. Table II.5 shows the consumption shocks used by Muellbauer (2020) to model the impact of the COVID-19 on quarterly US consumption. OECD (2020) provided two other sources, both reported in Table II.5. The first one is based on assumptions of shocks at the industry level, while Table shows assumptions of shocks by expenditure categories (COICOP: Classification of individual consumption by purpose).

II.6 Shock dampening effects

In Chapter 5 we diminished the employment shocks by roughly 0.33. In this section, test different factors of shock reductions. In particular, we assume a 0.5 and 0.1 diminishing factor, which leads to shocks that decrease the aggregate labor demand by 10% and 18% respectively. As expected, the unemployment rate increases more during the pandemic when we assume a lower diminishing factor. Fig. II.9 shows the aggregate unemployment rates. When we assume a 0.5 shock diminishing factor, the pandemic's unemployment rate increases up to roughly 13%. When we assume a 0.1 diminishing factor for shocks, the unemployment rate increases above 20%.

Similarly, at the occupation level, we find a larger increase in the unemployment rates when we diminish the shocks by a smaller amount (see Fig. II.10). Qualitatively, the results are the same as the ones presented in Chapter 5. Most occupations increase their unemployment rates, with some exceptions in the healthcare sector. Furthermore, for the complete network, we find that the occupation-specific shocks determine the change in the occupation's unemployment rates.

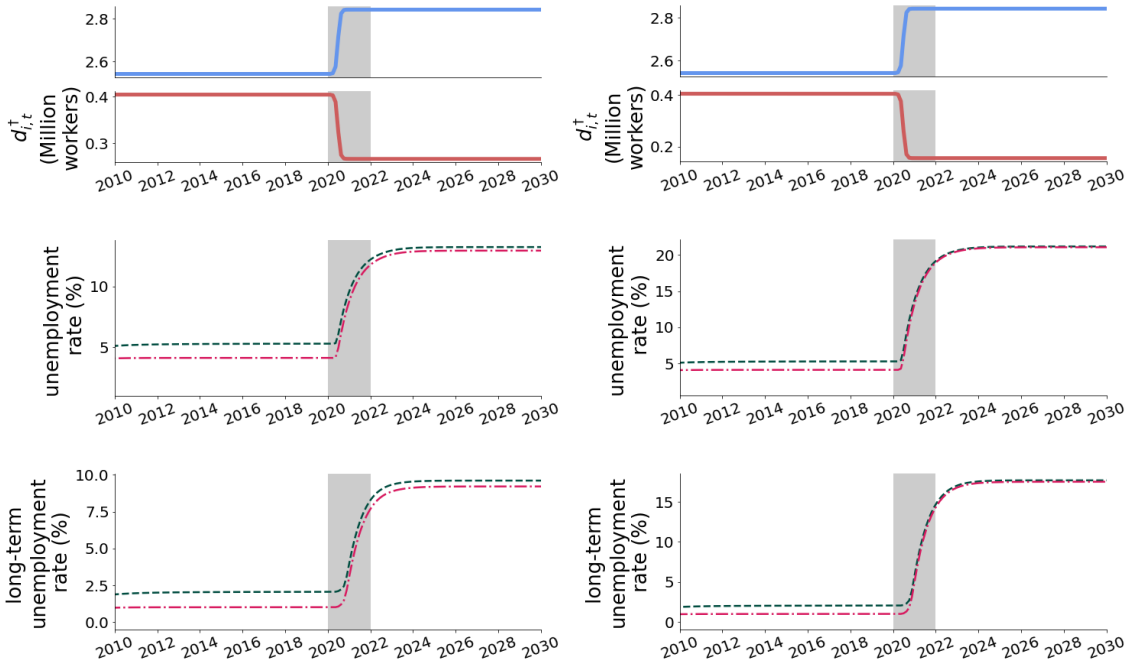


Figure II.9: **Unemployment and long-term unemployment rate at the aggregate level.** In all panels, the grey area denotes the period during which we measure the percentage change in unemployment and long-term unemployment. The *left* and *right* panels correspond to a 0.5 and 0.1 shock diminishing factor respectively. *Top* panels show the evolution of the target demand for two occupations. The occupation coloured in blue has a positive COVID-19 shock, while the occupation coloured in red a negative COVID-19 shock. *Middle and bottom* panels show the unemployment and long-term unemployment rates through time. The green line corresponds to the occupational mobility network while the red line to the complete network.

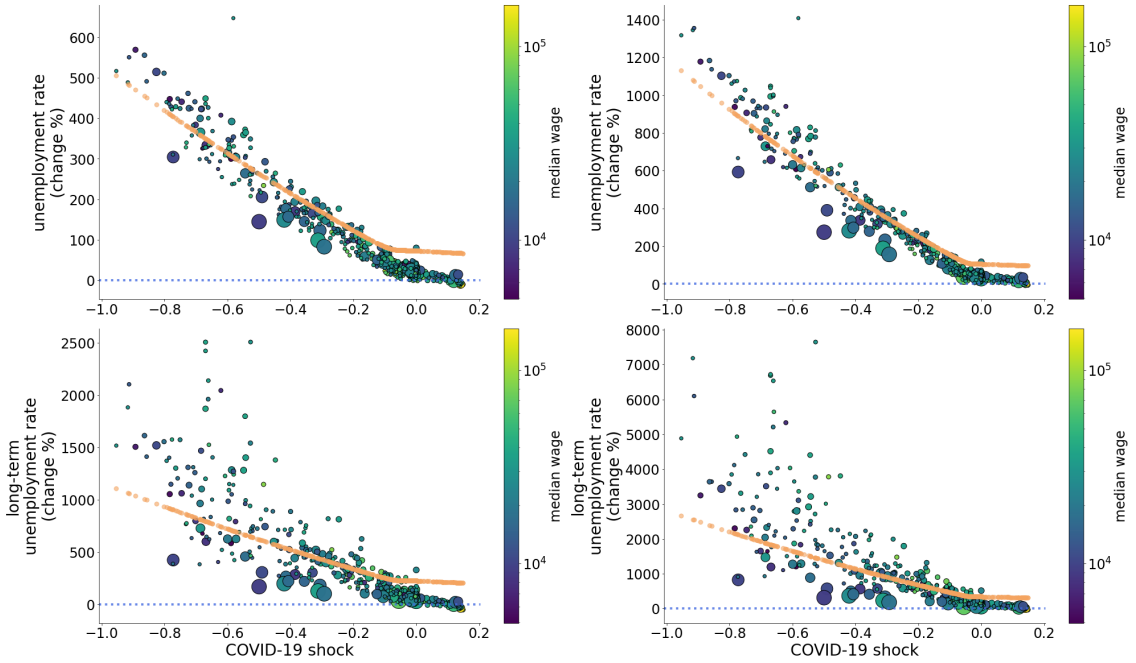


Figure II.10: **Occupation specific changes in unemployment and long-term unemployment.** The *left* and *right* panels correspond to a 0.5 and 0.1 shock diminishing factor respectively. The green dots are for the occupational mobility network and the orange dots are for the complete network. The size of the green dots is proportional to the employment of the occupation they represent and the colour gradient indicated the median wage of the occupation.

NAICS	NAICS	CBO	Severe	Mild
11	Agriculture, Forestry, Fishing and Hunting	Agriculture	-10	-3
21	Mining, Quarrying, and Oil and Gas Extraction	Mining	-10	-3
22	Utilities	Utilities	0	0
23	Construction	Construction	-10	-3
31	Manufacturing	Manufacturing	-10	-3
32	Manufacturing	Manufacturing	-10	-3
33	Manufacturing	Manufacturing	-10	-3
42	Wholesale Trade	Wholesale trade	-10	-3
44	Retail Trade	Retail trade	-10	-3
45	Retail Trade	Retail trade	-10	-3
48	Transportation and Warehousing	Transportation and warehousing (including air, rail and transit)	-67	-17
49	Transportation and Warehousing	Transportation and warehousing (including air, rail and transit)	-67	-17
51	Information	Information (Published, broadcast)	0	0
52	Finance and Insurance	Finance	0	0
53	Real Estate and Rental and Leasing	NA	0	0
54	Professional, Scientific, and Technical Services	Professional and business services	0	0
55	Management of Companies and Enterprises	NA	0	0
56	Administrative and Support and Waste Management and Remediation Services	NA	0	0
61	Educational Services	Education	0	0
62	Health Care and Social Assistance	Healthcare	15	4
71	Arts, Entertainment, and Recreation	Arts and recreation	-80	-20
72	Accommodation and Food Services	Accommodation/food service	-80	-20
81	Other Services (except Public Administration)	Other services except government	-5	-1
92	Public Administration (not covered in economic census)	Government	0	0

Table II.3: Mapping of CBO shocks to NAICS 2-digits

Industry	Consumption shock	Only postponed?
Food, drink, alcohol and tobacco	0	NA
Clothing and footwear	-50	yes
Housing, heating, etc.	0	NA
Goods and services (furniture, etc.)	-80	yes
Transport - buying cars	-100	yes
Transport services and car use	-50	no
Recreation and culture - durables	-100	yes
Recreation and culture - games and pets	0	NA
Recreation and culture - sport and culture	-100	no
Recreation and culture - newspapers and books	0	NA
Restaurants, hotels and net tourism	-100	no
Miscellaneous (incl health, communication education)	0	NA

Table II.4: Demand shock from Keogh-Brown et al. (2010). The first column gives the percentage decrease, while the second column gives the percentage of the first column which will be recouped in future quarters.

Category	Shock (%)
ISIC.Rev4 shock from OECD (2020)	
Manufacturing of transport equipment (29-30)	-100
Construction (VF)	-50
Wholesale and retail trade (VG)	-75
Air transport (V51)	-75
Accommodation and food services (VI)	-75
Real estate services excluding imputed rent (VL-V68A)	-75
Professional service activities (VM)	-50
Arts, entertainment and recreation (VR)	-75
Other service activities (VS)	-100
COICOP shock from OECD (2020)	
Clothing and footwear (3)	-100
Furnishings and household equipment (5)	-100
Vehicle purchases (7.1)	-100
Operation of private vehicles (7.2)	-50
Transport services (7.3)	-50
Recreation and culture excluding package holidays (9.1-9.5)	-75
Package holidays (9.6)	-100
Hotels and restaurants (11)	-75
Personal care services (12.1)	-100
Consumption shocks from Muellbauer (2020)	
Restaurants and Hotels	-71
Transport services	-70
Recreation services	-63
Food at home	43
Healthcare	18

Table II.5: Estimates of consumption shocks from various sources.

Bibliography

Abbe, E. & Sandon, C. (2015), Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery, *in* '2015 IEEE 56th Annual Symposium on Foundations of Computer Science', IEEE, pp. 670–688.

Acemoglu, D. & Autor, D. (2011), Skills, tasks and technologies: Implications for employment and earnings, *in* 'Handbook of Labor Economics', Vol. 4, Elsevier, pp. 1043–1171.

Acemoglu, D. & Restrepo, P. (2018), 'Artificial intelligence, automation and work', *NBER Working Paper No. w24196.* . Available at SSRN: https://ssrn.com/abstract=3101994.

Acemoglu, D. & Restrepo, P. (n.d.), 'Robots and jobs: Evidence from us labor markets', *Journal of Political Economy* **forthcoming**.

Adams-Prassl, A., Boneva, T., Golin, M. & Rauh, C. (2020), 'Inequality in the impact of the coronavirus shock: Evidence from real time surveys'.

Alabdulkareem, A., Frank, M. R., Sun, L., AlShebli, B., Hidalgo, C. & Rahwan, I. (2018), 'Unpacking the polarization of workplace skills', *Science Advances* **4**(7), eaao6030.

Alan, M. (2011), Imperfect competition in the labor market, *in* 'Handbook of labor economics', Vol. 4, Elsevier, pp. 973–1041.

Aleta, A. & Moreno, Y. (2019), 'Multilayer networks in a nutshell', *Annual Reviews of Condensed Matter Physics* **10**, 45–62.

Anand, K., Gai, P., Kapadia, S., Brennan, S. & Willison, M. (2013), 'A network model of financial system resilience', *Journal of Economic Behavior & Organization* **85**, 219–235.

Aoki, K., Benigno, G., Kiyotaki, N. et al. (2010), 'Adjusting to capital account liberalization', *CEPR Discussion Paper No. DP8087* . Available at SSRN: https://ssrn.com/abstract=1714859.

Arntz, M., Gregory, T. & Zierahn, U. (2016), The risk of automation for jobs in oecd countries: A comparative analysis, Technical Report 189, Organisation for Economic Cooperation and Development (OECD).

Arpaia, A., Kiss, A. & Turrini, A. (2014), Is unemployment structural or cyclical? main features of job matching in the eu after the crisis, Technical report, IZA Policy Paper.

Arrow, K. J. (1986), 'Rationality of self and others in an economic system', *Journal of Business* pp. S385–S399.

Arthur, W. B. (2014), Economic complexity: A different way to look at the economy, *in* 'Foundations & Frontiers of Complexity', Vol. 28, Santa Fe Institute for Complex Systems.

Arthur, W. B. (2021), 'Foundations of complexity economics', *Nature Reviews Physics* pp. 1–10.

Autor, D. (2013), 'The "task approach" to labor markets: an overview', *Journal for Labour Market Research* **46**(3), 185–199.

Autor, D. (2015*a*), 'Why are there still so many jobs? the history and future of workplace automation', *Journal of Economic Perspectives* **29**(3), 3–30.

Autor, D. (2015*b*), 'Why are there still so many jobs? the history and future of workplace automation', *Journal of Economic Perspectives* **29**(3), 3–30.

Autor, D. H., Levy, F. & Murnane, R. J. (2003), 'The skill content of recent technological change: An empirical exploration', *The Quarterly Journal of Economics* **118**(4), 1279–1333.

Autor, D., Katz, L. F. & Kearney, M. S. (2006), 'The polarization of the us labor market', *American Economic Review* **96**(2), 189–194.

Axtell, R., Guerrero, O. & López, E. (2016), 'The network composition of aggregate unemployment', *Saïd Business School WP 2016-02* . Available at SSRN: https://ssrn.com/abstract=2718244 or http://dx.doi.org/10.2139/ssrn.2718244.

Axtell, R. L., Guerrero, O. A. & López, E. (2019), 'Frictional unemployment on labor flow networks', *Journal of Economic Behavior & Organization* **160**, 184–201.

Bachmann, R. & Frings, H. (2017), 'Monopsonistic competition, low-wage labour markets, and minimum wages–an empirical analysis', *Applied Economics* **49**(51), 5268–5286.

Baker, S. R., Farrokhnia, R., Meyer, S., Pagel, M. & Yannelis, C. (2020), 'How does household spending respond to an epidemic? Consumption during the covid-19 pandemic.', *NBER WP 26949* .

Baldwin, R. & Weder di Mauro, B. (2020), Introduction, *in* R. Baldwin & B. Weder di Mauro, eds, 'Economics in the Time of COVID-19', CEPR, London, chapter 1, pp. 1–30.

Bank for International Settlements (2018), 'International banking statistics'. Available at https://www.bis.org/statistics/rppb1807.htm.

Bardoscia, M., Battiston, S., Caccioli, F. & Caldarelli, G. (2015), 'DebtRank: A microscopic foundation for shock propagation', *PloS ONE* **10**(6), e0130406.

Bardoscia, M., Caccioli, F., Perotti, J. I., Vivaldo, G. & Caldarelli, G. (2016), 'Distress propagation in complex networks: the case of non-linear DebtRank', *PloS ONE* **11**(10), e0163825.

Bargigli, L., Di Iasio, G., Infante, L., Lillo, F. & Pierobon, F. (2015), 'The multiplex structure of interbank networks', *Quantitative Finance* **15**(4), 673–691.

Barrot, J.-N., Grassi, B. & Sauvagnat, J. (2020), 'Sectoral effects of social distancing', *Covid Economics 3, 10 April 2020: 85-102* .

Bartik, A. W., Cullen, Z. B., Glaeser, E. L., Luca, M. & Stanton, C. T. (2020), What jobs are being done at home during the covid-19 crisis? evidence from firm-level surveys, Technical report, National Bureau of Economic Research.

Basel Committee on Banking Supervision (2019), 'Minimum capital requirements for market risk'. Available at https://www.bis.org/bcbs/publ/d457.htm.

Battiston, S., Farmer, J. D., Flache, A., Garlaschelli, D., Haldane, A. G., Heesterbeek, H., Hommes, C., Jaeger, C., May, R. & Scheffer, M. (2016), 'Complexity theory and financial regulation', *Science* **351**(6275), 818–819.

Battiston, S., Gatti, D. D., Gallegati, M., Greenwald, B. & Stiglitz, J. E. (2012), 'Default cascades: When does risk diversification increase stability?', *Journal of Financial Stability* **8**(3), 138–149.

Battiston, S., Puliga, M., Kaushik, R., Tasca, P. & Caldarelli, G. (2012), 'DebtRank: Too central to fail? Financial networks, the FED and systemic risk', *Scientific Reports* **2**, 541.

Becker, C., Noone, C. et al. (2009), Volatility in international capital movements, Technical report, Reserve Bank of Australia. Available at https://www.rba.gov.au/publications/rdp/2009/pdf/rdp2009-09.pdf.

Beveridge, W. H. (1944), *Full Employment in a Free Society (Works of William H. Beveridge): A Report*, W.W. Norton and Company, New York.

Bianconi, G. (2018), *Multilayer networks: structure and function*, Oxford university press.

Bonaccorsi, G., Riccaboni, M., Fagiolo, G. & Santoni, G. (2019), 'Country centrality in the international multiplex network', *Applied Network Science* **4**(1), 126.

Boone, L. (2020), Tackling the fallout from covid-19, *in* R. Baldwin & B. Weder di Mauro, eds, 'Economics in the time of COVID-19', CEPR, London, chapter 2, pp. 37–44.

Bravo, C. & Óskarsdóttir, M. (2020), 'Evolution of credit risk using a personalized pagerank algorithm for multilayer networks', *arXiv preprint arXiv:2005.12418* .

Brown, C. (2007), *Differential equations: A modeling approach*, number 150, Sage.

Brummitt, C. D. & Kobayashi, T. (2015), 'Cascades in multiplex financial networks with debts of different seniority', *Physical Review E* **91**(6), 062813.

Brummitt, C. D., Lee, K.-M. & Goh, K.-I. (2012), 'Multiplexity-facilitated cascades in networks', *Physical Review E* **85**(4), 045102.

Brynjolfsson, E. & Mitchell, T. (2017), 'What can machine learning do? workforce implications', *Science* **358**(6370), 1530–1534.

Brynjolfsson, E., Mitchell, T. & Rock, D. (2018), What can machines learn, and what does it mean for occupations and the economy?, *in* 'AEA Papers and Proceedings', Vol. 108, pp. 43–47.

Brynjolfsson, E., Rock, D. & Syverson, C. (2017), Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics, *in* 'Economics of Artificial Intelligence', University of Chicago Press.

Bullard, J. (2020), 'Expected US macroeconomic performance during the pandemic adjustment period', *VoxEU.org, 04 April* . https://voxeu.org/article/expected-us-macroeconomic-performance-during-pandemic-adjustment-period.

Caccioli, F., Barucca, P. & Kobayashi, T. (2018), 'Network models of financial systemic risk: A review', *Journal of Computational Social Science* **1**(1), 81–114.

Caccioli, F., Shrestha, M., Moore, C. & Farmer, J. D. (2014), 'Stability analysis of financial contagion due to overlapping portfolios', *Journal of Banking & Finance* **46**, 233–245.

Carvalho, V. M., Hansen, S., Ortiz, A., Garcia, J. R., Rodrigo, T., Rodriguez Mora, S. & Ruiz de Aguirre, P. (2020), 'Tracking the covid-19 crisis with high-resolution transaction data'.

Chen, H., Qian, W. & Wen, Q. (2020), 'The impact of the covid-19 pandemic on consumption: Learning from high frequency transaction data', *Available at SSRN 3568574* .

Chen, W.-C., Huang, A. S., Chuang, J.-H., Chiu, C.-C. & Kuo, H.-S. (2011), 'Social and economic impact of school closure resulting from pandemic influenza a/h1n1', *Journal of Infection* **62**(3), 200–203.

Chetty, R., Friedman, J. N., Hendren, N., Stepner, M. et al. (2020), How did covid-19 and stabilization policies affect spending and employment? a new real-time economic tracker based on private sector data, Technical report, National Bureau of Economic Research.

Claessens, S., Dooley, M. P. & Warner, A. (1995), 'Portfolio capital flows: hot or cold?', *The World Bank Economic Review* **9**(1), 153–174.

Congressional Budget Office (2006), 'Potential influenza pandemic: Possible macroeconomic effects and policy issues'. https://www.cbo.gov/sites/default/files/109th-congress-2005-2006/reports/12-08-birdflu.pdf.

Dachs, B., Hud, M., Koehler, C. & Peters, B. (2017), 'Innovation, creative destruction and structural change: firm-level evidence from european countries', *Industry and Innovation* **24**(4), 346–381.

de Domenico, M., Granell, C., Porter, M. A. & Arenas, A. (2016), 'The physics of spreading processes in multilayer networks', *Nature Physics* **12**, 901–906.

De Domenico, M., Solé-Ribalta, A., Omodei, E., Gómez, S. & Arenas, A. (2013), 'Centrality in interconnected multilayer networks', *arXiv preprint arXiv:1311.2906* .

De Haas, R. & Van Horen, N. (2012), 'International shock transmission after the Lehman Brothers collapse: Evidence from syndicated lending', *American Economic Review* **102**(3), 231–37.

del Rio-Chanona, R. M., Mealy, P., Pichler, A., Lafond, F. & Farmer, J. D. (2020), 'Supply and demand shocks in the covid-19 pandemic: An industry and occupation perspective', *Oxford Review of Economic Policy* **36**(Supplement_1), S94–S137.

Diamond, P. (2013), 'Cyclical unemployment, structural unemployment', *IMF Economic Review* **61**(3), 410–455.

Diamond, P. A. (1982), 'Aggregate demand management in search equilibrium', *Journal of Political Economy* **90**(5), 881–894.

Diamond, P. A. & Şahin, A. (2015), 'Shifts in the beveridge curve', *Research in Economics* **69**(1), 18–25.

Diem, C., Pichler, A. & Thurner, S. (2019), 'What is the minimal systemic risk in financial exposure networks?', *Journal of Economic Dynamics and Control* (2019-03), 1–25.

Dingel, J. & Neiman, B. (2020), 'How many jobs can be done at home?', *Covid Economics* **1**.

Diodato, D. & Weterings, A. B. (2014), 'The resilience of regional labour markets to economic shocks: Exploring the role of interactions among firms and workers', *Journal of Economic Geography* **15**(4), 723–742.

Dworkin, J. D. (2019), 'Network-driven differences in mobility and optimal transitions among automatable jobs', *Royal Society Open Science* **6**(7), 182124.

Eeckhout, J. & Lindenlaub, I. (2019), 'Unemployment cycles', *American Economic Journal: Macroeconomics* **11**(4), 175–234.

Eichenbaum, M. S., de Matos, M. G., Lima, F., Rebelo, S. & Trabandt, M. (2020), How do people respond to small probability events with large, negative consequences?, Technical report, National Bureau of Economic Research.

Eisenberg, L. & Noe, T. H. (2019), 'Systemic risk in financial systems', *Management Science* **47**(2), 236–249.

Elsby, M. W., Michaels, R. & Ratner, D. (2015), 'The beveridge curve: A survey', *Journal of Economic Literature* **53**(3), 571–630.

Fana, M., Tolan, S., Torrejón, S., Brancati, C. U. & Fernández-Macías, E. (2020), *The COVID confinement measures and EU labour markets*, Publications Office of the European Union Luxembourg.

Farhi, E. & Tirole, J. (2018), 'Deadly embrace: Sovereign and financial balance sheets doom loops', *The Review of Economic Studies* **85**(3), 1781–1823.

Farmer, J. D. & Foley, D. (2009), 'The economy needs agent-based modelling', *Nature* **460**(7256), 685–686.

Farmer, J. D., Kleinnijenhuis, A. M., Nahai-Williamson, P. & Wetzer, T. (2020), Foundations of system-wide financial stress testing with heterogeneous institutions, Technical report, Bank of England Working Paper. Available at https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2020/foundations-of-system-wide-financial-stress-testing-with-heterogeneous-institut pdf.

Farmer, J. D. et al. (2012), Economics needs to treat the economy as a complex system, *in* 'Paper for the INET Conference 'Rethinking Economics and Politics', Vol. 14.

Feinstein, Z. & Sojmark, A. (2019), 'A dynamic default contagion model: From Eisenberg–Noe to the mean field', *arXiv preprint arXiv:1912.08695* .

Fernando, C. S., May, A. D. & Megginson, W. L. (2012), 'The value of investment banking relationships: Evidence from the collapse of Lehman Brothers', *The Journal of Finance* **67**(1), 235–270.

Flood, S., King, M., Rodgers, R., Ruggles, S. & Warren, J. R. (2020), 'Integrated public use microdata series, current population: Version 8.0 [dataset]'. Minneapolis, MN: IPUMS, 2020., https://doi.org/10.18128/D030.V8.0.

Forbes, K. (2014), 'Financial 'deglobalization'? Capital flows, banks and the Beatles', *Speech to Queen Mary University, London* **18**. Available at https://www.bankofengland.co.uk/speech/2014/financial-deglobalization-capital-flows-banks-and-the-beatles.

Forbes, K., Reinhardt, D. & Wieladek, T. (2017), 'The spillovers, interactions, and (un) intended consequences of monetary and regulatory policies', *Journal of Monetary Economics* **85**, 1–22.

Fortunato, S. & Hric, D. (2016), 'Community detection in networks: A user guide', *Physics reports* **659**, 1–44.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E. et al. (2019), 'Toward understanding the impact of artificial intelligence on labor', *Proceedings of the National Academy of Sciences* p. 201900949.

Frey, C. B. & Osborne, M. A. (2017), 'The future of employment: how susceptible are jobs to computerisation?', *Technological Forecasting and Social Change* **114**, 254–280.

Gai, P. & Kapadia, S. (2010), Contagion in financial networks, *in* 'Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences', Vol. 466, The Royal Society, pp. 2401–2423.

Garrett-Peltier, H. (2017), 'Green versus brown: Comparing the employment impacts of energy efficiency, renewable energy, and fossil fuels using an input-output model', *Economic Modelling* **61**, 439–447.

Gleeson, J. P. (2013), 'Binary-state dynamics on complex networks: Pair approximation and beyond', *Physical Review X* **3**(2), 021004.

Gleich, D. F. (2015), 'PageRank beyond the Web', *SIAM Rev.* **57**(3), 321–363.

Goodhart, C. & Pradhan, M. (2020), 'Future imperfect after coronavirus', *VoxEU.org, 27 march* . https://voxeu.org/article/future-imperfect-after-coronavirus.

Gray, D. F. (2013), Modeling banking, sovereign, and macro risk in a cca global var, Technical Report WP/13/218.

Greenstone, M., Hornbeck, R. & Moretti, E. (2010), 'Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings', *Journal of Political Economy* **118**(3), 536–598.

Greenwood, R., Landier, A. & Thesmar, D. (2015), 'Vulnerable banks', *Journal of Financial Economics* **115**(3), 471–485.

Groes, F., Kircher, P. & Manovskii, I. (2014), 'The u-shapes of occupational mobility', *The Review of Economic Studies* **82**(2), 659–692.

Guerrero, O. A. & Axtell, R. L. (2013), 'Employment growth through labor flow networks', *PloS one* **8**(5), e60808.

Guerrieri, V., Lorenzoni, G., Straub, L. & Werning, I. (2020), Macroeconomic implications of covid-19: Can negative supply shocks cause demand shortages?, Technical report, National Bureau of Economic Research.

Haldane, A. G. & May, R. M. (2011), 'Systemic risk in banking ecosystems', *Nature* **469**(7330), 351–355.

Hall, R. E. (1979), 'A theory of the natural unemployment rate and the duration of employment', *Journal of monetary economics* **5**(2), 153–169.

Haltiwanger, J. C., Hyatt, H. R., Kahn, L. B. & McEntarfer, E. (2018), 'Cyclical job ladders by firm size and firm wage', *American Economic Journal: Macroeconomics* **10**(2), 52–85.

Hicks, M. J., Faulk, D. & Devaraj, S. (2020), 'Occupational exposure to social distancing: A preliminary analysis using o* net data', *Center for Business and Economic Research* . https://projects.cberdata.org/170/social-distance.

Inoue, H. & Todo, Y. (2020), 'The propagation of the economic impact through supply chains: The case of a mega-city lockdown against the spread of covid-19', *Covid Economics* **2**.

International Monetary Fund (2018*a*), 'Coordinated direct investment survey'. Available at https://data.imf.org/?sk=40313609-F037-48C1-84B1-E1F1CE54D6D5.

International Monetary Fund (2018*b*), 'Coordinated portfolio investment survey'. Available at https://data.imf.org/?sk=B981B4E3-4E58-467E-9B90-9DE0C3367363.

International Monetary Fund (2018*c*), 'Financial soundness indicators'. Available at https://data.imf.org/?sk=51B096FA-2CD2-40C2-8D09-0699CC1764DA.

Jeub, L. G., Mahoney, M. W., Mucha, P. J. & Porter, M. A. (2017), 'A local perspective on community structure in multilayer networks', *Network Science* **5**(2), 144–163.

Kahneman, D. (2011), *Thinking, fast and slow*, Macmillan.

Kaushik, R. & Battiston, S. (2013), 'Credit default swaps drawup networks: too interconnected to be stable?', *PloS one* **8**(7).

Keogh-Brown, M. R., Wren-Lewis, S., Edmunds, W. J., Beutels, P. & Smith, R. D. (2010), 'The possible macroeconomic impact on the uk of an influenza pandemic', *Health Economics* **19**(11), 1345–1360.

Kivelä, M. (2014), 'PYMNET: Multilayer networks library for Python'. Available at http://www.mkivela.com/pymnet.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y. & Porter, M. A. (2014), 'Multilayer networks', *Journal of complex networks* **2**(3), 203–271.

Kohlbrecher, B. & Merkl, C. (2016), 'Business cycle asymmetries and the labor market', *CESifo Working Paper Series* . Available at SSRN: https://ssrn.com/abstract=2764118.

Koren, M. & Pető, R. (2020), 'Business disruptions from social distancing', *Covid Economics* **2**.

Korniyenko, Y., Patnam, M., del Rio-Chanona, R. M. & Porter, M. A. (2018), Evolution of the global financial network and contagion: A new approach, Technical report, International Monetary Fund.

Krüger, J. J. (2008), 'Productivity and structural change: a review of the literature', *Journal of Economic Surveys* **22**(2), 330–363.

La Guardia, G. G. & Miranda, P. J. (2018), 'Lyapunov exponent for lipschitz maps', *Nonlinear Dynamics* **92**(3), 1217–1224.

Lane, P. R. R. & Milesi-Ferretti, G. M. M. (2017), International financial integration in the aftermath of the global financial crisis, Technical Report Working Paper, WP/17/115, International Monetary Fund.

Lazear, E. P. (1990), 'Job security provisions and employment', *The Quarterly Journal of Economics* **105**(3), 699–726.

Lee, K.-M., Brummitt, C. D. & Goh, K.-I. (2014), 'Threshold cascades with response heterogeneity in multiplex networks', *Physical Review E* **90**(6), 062816.

Lempel, H., Epstein, J. M. & Hammond, R. A. (2009), 'Economic cost and health care workforce effects of school closures in the us', *PLoS currents* **1**.

Leon, C., Berndsen, R. & Renneboog, L. (2014), 'Financial stability and interacting networks of financial institutions and market infrastructures', *Borradores de Economia* (848). Report available at https://econpapers.repec.org/paper/bdrborrec/848.htm.

Li, Y., Liu, G. & Pin, P. (2018), 'Network-based risk measurements for interbank systems', *PloS ONE* **13**(7), e0200209.

Lin, J. (2011), 'Technological adaptation, cities, and new work', *Review of Economics and Statistics* **93**(2), 554–574.

Lund, S., Daruvala, T., Dobbs, R., Härle, P., Kwek, J.-H. & Falcón, R. (2013), 'Financial globalization: Retreat or reset', *McKinsey Global Institute* p. 2. Report available at https://www.mckinsey.com/featured-insights/employment-and-growth/financial-globalization.

Mallaby, S. (2016), 'Globalization resets', *Finance and Development* **53**(4), 6–10. Available at http://www.contexto.org/pdfs/2016/IMFgloblwinersloosers.pdf.

Mann, C. (2020), Real and financial lenses to assess the economic consequences of COVID-19, *in* R. Baldwin & B. Weder di Mauro, eds, 'Economics in the time of COVID-19', CEPR, London, chapter 8, pp. 81–86.

Markose, S. M. (2012), Systemic risk from global financial derivatives: A network analysis of contagion and its mitigation with super-spreader tax, Technical report, International Monetary Fund.

McKibbin, W. J. & Fernando, R. (2020), 'The global macroeconomic impacts of covid-19: Seven scenarios'.

McKibbin, W. J. & Sidorenko, A. A. (2006), 'Global macroeconomic consequences of pandemic influenza', *Lowy Institute for International Policy Sydney, Australia* .

Mealy, P., del Rio-Chanona, R. M. & Farmer, J. D. (2018), 'What you do at work matters: New lenses on labour', *SSRN 3143064* . Available at SSRN: https://ssrn.com/abstract=3143064 or http://dx.doi.org/10.2139/ssrn.314306.

Meyn, S. P. & Tweedie, R. L. (2012), *Markov chains and stochastic stability*, Springer Science & Business Media.

Miles, D. & Scott, A. (2020), 'Will inflation make a comeback after the crisis ends?', *VoxEU.org, 04 April* . https://voxeu.org/article/will-inflation-make-comeback-after-crisis-ends.

Mokyr, J., Vickers, C. & Ziebarth, N. L. (2015), 'The history of technological anxiety and the future of economic growth: Is this time different?', *Journal of economic perspectives* **29**(3), 31–50.

Moretti, E. (2010), Local labor markets, Technical report, National Bureau of Economic Research.

Mortensen, D. T. (1982), The matching process as a noncooperative bargaining game, *in* 'The economics of information and uncertainty', University of Chicago Press, pp. 233–258.

Mortensen, D. T. (1999), 'Equilibrium unemployment dynamics', *International Economic Review* **40**(4), 889–914.

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A. & Onnela, J.-P. (2010), 'Community structure in time-dependent, multiscale, and multiplex networks', *science* **328**(5980), 876–878.

Muellbauer, J. (2020), 'The coronavirus pandemic and U.S. consumption', *VoxEU.org, 11 April* . https://voxeu.org/article/coronavirus-pandemic-and-us-consumption.

Nedelkoska, L., Diodato, D., Neffke, F. et al. (2018), Is our human capital general enough to withstand the current wave of technological change?, Technical report, Center for International Development at Harvard University.

Neffke, F. M., Otto, A. & Hidalgo, C. (2018), 'The mobility of displaced workers: How the local industry mix affects job search', *Journal of Urban Economics* **108**, 124–140.

Neffke, F. M., Otto, A. & Weyh, A. (2017), 'Inter-industry labor flows', *Journal of Economic Behavior & Organization* **142**, 275–292.

Newman, M. (2018*a*), *Networks*, Oxford university press.

Newman, M. E. (2006), 'Modularity and community structure in networks', *Proceedings of the national academy of sciences* **103**(23), 8577–8582.

Newman, M. E. J. (2018*b*), *Networks*, second edn, Oxford University Press, Oxford, UK.

Nimczik, J. S. (2017), Job mobility networks and endogenous labor markets, Annual conference 2017 (vienna): Alternative structures for money and banking, Verein für Socialpolitik / German Economic Association. Available at: https://EconPapers.repec.org/RePEc:zbw:vfsc17:168147.

OECD (2020), 'Evaluating the initial impact of COVID-19 containment measures on economic activity', *OECD Technical Report, March 2020* . https://read.oecd-ilibrary.org/view/?ref=126_126496-evgsi2gmqj&title=Evaluating_the_initial_impact_of_COVID-19_containment_measures_on_economic_activity.

Office for National Statistics (2020), 'Coronavirus and homeworking in the UK labour market: 2019', *ONS Technical Report* . https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/coronavirusandhomeworkingintheuklabourmarket/2019.

Okun, A. & Potential, G. (1962), Its measurement and significance, american statistical association, *in* 'Proceedings of the Business and Economics Statistics Section', pp. 98–104.

Pangallo, M., Heinrich, T. & Farmer, J. D. (2019), 'Best reply structure and equilibrium convergence in generic games', *Science advances* **5**(2), eaat1328.

Petrone, D. & Latora, V. (2018), 'A dynamic approach merging network theory and credit risk techniques to assess systemic risk in financial networks', *Scientific reports* **8**(1), 1–14.

Petrongolo, B. & Pissarides, C. A. (2001), 'Looking into the black box: A survey of the matching function', *Journal of Economic Literature* **39**(2), 390–431.

Pichler, A., Lafond, F. & Farmer, J. D. (2019), 'The aak paper revisited – i will work this out – i promise!', *xxx* **0**, 0.

Pichler, A., Pangallo, M., del Rio-Chanona, R. M., Lafond, F. & Farmer, J. D. (2020), 'Production networks and epidemic spreading: How to restart the uk economy?', *Production Networks and Epidemic Spreading: How to Restart the UK Economy* .

Pichler, A., Poledna, S. & Thurner, S. (2018), 'Systemic-risk-efficient asset allocation: Minimization of systemic risk as a network optimization problem', *arXiv preprint arXiv:1801.10515* .

Pissarides, C. A. (1979), 'Job matchings with state employment agencies and random search', *The Economic Journal* **89**(356), 818–833.

Pissarides, C. A. (1985), 'Short-run equilibrium dynamics of unemployment, vacancies, and real wages', *The American Economic Review* **75**(4), 676–690.

Pissarides, C. A. (2011), 'Equilibrium in the labor market with search frictions', *American Economic Review* **101**(4), 1092–1105.

Poledna, S., Hinteregger, A. & Thurner, S. (2018), 'Identifying systemically important companies in the entire liability network of a small open economy', *Entropy* **20**(10), 792.

Poledna, S., Molina-Borboa, J. L., Martínez-Jaramillo, S., Van Der Leij, M. & Thurner, S. (2015), 'The multi-layer network nature of systemic risk and its implications for the costs of financial crises', *Journal of Financial Stability* **20**, 70–81.

Porter, M. A. (2018), 'WHAT IS... a multilayer network?', *Notices of the AMS* **65**(11), 1419–1423.

Porter, M. A. & Gleeson, J. P. (2016), *Dynamical systems on networks: A tutorial*, Vol. 4 of *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*, Springer International Publishing, Cham, Switzerland.

Rahmede, C., Iacovacci, J., Arenas, A. & Bianconi, G. (2018), 'Centralities of nodes and influences of layers in large multiplex networks', *Journal of Complex Networks* **6**(5), 733–752.

Ramey, V. A. & Francis, N. (2009), 'A century of work and leisure', *American Economic Journal: Macroeconomics* **1**(2), 189–224.

Restrepo, P. (2015), 'Skill mismatch and structural unemployment', *Massachusetts Institute of Technology Job Market Paper* pp. 1–94.

Rogerson, R., Shimer, R. & Wright, R. (2005), 'Search-theoretic models of the labor market: A survey', *Journal of economic literature* **43**(4), 959–988.

Ross, S. M., Kelly, J. J., Sullivan, R. J., Perry, W. J., Mercer, D., Davis, R. M., Washburn, T. D., Sager, E. V., Boyce, J. B. & Bristow, V. L. (1996), *Stochastic processes*, Vol. 2, Wiley New York.

Rosvall, M., Axelsson, D. & Bergstrom, C. T. (2009), 'The map equation', *The European Physical Journal Special Topics* **178**(1), 13–23.

Santos, J. R., May, L. & Haimar, A. E. (2013), 'Risk-based input-output analysis of influenza epidemic consequences on interdependent workforce sectors', *Risk Analysis* **33**(9), 1620–1635.

Schmutte, I. M. (2014), 'Free to move? a network analytic approach for learning the limits to job mobility', *Labour Economics* **29**, 49–61.

Schumpeter, J. (1942), 'Creative destruction', *Capitalism, socialism and democracy* **825**, 82–85.

Schumpeter, J. A. (1962), *Capitalism, Socialism, and Democracy.*, Harper and Row.

Serfozo, R. (2009), *Basics of applied stochastic processes*, Springer Science & Business Media.

Serrano, M. Á. & Boguñá, M. (2005), Weighted configuration model, *in* 'AIP Conference Proceedings', Vol. 776, American Institute of Physics, pp. 101–107.

Silva, T. C., da Silva Alexandre, M. & Tabak, B. M. (2018), 'Bank lending and systemic risk: A financial-real sector network approach with feedback', *Journal of Financial Stability* **38**, 98–118.

Sniekers, F. (2018), 'Persistence and volatility of beveridge cycles', *International Economic Review* **59**(2), 665–698.

Starnini, M., Boguñá, M. & Serrano, M. Á. (2019), 'The interconnected wealth of nations: Shock propagation on global trade-investment multiplex networks', *Scientific reports* **9**(1), 1–10.

Stevens, M. (2007), 'New microfoundations for the aggregate matching function', *International Economic Review* **48**(3), 847–868.

Stoneman, P. (2001), *The economics of technological diffusion*, Wiley-Blackwell.

Tebogo, B. (2012), 'The failure and collapse of the Lehman Brothers'. Available at SSRN at https://ssrn.com/abstract=2060758.

Thaler, R. H. & Sunstein, C. R. (2009), *Nudge: Improving decisions about health, wealth, and happiness*, Penguin.

Turrell, A., Speigner, B., Djumalieva, J., Copple, D. & Thurgood, J. (2018), Using job vacancies to understand the effects of labour market mismatch on uk output and productivity, Technical report, Bank of England Working Paper.

van Lidth de Jeude, J. A. (2019), 'Analysing mesoscale structures in economic and financial networks'. PhD. thesis available at http://e-theses.imtlucca.it/id/eprint/266.

Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P. G., Fu, H. et al. (2020), 'Estimates of the severity of coronavirus disease 2019: a model-based analysis', *The Lancet Infectious Diseases* .

Vermeulen, B., Kesselhut, J., Pyka, A. & Saviotti, P. P. (2018), 'The impact of automation on employment: Just the usual structural change?', *Sustainability* **10**(5), 1661.

Vona, F., Marin, G., Consoli, D. & Popp, D. (2018), 'Environmental regulation and green skills: an empirical exploration', *Journal of the Association of Environmental and Resource Economists* **5**(4), 713–753.

Wallis, J. J. (1989), 'Employment in the great depression: New data and hypotheses', *Explorations in Economic History* **26**, 45–72.

Watts, D. J. (2002), 'A simple model of global cascades on random networks', *Proceedings of the National Academy of Sciences of the United States of America* **99**(9), 5766–5771.

Wiersema, G., Kleinnijenhuis, A. M., Wetzer, T. & Farmer, J. D. (2019), Inherent instability: Scenario-free analysis of financial systems with interacting contagion channels, Technical report, INET Oxford Working Paper. Available at https://www.inet.ox.ac.uk/publications/no-2019-10-inherent-instability-scenario-free-analysis-of-financial-systems-wit

Yağan, O. & Gligor, V. (2012), 'Analysis of complex contagions in random multiplex networks', *Physical Review E* **86**(3), 036103.

Zhang, S. X., Wang, Y., Rauch, A. & Wei, F. (2020), 'Health, distress and life satisfaction of people in china one month into the covid-19 outbreak', *medRxiv preprint* . https://www.medrxiv.org/content/10.1101/2020.03.13.20034496v1.full.pdf.