An agent-based model of the labour market with applications to the post-carbon transition

Anna K. Berryman
Balliol College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary 2025

In loving memory of my Grandpa Martin and my godfather Keith.

Acknowledgements

I would like to thank the following people:

My supervisors; Maria del Rio-Chanona for her weekly guidance, from COVID lockdown meetings at Balliol to endless google meets, Doyne Farmer for his expert guidance, and Sam Cohen for his wisdom, mathematical and otherwise.

Everyone who made InFoMM, particularly Chris Breward, Colin Please, Cohort 6, and especially James and Brady, who survived the grind to the end beside me. The brilliant members of OCIAM who supported me through coffee, cake, and board games. My S-3.14 office mates throughout the years, particularly Andrew, Pavan, and Georgina. The colleagues over at INET, notably my co-authors Joris, Fernanda, Pete, and Penny. Dorothy and Sophie who were shining lights during my days at both offices.

Kathryn Pearson, without whose weekly study skills sessions I would not have finished the Brazil paper, or this thesis. Coralia Cartis who was always ready with a caring ear and the academic support I needed. My industrial supervisor at BEIS, James Foster, all those involved in the EEIST project for Brazil, China, and India, particularly those I met in Delhi, and Marek Hanusch and Shireen Mahdi at the World Bank.

Completing this DPhil would not have been possible without daily penguin jokes, countless tears, and the following people in my life:

Everyone at RMS, guiding me from a reckless 11 year old to a confident mathematician, particularly Mr Broome and the ‘Year 9 Maths Prize’.

My rowing teammates, and dear friends, at both BCBC and OUBC. Rowing was a necessary constant throughout my PhD and I will treasure all the memories we made on the river, and off; not least Spoons to Blades in Summer Eights, surviving a thunderstorm in Wallingford, and late nights singing at Trinity.

The community of Balliol MCR; I will cherish the times spent around Holywell Manor, the BOPs, formals, and Brett's yoga giving me one hour of peace a week.

My friends, both near and far, particularly Tom Grein, Balliol women Vicky, Minnie, and Sophie, and Moa, whose motivational post-it notes now adorn my computer screen and our endless catch-ups that continue to fill me with gratitude and love.

And finally, my beloved family. Polly who is always on the other end of FaceTime for a gossip or last-minute outfit check. Elizabeth, my older-younger sister whose library sessions, knowledge, and exasperation keep me a little more sane everyday. Harry, whose calm, caring confidence binds our quad together. My father, Jonathan, for his never-ending wisdom, competitiveness, and ambition, and my mother, Kate, for her patience, strength, and devotion, who didn't know what she was signing up for when I came to Oxford but has been unwavering in her support of everything I do; both of whom I am eternally grateful to for everything they have given me, and my siblings, to enable us to flourish into a pretty cool bunch.

I did it!

Abstract

Many countries around the world are facing major economy-wide transitions over the next few decades in order to meet the goals of the Paris Agreement. While much modelling has been done to understand the economic and environmental implications of different transition scenarios, many of these models do not account for labour market frictions that limit workers' mobility. We make a technical and empirical contribution to the question of how worker mobility plays a role in such economic transitions.

The technical contribution of this thesis is to extend a data-driven labour market model that accounts for labour mobility and second order frictions in the labour market. We extend the network-based model to include regional mobility, and also to include multiple job applications, on-the-job search, and wage driven dynamics. We present a deterministic approximation to the full model, notably for multiple job applications in such a model. This approximation enables fast evaluations of the model, which in turn enables exploration of the parameter space, and the use of detailed occupations and multiple regions.

Empirically, we visualise and analyse the network of occupational and regional mobility for Brazil for the first time. We then combine our model with two development pathways for Brazil to study how a productivity shift in either manufacturing or agriculture may affect occupation-level unemployment. We find that a poorly managed productivity shift could increase inequality.

The flexible, data-driven model framework we present can also be used to understand the labour market impacts of other net zero transition policies, as well as other economic transitions, such as automation.

Contents

1 Introduction **1**

1.1 Publications and presentations 4

1.2 Thesis overview 5

2 Literature review **6**

2.1 Historic environmental policy impacts on labour 6

2.2 Labour market modelling 7

2.3 Skills 9

2.4 Geography 11

2.5 Networks 13

2.6 Agent-based models 15

2.7 Definitions 17

3 Labour market model with geography **19**

3.1 Del Rio-Chanona model 20

3.1.1 Model definition 20

3.1.2 Deterministic approximation for large populations 24

3.1.3 Limitations 29

3.2 Regional occupational mobility 30

3.2.1 Toy model 40

3.3 Conclusions 46

4 Brazil case study **47**

4.1 Introduction 47

4.2 Methods 49

4.2.1 Transition scenarios 49

4.2.2 Unemployment and unfilled vacancy rates 53

4.2.3 Occupation-related versus region-related network effects 53

4.3 Results 54

4.3.1 Inequality of unemployment outcomes 54
4.3.2 Unfilled job vacancies risk slowing down development 60
4.4 Discussion 64
4.5 Conclusions 67

5 A triad of model extensions 68
5.1 Introduction 68
5.2 On-the-job search 69
5.2.1 Deterministic approximation for large populations 71
5.2.2 Toy model 75
5.3 Multiple job applications 80
5.3.1 Deterministic approximation for large populations 83
5.3.2 Toy model 87
5.4 Wage driven dynamics 89
5.5 Approximation tests 92
5.5.1 Non-empty vacancies 93
5.5.2 Accepted job offers 94
5.5.3 Applicants as constant in multiple applications 96
5.6 Conclusions 97

6 Automation case study 98
6.1 US network 98
6.1.1 CPS v. ASEC 99
6.2 Demand shock 102
6.3 Results 103
6.3.1 On-the-job search 104
6.3.2 Multiple job applications 105
6.3.3 Wage driven dynamics 106
6.4 Conclusions 107

7 Conclusions 109

Bibliography 113

List of Figures

3.1 **Regional occupational mobility network for Brazil.** In (a) the full regional occupational mobility network, and in (b) a map of Brazil with states coloured similarly to (a). 34

3.2 **National occupational mobility network for Brazil.** In (a) the national occupational mobility network, without disaggregation by state, and in (b) the full regional occupational mobility network, coloured by occupational group. 35

3.3 **Regional mobility network and gravity model.** In (a) the adjacency matrix for the regional mobility network of Brazil, and in (b) the matrix of relationships proposed by the gravity model. 38

3.4 **Demographics on the occupational mobility network.** Each node is coloured by (a) median wage, (b) proportion of workers who are female, and (c) median age from 2018. 39

3.5 **Toy model occupational mobility network.** The occupational mobility network for the toy model with two occupations. 41

3.6 **Toy model regional occupational mobility network.** In (a) the regional occupational mobility network for the toy model with two occupations and two regions, in (b) the network summed over occupations (and row-normalised), and in (c) the network summed over regions (and row-normalised). 41

3.7 **Target demand profiles for toy model for the five scenarios.** In (a) Universal 1 without regions then, in (b) Regional 1A,2A with regional heterogeneity, (c) Regional 1B,2B, and in (d) Split 1A,2B has both occupational and regional heterogeneity, as does (e) Split 1B,2A. 42

3.8 **Aggregate unemployment rate for Regional 1A,2A.** The unemployment rate for the Regional 1A,2A scenario with the full simulation and deterministic approximation. 46

4.1 **Occupation level labour demand change.** Labour demand change per occupational group for (a) the Manufacturing growth path, and (b) the Agriculture growth path. 51
4.2 **Demand change on the occupational mobility network.** 2030 demand change compared to the baseline for (a) the Manufacturing growth path, and (b) the Agriculture growth path. 52
4.3 **Brazil demand and unemployment rate.** In (a) the volume of labour reallocation (Definition 6) per year, and in (b) the aggregate unemployment rate between 2018 and 2030. 55
4.4 **Manufacturing growth path unemployment outcomes.** In (a) the percentage-point change of the 2018–2030 average unemployment rate from the baseline against percentage demand change compared to baseline in 2030, each data point represents an occupation-region pair, shaped by region and sized proportional to employment in 2018 with inset labels identifying different occupations, in (b) the percentage-point change of the 2018–2030 unemployment rate from the baseline for each occupational group and region, and in (c) the regional percentage-point change of the average unemployment rate from baseline. 56
4.5 **Agriculture growth path unemployment outcomes.** In (a) the percentage-point change of the 2018–2030 average unemployment rate from the baseline against percentage demand change compared to baseline in 2030, each data point represents an occupation-region pair, shaped by region and sized proportional to employment in 2018 with inset labels identifying different occupations, in (b) the percentage-point change of the 2018–2030 unemployment rate from the baseline for each occupational group and region, and in (c) the regional percentage-point change of the average unemployment rate from baseline. 57
4.6 **Variance decomposition.** Variance decomposition of the unemployment and 6-month unfilled vacancy rate outcomes for the Agriculture and Manufacturing growth paths. 60

4.7 **Manufacturing growth path vacancy outcomes.** In (a) the percentage-point change of the 2018–2030 average unfilled vacancy rate from the baseline against percentage demand change compared to baseline in 2030, each data point represents an occupation-region pair, shaped by region and sized proportional to employment in 2018 with inset labels identifying different occupations, in (b) the percentage-point change of the 2018–2030 unfilled vacancy rate from the baseline for each occupational group and region, and in (c) the regional percentage-point change of the average unfilled vacancy rate from baseline. 61

4.8 **Agriculture growth path vacancy outcomes.** In (a) the percentage-point change of the 2018–2030 average unfilled vacancy rate from the baseline against percentage demand change compared to baseline in 2030, each data point represents an occupation-region pair, shaped by region and sized proportional to employment in 2018 with inset labels identifying different occupations, in (b) the percentage-point change of the 2018–2030 unfilled vacancy rate from the baseline for each occupational group and region, and in (c) the regional percentage-point change of the average unfilled vacancy rate from baseline. 62

5.1 **On-the-job search toy model shock demand and unemployment rate.** In (a) the occupation level demand profiles for $\mathcal{L} = 500$, and in (b) the aggregate unemployment rate with, and without, on-the-job search. 76

5.2 **Simulations and approximation for the toy model with on-the-job search, $\mathcal{L} = 500$.** Unemployment rates of the deterministic approximation and for 10 runs of the full model. 77

5.3 **Simulations and approximation for the toy model with on-the-job search, $\mathcal{L} = 50{,}000$.** Unemployment rates of the deterministic approximation and for 10 runs of the full model. 78

5.4 **Simulations and approximation for the toy model with on-the-job search, $\mathcal{L} = 5{,}000{,}000$.** Unemployment rates of the deterministic approximation and for 10 runs of the full model. 79

5.5 **Flow chart of the full, extended agent-based model.** *Top*: transition pathway of a worker. *Bottom*: transition pathway of a job vacancy. 82
5.6 **Offer sending process schematic.** Each square represents an application sent to occupation j at time $t+1$, and each column corresponds to a worker. Blue squares in the right hand grid are successful applications. 84
5.7 **Multiple applications toy model unemployment rate.** The aggregate unemployment rate with, and without, multiple applications and on-the-job search. 87
5.8 **Simulations and approximations for the toy model with multiple applications.** Aggregate unemployment rate of the deterministic approximation and 10 runs of the full agent-based model for different values of $\mathcal{L}$. 88
5.9 **Random occupational mobility network for 10 occupations.** The size of the nodes is proportional to the size of initial employment. 92
5.10 **Simulation average for non-empty vacancies.** The number of non-empty vacancies, $M_{j,t+1}$ for (a) 100 runs, and (b) 200 runs of the full agent-based model with $\mathcal{L} = 100{,}000$ and $N = 10$, and in (c), the approximation error. 93
5.11 **Effect of λ, β_u, and β_e on non-empty vacancies.** The number of non-empty vacancies, $M_{j,t+1}$. 94
5.12 **Effect of λ, β_u, and β_e on accepted job offers.** The number of accepted job offers, $F_{j,t+1}$. 95
5.13 **Effect of occupancy distribution method on the number of job offers.** The number of accepted offers for (a) scaled normal density approximation to occupancy distribution (5.38), and (b) recursion algorithm using the reduced occupancy distribution (5.52) for the agent-based model with $\mathcal{L} = 100{,}000$ and $N = 10$, and in (c), the approximation error. 96
5.14 **Job offers accepted when applications are sampled.** The number of accepted job offers and the standard error of the approximation for 10, 100, and 1,000 samples of the multinomials $A^{(u)}_{ij,t+1}$ and $A^{(e)}_{ij,t+1}$ for $i \in 0, \ldots, 9$. 97

6.1 **Comparison of occupational mobility networks using CPS and ASEC data.** (a) The number of edges in the networks, and (b) The number of transitions recorded in each year, i.e. the sum of all the weights in the network for networks from 2003 to 2020 for CPS and 2003 to 2021, excluding 2014, for ASEC. 100
6.2 **Stacked bar plot for ASEC migration status.** Proportion of respondents with different migration status answers for 2010 to 2021 in ASEC. 101
6.3 **Histogram of diagonal entries of Brazil's National Occupational Mobility Network.** 102
6.4 **Unemployment rate with the approximation and 10 runs of the full agent-based model.** The unemployment rate as a function of time for the occupational mobility network and the Frey and Osborne automation shock for $\lambda = 0.001$, $\beta_u = 5$ and $\beta_e = 1$. . . 104
6.5 **Unemployment rate for on-the-job search.** The unemployment rate as a function of time for the Frey and Osborne automation shock for varying values of λ, the probability that an employed worker is looking for a job. 105
6.6 **Unemployment rate for multiple applications without mobility frictions.** The unemployment rate as a function of time for the complete network and Frey and Osborne automation shock for varying values of β_u, the number of applications unemployed workers send. 106
6.7 **Unemployment rate for multiple applications with mobility frictions.** The unemployment rate as a function of time for the occupational mobility network and the Frey and Osborne automation shock for varying values of β_u, number of applications unemployed workers send. 107
6.8 **Effect of adding wage dynamics on the percentage change in the unemployment rate per occupation.** The difference between the percentage change in the unemployment rate without and with wage pressure. The size of the dot is proportional to the in-degree in each network and the colour represents the log of the median wage. . 108

List of Tables

3.1 **Community detection.** Allocation of regions within the clusters with size over 100, and the remaining counts in 'Other'. 37

3.2 **Unemployment rate change for toy model at the occupation level.** Percentage change of the average occupation level unemployment rate. 45

3.3 **Unemployment rate change for toy model at the occupation-region level.** Percentage change of the average occupation-region level unemployment rate. 45

4.1 **Aggregate outcomes.** Average unemployment and unfilled vacancy rate change compared to the baseline between 2018 and 2030 by occupational group. 63

4.2 **Summary table.** Top five affected occupation-regions. 65

5.1 **Comparison of average unemployment rate for on-the-job search.** Decreasing standard error and root mean squared (RMS) difference between the agent-based model (ABM) mean and the deterministic approximation (DA), for different values of $\mathcal{L}$. 80

5.2 **On-the-job search toy model time taken.** Time taken for one deterministic approximation and the average for 10 runs of the agent-based model (ABM), for different values of $\mathcal{L}$. 80

5.3 **Comparison of average unemployment rate for multiple applications.** Decreasing standard error and root mean squared (RMS) difference between the agent-based model (ABM) mean and the deterministic approximation (DA), for different values of $\mathcal{L}$. . . . 89

5.4 **Variables and parameters.** 91

Mathematical notations

Throughout, we use the following notation for the stochastic agent-based model:

- Upper-case letters denote occupation-region pair level stochastic variables, for example $U_{i\alpha;t}$ is the number of workers unemployed in occupation i and region α at time t.
- Lower-case letters denote conditional probabilities.
- Calligraphic letters denote parameters, e.g. the empirical mobility network $\mathcal{A}$.
- I represents the set of occupations, and R, the set of regions.
- $\mathbb{E}$ is the expected value of a random variable.

Specifically for the deterministic approximation to the full agent-based model, we use the following:

- A bar on lower-case letters denotes the occupation-region pair level expected value, e.g. $\bar{u}_{i\alpha;t}$ (see Section 3.1.2 for details).
- Lower-case bold letters denote the set of expected values for all pairs of occupations in I and regions in R, that is $\mathbf{u}_t = \{\bar{u}_{i\alpha;t}\}_{i\alpha \in I \times R}$.

Chapter 1

Introduction

The United Nations believes that ‘climate change is the defining issue of our time’ and if we are to achieve many of the net zero carbon goals set out by countries across the globe, we need to understand how a rapid post-carbon transition will affect every aspect of society. There is significant concern about the possible impact on employment, particularly for workers currently employed in emissions-intensive jobs [1, 2, 3].

Previous studies aiming to quantify the employment impact of environmental or emissions reduction policies often rely on models in equilibrium that assume labour to be unrealistically flexible [4, 5, 6]. There is also little regard given to the geographic concentration of employment in emissions-intensive industries, the challenges that these workers may face in transitioning into other lines of work, or how this transition may exacerbate inequality [7, 8, 9].

Arthur [10] argues that Complexity Economics is the way forward to address a number of these shortcomings; to understand how agents behave in many subfields of macroeconomics without assuming equilibrium, including the labour market. So, we extend the labour market model developed by del Rio-Chanona, Mealy, Beguerisse-Díaz, Lafond, and Farmer [11] to study the low carbon transition. Their agent-based model of the labour market uses a network and is designed to be able to investigate any external labour market shock.

The flexibility of the del Rio-Chanona et al. framework allows the agent-based model to be applied to different economic transitions and the deterministic approximation enables quick evaluations. This enables parameter exploration and calibration without the need for a scaled down population. The full agent-based model is described by a set of stochastic equations and processes. To calculate a deterministic approximation, we take the expectation of every stochastic process at each time step. This gives us a system that runs up to 20 times faster than the full agent-based model. We give due attention in this thesis to deriving this approximation as we add complexity to the model, so that it remains practical to use, especially for our analysis of detailed occupations.

The term agent-based model is used in a loose way to describe the del Rio-Chanona model. We acknowledge that the model does not fulfil the criteria of an agent-based model set out in the literature. Wooldridge and Jennings [12], in computer science, state that agents must have autonomy, social ability, reactivity, and pro-activeness. In social science, an agent-based model is defined as a model with heterogeneous, autonomous agents that interact with each other and within some social structure, and are boundedly rational [13, 14, 15]. For now, the stochastic model presented in this thesis does not fulfil these criteria however, it is able to provide insights into questions of labour mobility frictions and their interaction with demand scenarios and the framework is set up to be extended in these ways where further complexity can be added to the agents (workers). In this thesis, we use the term 'stochastic agent-based model' for the full stochastic model and 'deterministic approximation' for describe the deterministic system of equations that can be used to approximate the full model.

Moretti [16] labels the increase in the inequality of job opportunities as 'the Great Divergence' of US cities, emphasising how much one's geographic location influences one's employment prospects in the US. Therefore firstly, we introduce regional mobility into the del Rio-Chanona et al. model. Understanding migration between regions, and countries, has many lenses and, as better data becomes available, so do our models. We discuss different regional mobility models before extending the del Rio-Chanona et al. model to include both occupational and regional mobility.

Next, we discuss a case study modelling the transition to a low carbon economy in Brazil. Although the Amazon Rainforest is one of the world's largest carbon sinks [17], deforestation and agricultural activity mean that Brazil is among the largest greenhouse gas emitters [18]. Additionally, Brazil's energy matrix is one of the least carbon-intensive in the world. A key focus for Brazil, therefore, is to reduce deforestation and develop sustainably over the next decade, with a commitment to lower emissions by 53% compared to 2005 levels by 2030 and a 2050 goal of carbon-neutrality, as set out in its Nationally Determined Contribution [19]. To help understand this pathway for Brazil, Ferreira Filho and Hanusch [20] model productivity pathways with a 0.5% total factor productivity increase, including one in the manufacturing sector and the other in the agriculture sector. Using these scenarios as target demand for our labour market model, we identify the possible skill and spatial mismatches involved in such pathways for Brazil.

To incorporate both skill and spatial mismatches, we construct an empirical mobility network with regional and occupational mobility. Through network analysis, we are able to understand the mobility patterns observed in Brazil, as well as how closely different occupations, regions, and demographics are represented in the data. We find that the regional mobility of workers across Brazil mirrors the geographical relationships between regions, and that transitions happen mostly within, rather than between, regions. When looking at occupational mobility alone, we find that similar occupations are close to each other, as are occupations with higher wages. Using this network, we are able to understand the rich structure of the mobility data, before using it in our modelling.

Applying the two distinct growth pathways to the extended del Rio-Chanona et al. model gives quantitative forecasts of detailed occupation and region unfilled vacancy rates, unlike much of the literature that identifies such 'hard-to-fill' vacancies using qualitative methods such as surveys [21, 22]. We find that a productivity increase in the manufacturing sector will be smoother for most workers, as the unemployment rate is lower than the baseline in a large majority of occupations and regions. However, there are also many occupations and regions where frictions created by a shortage of workers may affect the manufacturing growth path. We find a disconnect between occupations and regions with excess labour supply or demand which means that labour frictions are likely to slow down these growth pathways. Retraining and relocation policies need to be considered to enable smoother development in Brazil and our results identify where these should be targeted. Including geography in the model enables us to highlight to policy makers to what extent skill or spacial frictions are the key to enabling a smooth transition.

Finally, we make enhancements to the model by dropping some simplifying assumptions from the del Rio-Chanona et al. model as three important mechanisms in the labour market are currently missing. These are employer-to-employer transitions [23, 24, 25], which account for a significant portion of job transitions, multiple job applications to allow for the different search intensities of employed versus unemployed workers [26, 27], and wage driven dynamics [28, 29]. We compute the deterministic approximation of the extended full agent-based model by taking expectations of all the stochastic processes that determine the agents' movements, to ensure the extended model remains practical for use with hundreds of occupations and multiple regions. We show that the approximation accurately represents the agent-based model with toy model scenarios and analysis of the approximation error. We finish with analysis of the impact of the new parameters on the model outputs, using the automation shock studied by del Rio-Chanona et al. Thus showing that the resulting extended model is ready to be applied to other economic transitions, as well as for improvements to calibration.

Ultimately, this thesis contributes a granular, flexible, data-driven agent-based model of the labour market. We present multiple extensions to a unique stochastic model, including regional mobility, and calculate a deterministic approximation to the model. We construct and study the mobility network for Brazil, and discuss the labour market impacts of two growth pathways. This thesis adds to a growing body of literature using data-driven, agent-based models to study economics, and particularly the impact of the post-carbon transition, which is needed if we are to mitigate the human cost of the climate emergency we face.

1.1 Publications and presentations

During the course of this DPhil, this work was presented in various forms at the following:

- April 2022 – Poster (winning Best Student Poster) at the British Applied Mathematics Colloquium, University of Loughborough – Modelling the labour market: Can we predict occupation transitions?
- June 2022 – Talk at the Workshop on Economic Science with Heterogeneous Interacting Agents, University of Catania, Italy – Modelling the labour market and understanding occupational mobility in the US
- July 2023 – Talk at Network Science, University of Vienna, Austria – Measuring labour mobility frictions of the green transition in Brazil
- Autumn 2023 – Capacity Building sessions, Delhi, India and London, UK (in-person), and online for Brazil and China – Labour market modelling
- February 2024 – Talk at Women and Non-Binary People in Mathematics Day, University of Oxford – Region and occupation bottlenecks of the net zero transition

The work presented in Chapters 3 and 4 is based on a manuscript that is currently accepted, pending minor revisions, at Nature Sustainability. The first version of this work is available on arxiv at `https://arxiv.org/abs/2503.05310` under the title 'Skill and spatial mismatches for sustainable development pathways in Brazil'.

1.2 Thesis overview

In this thesis, we discuss work carried out for The Economics of Energy Innovation and System Transition (EEIST) project, working at the time with the UK Government Department for Business, Energy, and Industrial Strategy,[1] and a collaboration with the World Bank for the Brazil application. The work in Chapter 4 does not represent the views of the World Bank, or the countries it represents.

After discussing general labour market literature including networks and agent-based modelling in Chapter 2, we present the del Rio-Chanona et al. agent-based model in Chapter 3 with the deterministic approximation to the full stochastic agent-based model. We then discuss modelling regional mobility, construct a network of occupations and regions which allows us to use network analysis techniques to study worker mobility in Brazil, and present a simple toy model with the regional del Rio-Chanona model. In Chapter 4, using the regional occupational mobility network, we apply our framework to two growth pathways for Brazil, finding that an unmanaged productivity increase could increase wage inequality and lead to unemployment and unfilled vacancies in some occupations and regions.

In Chapter 5, we extend the del Rio-Chanona et al. model further, adding complexity to three modelling assumptions; we add on-the-job search, allow agents to send multiple job applications, and include wage driven dynamics. We derive the new deterministic approximation for the extended model and show qualitatively that it reflects the full stochastic model. We present a final case study in Chapter 6 to illustrate the impact of the new parameters introduced through the model extensions on an automation shock, and finally, we conclude in Chapter 7.

[1] Now split into three different departments, including the Department for Energy Security and Net Zero.

Chapter 2

Literature review

Previous work to understand the labour market and the impact of transitions ranges from retrospective analysis of specific policies to macroeconomic models for predicting future labour demand. In this section, we first discuss the broad literature reviewing historic labour market impacts from external shocks aimed at a greener economy. We then focus on various predictions for the labour market impacts of a transition to net zero, with much research carried out using input-output models and computable general equilibrium models. We then go on to discuss research into how skills and geography might affect the transition. Much of the recent research into labour market impacts represents the labour market structure with a network, and we discuss the various network approaches next. Finally, we discuss agent-based models of the labour market and our motivation to use and extend the model presented by del Rio-Chanona et al. [11].

2.1 Historic environmental policy impacts on labour

To understand the labour market impacts of specific policies, some studies have looked into the impact of previous environmental regulations on employment. There is evidence of mixed employment impacts after environmental regulations and policies have been enacted. Often, research finds no dramatic impact on employment. For example, Berman and Bui [4] find no evidence that local air quality regulation in Los Angeles substantially reduced employment. Gray et al. [30] investigate the impact of the Cluster Rule on employment. The Cluster Rule was imposed on the pulp and paper industry in the US in 1993 to reduce toxic releases into the air and water by pulp and paper plants. Although there was one group of plants with a statistically significant reduction in employment, in general the Cluster Rule had only very small effects on employment. Also, Morgenstern et al. [31] find that increased environmental spending in the US in four heavily polluting industries did not cause a significant change in employment. Looking at the industry level, they find not only job losses but job creation at emissions-intensive factories due to increased environmental spending and so, the opposition to climate policies on the basis of 'jobs versus the environment' is likely to be unfounded.

However, there is some evidence from other environmental regulations that might back up this opposition. Liu et al. [1] look at the impact of a new wastewater standard on textile printing and dyeing enterprises in the Jiansu region in China. They find that labour demand decreased overall by 7% and that the effects felt by different types of enterprises were heterogeneous. Considering the US Clean Air Act, Walker [3] estimates a 15% decrease in employment in non-attainment[1] areas, relative to that in attainment counties, and Greenstone [32] estimates 590,000 job losses in non-attainment counties.

These studies use typical economic methods such as difference-in-difference analysis and instrumental variable techniques. They seek to identify the causal effect of the regulation of interest. These studies are somewhat focussed, each considering one environmental policy with some identifying impacts at the plant-level [30], and others at the regional level [1]. These studies, that aim to quantify the employment impacts of historic environmental policies, highlight the nuances within the labour market that need to be considered when thinking about environmental policies. That is, such policies aimed at facilitating the transition to a green economy do not always create, nor always destroy, jobs and so the impacts of future policies needs to be considered.

2.2 Labour market modelling

Adding to the literature reviewing historic environmental policies, modelling efforts try to enable predictions of the impact of future policies. In this thesis, we focus on these models developed to look specifically at the labour market impacts of the post-carbon transition. While many researchers find that the post-carbon transition is likely to result in net job creation, shifts towards more sustainable development pathways will likely create and destroy jobs; a number of methods are used, including input-output models and computable general equilibrium (CGE) models.

[1]US counties with emissions above the level required by the act were labelled non-attainment.

Garrett-Peltier [33] develops an input-output model to understand the effect of a \$1 billion government spending shift from fossil fuels to clean energy. Garrett-Peltier finds that every \$1 million shifted from brown to green energy destroys 2.65 brown jobs and creates 7.49 green jobs, causing a net increase of five jobs. However, this model can only be used for short term prediction as it does not account for changes in the input-output structure of an industry, which is inevitable for new, fast-growing industries (such as solar energy). Additionally, there is no consideration of skill requirements or ease of transition for workers moving between the industries where the increase or decrease in labour demand is concentrated.

Similarly, Montt et al. at the International Labour Organisation [34] use an input-output model to investigate the global employment impacts of keeping global temperatures to 2 degrees above pre-industrial levels. They find a net increase in employment but that to realise this increase the necessary job reallocation across industries could be significant. Again, this reallocation does not consider the occupation or skills of workers required to transition.

These input-output models are good for short term estimates and understanding indirect effects that propagate through the input-output network. However, these models are limited because they ignore changes in input structure and, in any reallocation, do not consider the difficulty faced by workers required to change jobs. In this thesis, we aim to tackle the question of the difficulties faced by workers but our model is also limited by the changes to the underlying mobility network used. In Section 2.6, we discuss a model that does manage to overcome these limitations [35] and in Chapter 7, we detail how this reliance on a static network might be removed.

Analysis using computable general equilibrium (CGE) models to understand the net zero transition also find different aggregate labour market impacts. For example, Fragkos and Paroussos [36] develop a CGE model to evaluate the recent European Union (EU) Energy and Climate policy framework [37]. Using the employment factor approach, they predict that the transition to a low-carbon economy will directly create 200,000 jobs in the energy sector. Further analysis using their CGE model finds that the transition could lead to a reallocation of about 1.3% of jobs in the EU by 2050. Ram et al. [38] also use the employment factor approach and find that a 100% renewable power system may result in an increase in jobs in the energy sector from 21 million in 2015 to 35 million in 2050. However, this research does not consider the policies required to enable the transition of workers into these new jobs.

Not all CGE models conclude that the post-carbon transition will result in net job creation. Castellanos and Heutel [5] develop a CGE model and consider the effects of a carbon tax under perfect mobility and perfect immobility of labour. In both scenarios, they find that aggregate unemployment could increase by 0.2–0.4 percentage points. In the perfect immobility scenario, the effect on unemployment in fossil fuel sectors is much larger, reaching a 24 percentage-point increase for the coal sector. They conclude that by not considering mobility frictions, models risk greatly under-predicting sectoral labour impacts. Hafstead and Williams [6] develop a two-sector computable general equilibrium model to compare the employment effects of environmental performance standards and environmental taxes. They find that, while aggregate impacts on employment are small, a substantial employment shift from the polluting sector to the non-polluting sector will be required. Although the authors consider two sectors, there is no dissagregation to occupation level effects.

AlShehabi [39, 40] develops a CGE model focused on modelling energy, crude oil, and production in Iran. They show that returning crude oil and fuel subsidies to households has adverse labour market effects while investing the subsidies improves the labour market. Liyanaarachchi, Naranpanawa, and Bandara [41] also use a CGE model to study the impact of a trade liberalisation policy in Sri Lanka. They find that such a policy would have mixed impacts such as increasing economic growth but also increasing income inequality. Neither of these models consider worker mobility frictions when assessing the labour market impacts of different policies.

Collectively, these CGE models are good for their flexibility in assessing different policies and providing quantitative estimates of the aggregate demand impacts of these policies. However, they assume equilibrium and largely overlook labour market frictions (Definition 2) or the difficulties workers might face in finding a new job. It has been shown repeatedly that labour market frictions are important when studying the impacts of economic policy relating to sustainable development [42] and sustainability transitions [43], such as net zero [44]. This is to avoid missing complex dynamics, such as the transitions workers make in reaction to labour demand changes and the indirect effect of these job transitions on neighbouring occupations. Therefore, to answer specific questions relating to labour market frictions and the impacts of economic policy on workers, we next consider research into key mechanisms within the labour market.

2.3 Skills

One of these key mechanisms is the consideration of workers' skills, and the possible disparity in skill requirements of the occupations that workers need to transition between. The OECD highlights the need for further research in this area [45].

Vona [2] emphasises the importance of research into so-called green jobs, due to the disparity between the skill requirements of emissions-intensive jobs and green jobs. Bowen et al. [46] use the O*NET database and its classification of green jobs to estimate the impacts that these green jobs will have on the labour market. They find that the jobs classified as 'green' by O*NET require a range of green and non-green tasks, concluding that jobs should be defined using a continuum, rather than a binary characteristic. However, in contrast to Vona, they conclude that the job transitions induced by the post-carbon transition are likely to be similar to existing job transitions.

More recent analysis of job vacancies and skill descriptions agrees with Vona, and suggests that skills and wage gaps can arise during the post-carbon transition. Sato et al. [8] find that low-carbon jobs are more skills intensive but, in recent years, have not come with a wage premium. Saussay et al. [47] also find that the skill requirements of low-carbon jobs differ from high-carbon jobs in important ways. Both of these studies conclude that careful consideration will be needed to facilitate the transition to a low-carbon economy if we want to ensure workers of all skill levels can find employment.

If the new skill requirements of low-carbon jobs are not addressed, economies may face a skill mismatch (Definition 3) between available workers and available jobs. Bücker et al. [9] find strong evidence of a skill mismatch between jobs created and jobs destroyed during a rapid decarbonisation of the power sector, with industries struggling to find suitably skilled workers and displaced workers struggling to find employment. As well as possibly impacting the transition to a low carbon economy, skill mismatches have been shown to negatively impact displaced workers' earnings [48] and workers particularly affected by skill mismatch early in their career are at risk of persistent lower career earnings [49, 50].

A policy-based solution to combat skill mismatches has, so far, been difficult to pin down. McGuinness, Pouliakas, and Redmond [51] find that there are many different causes of skill mismatch and more work is needed to find effective policy solutions. Adalet McGowan and Andrews [52] compare skill mismatch with policies in 22 OECD countries, and find certain policies that are associated with lower skill mismatch, such as laws that do not over-penalise business failure, as well as specific job-related factors such as negotiable wages and continued education during employment. However, Kupets [53] warns that over-education and skill mismatches can coincide during economic transition, therefore, more education is not necessarily the solution.

Skill mismatches can lead to 'hard-to-fill' vacancies in specific occupations, slowing down transition pathways. Much of the current work looking to identify occupations with 'hard-to-fill' vacancies uses anecdotal reports by employers [21]. Attstrom et al. [22] go further and categorise a vacancy as 'hard-to-fill' if open for at least three or six months, or is expected by employers to be 'hard-to-fill' in the upcoming year. Lankhuizen et al. [54] identify that geographical and inter-industry differences between available workers and open job vacancies might slow down the energy transition in the Netherlands but there is little occupation level quantitative research into these unfilled vacancies. In this thesis, we address this gap in the literature by identifying occupations with unfilled vacancies using a quantitative model of occupations and regions.

2.4 Geography

As well as skills, it is well studied that geography plays a key role in labour market outcomes for workers and employers. The research considers all aspects, including regional differences in unemployment, firm success, and inequality [16]. We start this section with a look at some analytic studies of these regional differences.

Bilal [55] identifies that the geographical concentration of employers in similar industries leads to increased spatial disparity of job opportunities and unemployment for workers. Jara-Figueroa et al. [56] study the importance for new firms of hiring workers with geographical, industry, or occupation knowledge and find that hiring for geographical and industrial knowledge increases their chance of firm success. Overman and Zu [57] find that spatial differences between regions in the UK lead high-wage and high-skill workers to be geographically concentrated and policies to combat inequality of job opportunities need to carefully consider workers at both ends of the skill spectrum. Dickey and Magante [58] look further at the drivers of regional migration in the UK and find that skills, housing, and wages are among the key motivations. Overman and Zu, and Dickey and Magante's findings are complemented by the work around urban wage premiums and their complex causes [59, 60, 61]. These studies of existing data identify the spatial disparities in factors such as job opportunities, wages, and skill requirements, and demonstrate the importance of geography when thinking about the labour market.

In order to investigate the causes of spatial differences and the possible impacts of future regional shocks, other researchers model geography in the labour market, with a range of different approaches. Some CGE models have been developed with a spatial dimension, these models allow for different region-level effects not previously modelled in national-level CGE models [62, 63]. Specifically, these models account for location-specific differences, such as infrastructure, environment, and production. However, similar to the work described in Section 2.2, these CGE models have limitations, including assuming equilibrium and overlooking occupation level frictions.

Other studies use models that were specifically developed to study the labour market. Kuhn, Manovskii, and Qiu [64] extend the classic DMP model to include on-the-job search and study the causes of spatial differences in the labour market. The DMP model, named after the 2010 Nobel prize winners Diamond, Mortensen, and Pissarides, is itself a search and matching model, presented in full in the second edition of Pissarides' book [65]. Briefly, unemployed workers search for jobs and firms post vacancies. A match is dependent on some wage bargaining in which the firm and the worker compare the benefit of a match with the cost of no match. With the extended model, Kuhn, Manovskii, and Qiu find that two thirds of geographical differences in unemployment can be attributed to different job-separation rates, with the remaining one third attributed to different job-finding rates, highlighting the need for regional modelling of labour market dynamics.

Another widely used regional labour market model is Blanchard and Katz' [66] 1992 model of labour mobility. They model worker movements between regions to understand unemployment and employment differences. They do not consider mobility frictions or different occupations, but their model has been extended in many directions, notably by Vega and Elhorst [67], who include spatial relationships. In the Blanchard-Katz model, regions are independent. Vega and Elhorst include geography to relate regions to each other in spatial ways, for example sharing a border, or journey time. They find that regional shocks hit the primary region hardest but also dissipate into neighbouring regions, with some time delay. This shows that analysing demand shocks at the regional level can give insight into how impacts dissipate across a country; this motivates our extension of the labour market model in this thesis to include geography.

Others modelling regional labour mobility look at the decision making processes of individuals, or households. Kern and Stein [68] model the decision making process of couples using multilevel structural equation modelling. They show that including partnerships and two-step decision making in their model improves their ability to explain the mobility data, concluding that this is evidence of partner and contextual effects in decision making about regional mobility by couples. Modelling decision making gives an insight into motivations behind migration, and job switching, however, in this thesis we do not model decision making explicitly, rather abstracting this process with a network.

Remote work has changed the impact of geography on some jobs. The COVID-19 pandemic accelerated remote working and since, there have been numerous studies to quantify occupations which can be performed remotely [69, 70, 71] and to understand the impact of remote work on the labour market. Among these, Althoff et al. [72] identify that, in the US, cities are most likely to feel the effects of remote work as they have the highest concentration of jobs with the possibility for remote work. They also find that remote work has one-sided consequences for service workers in cities, who depend on business workers' spending. Complementary to the US, Luca, Özgüzel and Wei [73] find further evidence that cities are at risk in 30 European countries. However, they note that a city exodus could give rural areas an opportunity to attract remote workers, potentially reducing spatial differences in the labour market.

Looking at the transition to net zero, Lim, Aklin, and Frank [7] find that understanding the geographical distribution of possible inequality is important for policy makers to navigate the net zero transition. Additionally, While and Eadson [74] study the economic restructuring needed to transition to a low-carbon economy and emphasise the possibility of further geographic inequality as a consequence of such a transition.

We have reviewed research that shows that both skills and geography need to be considered by policy makers in the net zero transition [47, 44], however, they are often treated separately in the modelling literature. Such research does not enable modelling of the interaction of skills and geographical disparities or quantify which is more important for labour market outcomes. In this thesis we address this gap by studying skill and spatial mismatches together, using a network.

2.5 Networks

Networks have been used to represent many different systems, from social networks to physics. A network is a way to represent relationships between objects. We represent the objects as nodes, say people in a karate club, and the relationships are represented with edges, say the presence of interactions outside of the club between two members. Using networks gives us a powerful way to analyse the structure of these systems, such as finding central objects, identifying communities, and calculating connectedness and relatedness of people or objects [75, 76].

In the labour modelling literature, much of the work incorporating networks started with social networks. A worker's social network is a key tool they can leverage when looking for a new job. In 1991, Montgomery [77] developed a model with workers and firms to understand how finding jobs through connections and hiring workers through referrals affects wages and profits. Using a random network for social connections between workers, they find that better connected workers achieve higher wages and firms that hire more through referrals make more profit. Calvó-Armengol and Jackson [78] find it is beneficial for unemployed workers to have more employed workers in their social network, in order to hear about more job openings. They later extend their model [79] to include wages and find persistent wage differences for agents starting as employed versus unemployed.

In this thesis, we expand on the more recent work that constructs networks to model the structure of occupations, regions, and industries instead of incorporating social networks into our model of labour market dynamics. This area of work began with the doctoral thesis of Gianelle [80] in 2010 who constructed the first network of job transitions. From there, as the field of network science was growing, Guerrero and Axtell built a national firm-level labour-flow-network (LFN) [81] and used network techniques to analyse worker mobility. In a series of studies, Guerrero and López developed this field, leading to using the LFNs as the basis for a stochastic model, the first of its kind to incorporate the rich network structure we discuss in this section into a labour market model [82, 83, 84].

There are many different ways to construct these networks, with edges able to represent anything such as co-occurrences, worker flow between nodes, and similarity of nodes. For example, Hartmann et al. [85] use a bipartite network of occupations and industries to construct a network of occupations, based on their co-appearance in industries. They find strong community structure which exposes a gender split, as well as differences in education, wages, and racial diversity. Schmutte [86] also constructs a bipartite network of occupations and industries, providing insights into the structure of mobility in the labour market. In fact, much analysis of these networks finds rich structure and clustering of different demographics, indicating the need to consider the relationships between occupations when modelling the labour market.

Another way to construct an occupation level network is to look at the similarity of the work carried out by occupations. Christenko [87] develops a network where nodes are connected based on the tasks workers perform in each occupation. An edge is present if two occupations share work activities, normalised by the importance of each activity and the knowledge required. This work closely follows the work by

Mealy et al. [88] who construct a 'job space' network using a broader definition of work activities and find that the occupations split into nine clusters.

As well as building links using characteristics such as work activities and industry co-appearances, a popular construction of labour networks is to measure labour mobility. At the industry level, relationships between industries can be identified. Neffke, Otto, and Weyh [89] study inter-industry labour flows and find the network is sparse, stable over time, and predictive of industry growth. O'Clery and Kinsella [90] also find a predictor of industry growth using a clustering of such an industry network. They find that clustered industries require similar skills, and increased worker availability within an industry cluster is a predictor of employment growth.

A network of labour mobility between firms can also be constructed. Nimczik [91] constructs such a network for Austria and partitions the firms into 'markets' using a stochastic block model. While these markets are geographically clustered, they do not necessarily align with administrative regions, such as counties or states. This new clustering can give insights into the reaction of workers facing demand shocks, that were previously not captured when looking by prescribed geographic areas.

Lastly, occupation level mobility networks are widely used and researched to understand the structure of the labour market. For example, Cheng and Park [92] construct a US mobility network for worker transitions between occupations and analyse the structure, finding that communities of occupations have become more separated over time. Toubøl and Larsen [93] construct a mobility network for the Danish labour market and use a clustering algorithm to examine the class structure of occupations. Also, del Rio-Chanona et al. [11] use a US occupation transition network to capture the labour market structure in their agent-based model.

The labour market network literature is extensive, with many different mechanisms to consider, such as social networks and relationships between industries. As well as network analysis, many papers have used their network as the structural basis of different economic models of the labour market. A subset of these models are agent-based models, which we focus on in this thesis.

2.6 Agent-based models

Arthur [10] argues that we cannot fully understand the economy if we assume equilibrium, therefore, modelling the labour market without equilibrium is a necessary focus for research. Arthur argues that Complexity Economics is the way forward to understanding how agents behave in many subfields of macroeconomics without assuming equilibrium, including the labour market. Arthur explains that removing the assumption of equilibrium in economic modelling enables models to capture emergent behaviour and that using agent-based models, with heterogeneous agents that interact and make decisions yields a more fluid, unrestricted model of the fluid, unrestricted economy they are trying to emulate. This motivates us to concentrate on agent-based labour market models in this thesis.

In the early days of the COVID-19 pandemic, Pichler et al. [94] developed an agent-based model and used it to accurately predict the GDP impact of UK lockdowns. They use a combination of micro-data, an agent-based model with agents of firms and workers, and a disequilibrium input-output approach. This model provides good evidence for using agent-based models to understand the economy. A key feature of agent-based models is emergent behaviour. Agents are given specific characteristics and behaviours, and, through interactions, emergent phenomena appear. In the Pichler et al. model, one emergent behaviour was the aggregate impact on GDP. Other successful agent-based models range from the simple, such as modelling bird flight [95], to the complex economic models discussed in this section.

Looking specifically at labour market agent-based models, Axtell [96] presents an agent-based model of the labour market that required years of work constrained by computer power, during which Axtell was unable to run his model with the desired 120 million workers. Axtell uses his agent-based model to understand the creation, growth, and survival of firms. The model is shown to reproduce key stylised facts such the firm size distribution and firm survival probability. After discussing the emergent firm level behaviour of the model, and what the results tell us about how firms work in the economy, Axtell concludes that economics should follow the other social sciences in leveraging big data and computer simulation with such complex agent-based models.

In Axtell's model, although workers have different effort levels, and desire for income and leisure, there are no occupations, or regions. Moro et al. [97] develop another agent-based labour market model with occupations, and consider the impact of demand shocks on the labour market. Using a network where edges are based on the similarity of skills, their work shows that the connectivity of occupations within a city can influence unemployment. That is, during a demand shock, workers in a city with more connections to their occupation are able to find employment more easily than a less connected city. While this model provides a new, occupation level analysis, Mealy et al. [88] found that a task-based network describes worker's job mobility behaviour better than a skill-based network. Therefore, instead of using a network of skill similarities between occupations, we focus on mobility networks.

Building on such mobility networks, Fair and Guerrero [35] develop an agent-based model of the labour market which, for the first time, is able to endogenously reproduce the network of transitions between occupations, industries and regions in the UK. Additionally, Fair and Guerrero include data at the agent-level on wages, non-labour income, and consumption preferences for the first time. They also show that their calibrated model can then be used to study future demand shocks without the dependence on historic job transitions. This advanced agent-based model marked a key step forward in labour market modelling.

The agent-based model developed by del Rio-Chanona et al., and extended in this thesis, while more primitive than the agent-based models discussed in this section, has the special advantage of a deterministic approximation that can be used quickly for parameter exploration.

All countries need to undertake some degree of structural change in order to transition to a low carbon economy. Although much of the research above finds that the post-carbon transition is expected to generate a net increase in jobs, there is little investigation into how friction within the labour market will affect this transition, and whether any specific detailed occupations might experience a negative effect. In this thesis, we extend the del Rio-Chanona model [11] to address some of these shortcomings, presenting a case study for Brazil that investigates the possible skill and spatial mismatches associated with two growth pathways.

2.7 Definitions

Definition 1. Beveridge curve

The Beveridge curve is the name given to the relationship between the unemployment and vacancy rates [98]. The curve typically moves anti-clockwise; during an economic upturn, there is an efficient labour market with job vacancies being filled more quickly, however, during a recovery period a shift away from the origin normally occurs, caused by a mismatch between unemployed workers and open vacancies.

Definition 2. Labour market frictions

In this thesis, labour market frictions refer to the limits that the labour market structure (captured in our model by the occupational mobility network) impose on a worker's mobility between occupations, or regions.

Definition 3. Skill mismatch
In this thesis, skill mismatch refers to a difference between the skills required by the open vacancies and the skills of worker's looking for a new job, as the ILO define [99].

Definition 4. Steady state
An economy is at a steady state when all economic variables are not changing over time.

Definition 5. Equilibrium
A system is in economic equilibrium if supply equals demand. The system does not have to be in a steady state to be at equilibrium, as various economic variables could be changing with supply still equal to demand.

Definition 6. Labour reallocation
Labour reallocation for a scenario (or occupation, region, or otherwise), is defined in this thesis as the number of jobs created in said scenario for a given time period.

Chapter 3

Labour market model with geography

This chapter addresses the questions:

- How do workers transition between regions and occupations in Brazil?
- What modelling approaches can we use to capture these transitions?

The first section is continued literature review describing the novel, data-driven network model of the labour market developed by del Rio-Chanona et al., following [11]. We discuss the full stochastic agent-based model and the random processes involved, before presenting the novel calculation by del Rio-Chanona et al. of the expectation of all these processes, to reach a deterministic approximation to the full model.

One of the limitations of the model is that it only separates workers by occupation. Assuming that workers can take up any job that becomes available, with no relocation considerations restricts the analysis to a single city. This limits the applicability of the model and so, for the first time, we extend the model framework to include regional mobility. The extension to geography gives the model greater explanatory power and gives us the unique ability to compare regional frictions with occupational frictions. In later sections of this chapter, we introduce regional mobility into the model using a mobility network and, discuss and analyse the empirical network for Brazil. Finally, we discuss a toy model to show the effect that regional mobility can have on the model outputs, by studying the unemployment impact of different future demand profiles.

3.1 Del Rio-Chanona model

In this section, we introduce the del Rio-Chanona et al. [11] labour market model and its deterministic approximation. We discuss the full stochastic agent-based model and the random processes involved, before presenting the deterministic approximation to the full model, reached by taking the expectation of all the random processes.

3.1.1 Model definition

To model the flow of workers around the labour market network, del Rio-Chanona et al. [11] use three discrete-time stochastic equations for each occupation. These model the number of employed and unemployed workers, as well as the number of vacancies. Workers can then be hired or separated (fired) at each time step, and vacancies can be opened and filled.

At time t, let $E_{i,t}$ be the number of workers employed in occupation i, $U_{i,t}$ be the number of workers unemployed in occupation i, and let $V_{i,t}$ be the number of vacancies open in occupation i. An unemployed worker is considered to be in the occupation in which they were most recently employed. We describe the labour flow by $F_{ij,t+1}$ which is the number of workers hired in occupation j who were previously employed in occupation i. Let $B_{i,t}$ be the number of workers separated from occupation i at time t, and let $C_{i,t}$ be the number of vacancies that are opened in occupation i. Then the equations governing the flow of workers around the network are

$$E_{i,t+1} = E_{i,t} - \underbrace{B_{i,t+1}}_{\text{separated workers}} + \underbrace{\sum_j F_{ji,t+1}}_{\text{hired workers}}, \tag{3.1}$$

$$U_{i,t+1} = U_{i,t} + \underbrace{B_{i,t+1}}_{\text{separated workers}} - \underbrace{\sum_j F_{ij,t+1}}_{\text{transitioning workers}}, \tag{3.2}$$

$$V_{i,t+1} = V_{i,t} + \underbrace{C_{i,t+1}}_{\text{opened vacancies}} - \underbrace{\sum_j F_{ji,t+1}}_{\text{hired workers}}. \tag{3.3}$$

These three equations describe the conservation laws governing the flow of workers in occupation i. In (3.1), the change in employment is equal to the number of workers hired into occupation i minus the number of workers separated from i. In (3.2), the change in unemployment is equal to the number of workers separated in occupation i minus the number of workers hired from occupation i. In (3.3), the change in the number of vacancies is equal to the number of vacancies opened minus the number of positions filled in occupation i. The time step is chosen so that only one event happens in each time step. A worker cannot transition from employment to unemployment and then back to employment within one time step.

We model the number of workers separated and the number of vacancies opened at time $t+1$ using a binomial distribution. The number of workers separated from occupation i, $B_{i,t+1}$, is drawn from the binomial distribution

$$B_{i,t+1}|(E_{i,t}, p_{u,i,t}) \sim \mathrm{Bin}(E_{i,t}, p_{u,i,t}), \tag{3.4}$$

where $\mathrm{Bin}(m, p)$ is the binomial distribution with m trials and success probability p, and the number of vacancies opened in i, $C_{i,t+1}$, is drawn from the binomial distribution

$$C_{i,t+1}|(E_{i,t}, p_{u,i,t}) \sim \mathrm{Bin}(E_{i,t}, p_{v,i,t}). \tag{3.5}$$

The conditional probabilities $p_{u,i,t}$ and $p_{v,i,t}$ are calculated by two separate processes. There is a spontaneous process in which workers are separated and vacancies are opened at random, and a state-dependent process which is reacting to supply and demand within the labour market. We calculate these probabilities in the next section.

Finally, the flow of workers $F_{ij,t+1}$ depends on the structure of the network, the number of workers unemployed, the number of vacancies open, and the job searching and job matching processes. We liken this to an urn problem. Suppose each unemployed worker has a ball with their name on and each vacancy is represented by a bin.[1] A worker in occupation i chooses a bin representing a vacancy open in occupation j with conditional probability $q_{ij,t+1}$, and places their ball inside. After all unemployed workers have placed their balls, a ball is selected uniformly at random from each bin and the corresponding worker is hired. If no balls have been placed in a bin, no applications have been made so the vacancy represented by that bin is kept open for another time step.

This results in $F_{ij,t+1}$ being a stochastic variable that follows a multinomial distribution where the number of trials is the number of workers in occupation i that choose to apply to occupation j, $A_{ij,t+1}$, and with success probabilities, $P_{j,t+1}$, which we make explicit in Section 3.1.2, that are derived from the urn model. For fixed i, the number of workers that choose to apply to j, $A_{ij,t+1}$, also follows a multinomial

[1] For simplicity, for now, we assume each worker only sends out one job application per time step, and we relax this assumption in Section 5.3.

distribution, with $U_{i,t}$ trials and success probability $q_{ij,t+1}$ for $j \in I$, where

$$q_{ij,t+1}|\mathbf{V}_t = \frac{V_{i,t}\mathcal{A}_{ij}}{\sum_l V_{l,t}\mathcal{A}_{il}}. \tag{3.6}$$

Supply and demand for workers

Vacancy openings and worker separations are each broken into two separate random processes, one spontaneous process and one state-dependent process.

In the spontaneous process, separations and openings occur at random. We assume that the rates are constant over all occupations. Let the probability that any worker is randomly separated be δ_u. Let the probability that a vacancy randomly opens be δ_v.

The state-dependent process reacts to supply and demand in the labour market. Let $r_{u,i,t}$ be the conditional probability that a worker from occupation i is separated at time t by the state-dependent process and let $r_{v,i,t}$ be the conditional probability that a vacancy is opened in occupation i at time t by the state-dependent process. Let $\mathcal{D}^{\dagger}_{i,t}$ be the target labour demand for occupation i at time t.[2] We denote the model's realised demand as $D_{i,t} = E_{i,t} + V_{i,t}$.

We must have that $r_{u,i,t}$ and $r_{v,i,t}$ satisfy the following conditions:

1. If target demand and realised demand are equal, no adjustment is made. In other words, if $D_{i,t} - \mathcal{D}^{\dagger}_{i,t} = 0$ then $r_{u,i,t} = r_{v,i,t} = 0$.
2. $r_{u,i,t}$ is an increasing function of $D_{i,t} - \mathcal{D}^{\dagger}_{i,t}$. (When realised demand is greater than target demand, workers are more likely to be separated, thus decreasing realised demand towards target demand). Similarly, $r_{v,i,t}$ is an increasing function of $\mathcal{D}^{\dagger}_{i,t} - D_{i,t}$. (When target demand is greater than realised demand, more vacancies will appear to bring realised demand up to target demand).
3. $r_{u,i,t}$ and $r_{v,i,t}$ are probabilities and must lie in the interval $[0, 1]$.

In order for the conditional probabilities $r_{u,i,t}$ and $r_{v,i,t}$ to satisfy the above conditions and adjust at a linear rate with respect to the difference between the

[2] Target labour demand is a parameter of the model. In the toy model at the end of this chapter, we impose a sigmoid shaped demand change on the occupations, and in Chapter 4, we use an external macro model for target demand.

target demand and realised demand, following del Rio-Chanona et al., we propose the following functional forms for $r_{u,i,t}$ and $r_{v,i,t}$,

$$r_{u,i,t}|(E_{i,t}, V_{i,t}) = \frac{\gamma_u \max\{0, D_{i,t} - \mathcal{D}^{\dagger}_{i,t}\}}{E_{i,t}}, \tag{3.7}$$

$$r_{v,i,t}|(E_{i,t}, V_{i,t}) = \frac{\gamma_v \max\{0, \mathcal{D}^{\dagger}_{i,t} - D_{i,t}\}}{E_{i,t}}, \tag{3.8}$$

where γ_u and γ_v are parameters determining the speed of adjustment towards the target demand, with γ_u and $\gamma_v \in [0,1]$. $\gamma_u = \gamma_v = 1$ corresponds to the maximum adjustment speed and $\gamma_u = \gamma_v = 0$ corresponds to no adjustment at all. As they are probabilities, we require $0 \leq r_{u,i,t} \leq 1$ and $0 \leq r_{v,i,t} \leq 1$. These conditions are normally satisfied but in some extreme cases, r exceeds the upper bound. In these cases, we set $r = 1$. We also let $\gamma_u = \gamma_v = \gamma$.

The probabilities governing the spontaneous process (δ_u and δ_v) and the state dependent process ($r_{u,i,t}$ and $r_{v,i,t}$) are independent. The probability that a worker in occupation i is not separated at time t by the spontaneous process is $1 - \delta_u$, and the probability that a worker in i is not separated at time t is $1 - r_{u,i,t}$. So, the probability that a worker in i is not separated by either process is $(1 - \delta_u)(1 - r_{u,i,t})$, since they are not spontaneously separated and they are not separated by labour demands. Therefore, the conditional probability that a worker in occupation i *is* separated at time t is

$$p_{u,i,t}|(E_{i,t}, V_{i,t}) = 1 - (1 - \delta_u)(1 - r_{u,i,t}) = \delta_u + r_{u,i,t} - \delta_u r_{u,i,t}. \tag{3.9}$$

The last term avoids counting a worker as separated twice from the two processes. Similarly, the conditional probability that a vacancy opens in occupation i at time t is

$$p_{v,i,t}|(E_{i,t}, V_{i,t}) = \delta_v + r_{v,i,t} - \delta_v r_{v,i,t}. \tag{3.10}$$

We do not have any empirical data with which to calibrate the parameter γ, therefore, let $\gamma = 10\delta_u$. Results of the model appear to be fairly insensitive to the value of γ (as demonstrated with simulations by del Rio-Chanona et al. [11]). However, δ_u, δ_v and Δt can be calibrated using the Beveridge curve (Definition 1). See del Rio-Chanona et al. [11] for details.

Throughout this thesis, we will use the unemployment rate to analyse different toy model scenarios, two growth pathways for Brazil, and an automation shock in the US. The unemployment rate of occupation i represents the proportion of workers in occupation i that have lost their job, and have subsequently not found new employment. The average unemployment rate over a period, T, for occupation i is defined as

$$u_{i,average}(T) = \frac{\sum_{t\in T} U_{i,t}}{\sum_{t\in T}(U_{i,t} + E_{i,t})}. \tag{3.11}$$

3.1.2 Deterministic approximation for large populations

Running the agent-based model governed by (3.1)−(3.3) is computationally expensive and not practical for extended analysis. However, when the overall population, $\mathcal{L}$, is large, we can use the law of large numbers and multivariate Taylor expansions to calculate a deterministic approximation of the system using expected values. Expected values, conditional on the previous state of the system are denoted with lowercase letter and a bar. For example,

$$\bar{u}_{i,t+1} = \mathbb{E}[U_{i,t+1}|\mathbf{E}_t = \mathbf{e}_t, \mathbf{U}_t = \mathbf{u}_t, \mathbf{V}_t = \mathbf{v}_t] \tag{3.12}$$

where $\mathbf{u}_t = \{\bar{u}_{i,t}\}_{i\in I}$, and similarly for $\mathbf{e}_t$ and $\mathbf{v}_t$.

In this section, following del Rio-Chanona et al. [11], we present their approximations for separated workers and opened vacancies before moving onto the calculation for the flow of workers. In Section 3.2.1 we illustrate the results from the full agent-based model and the approximation for a toy model, where the approximation runs over 20 times faster than the full agent-based model. In Chapter 5, we introduce three model extensions, calculate their deterministic approximation, and give more mathematical justification for the approximation, as well as further discussion of the approximation error and computational gains.

Separated workers and opened vacancies

Workers are separated and vacancies are opened according to the binomial distribution. The expected value of the binomial distribution is the number of trials times the success probability. Therefore, given (3.4), (3.7), and (3.9) for $B_{i,t}$ and (3.5), (3.8), and (3.10) for $C_{i,t}$, the expected values for separated workers and opened vacancies at time $t+1$, conditional on employment at time t, are

$$\bar{b}_{i,t+1} = \bar{e}_{i,t} p_{u,i,t} = \delta_u \bar{e}_{i,t} + (1-\delta_u)\gamma \max\{0, \bar{d}_{i,t} - \mathcal{D}^\dagger_{i,t}\}, \tag{3.13}$$

$$\bar{c}_{i,t+1} = \bar{e}_{i,t} p_{v,i,t} = \delta_v \bar{e}_{i,t} + (1-\delta_v)\gamma \max\{0, \mathcal{D}^\dagger_{i,t} - \bar{d}_{i,t}\}. \tag{3.14}$$

Flow of workers

The flow of workers $F_{ij,t+1}$ is the number of workers in occupation i who apply to occupation j and are successful. This follows a multinomial distribution where the number of trials is the number of applicants from occupation i to j, $A_{ij,t+1}$. Recall, the random variable for the number of workers applying from occupation i to occupation j, $A_{ij,t+1}$, is also a multinomial with expectation given by

$$\mathbb{E}[A_{ij,t+1}|\mathbf{u}_t] = q_{ij,t+1}\bar{u}_{i,t}. \tag{3.15}$$

While workers only sent one applications per time step, the number of *applicants* is the same as the number of *applications*, but for consistency of notation, we introduce the random variable $S_{ij,t+1}$ here, for the number of applications sent from workers in occupation i to j and for now,

$$S_{ij,t+1} \equiv A_{ij,t+1} \tag{3.16}$$

for all $ij \in I$, and the expectation is given by

$$\bar{s}_{ij,t+1} = q_{ij,t+1}\bar{u}_{i,t}. \tag{3.17}$$

The probabilities for the multinomial distribution for the flow of workers $F_{ij,t+1}$ is the probability of any application to occupation j being successful, $P_{j,t+1}$. It does not depend on occupation i because a successful application is chosen independently of its origin. Let $S_{j,t+1}$ be the total number of applications occupation j receives, so

$$S_{j,t+1} = \sum_k S_{kj,t+1}. \tag{3.18}$$

Then let $M_{j,t+1}$ be the number of vacancies filled in occupation j. The probability of having a successful application, $P_{j,t+1}$, is the fraction of the number of vacancies filled out of the total number of applications received, given by

$$P_{j,t+1} = \frac{M_{j,t+1}}{S_{j,t+1}}. \tag{3.19}$$

Therefore, the multinomial $F_{ij,t+1}$ for the flow of workers from occupation i has $S_{ij,t+1}$ trials and success probability $P_{j,t+1}$ for $j \in I$ and the expected value is

$$\bar{f}_{ij,t+1} = \mathbb{E}[S_{ij,t+1}P_{j,t+1}|\mathbf{u}_t, \mathbf{v}_t, \mathbf{e}_t]. \tag{3.20}$$

Matches The random variable $M_{j,t+1}$ represents the number of vacancies filled in occupation j. Since any vacancy that receives at least one application is filled at the next time step (any non-empty vacancy accepts a candidate uniformly at random, and every chosen candidate accepts the job offer), $M_{j,t+1}$ can also be thought of as the number of job openings that receive at least one application.

The probability that a worker applying to occupation j applies to a specific vacancy is $1/\bar{v}_{j,t}$. So, the probability that a worker does not apply to said vacancy is $1 - 1/\bar{v}_{j,t}$. Therefore, the probability that none of the $\bar{s}_{j,t+1}$ workers applying to occupation j apply to said vacancy is $(1 - 1/\bar{v}_{j,t})^{\bar{s}_{j,t+1}}$. This means the probability that of at least one worker applying to specific vacancy is $1 - (1 - 1/\bar{v}_{j,t})^{\bar{s}_{j,t+1}}$. The expected number of vacancies filled in occupation j is the probability that of a specific vacancy being filled, times the number of vacancies in occupation j, therefore

$$\bar{m}_{j,t+1} = \bar{v}_{j,t}\left[1 - \left(1 - \frac{1}{\bar{v}_{j,t}}\right)^{\bar{s}_{j,t+1}}\right]. \tag{3.21}$$

Using the approximation $(1-a)^b \approx e^{-ab}$, for large a and b, yields

$$\bar{m}_{j,t+1} = \bar{v}_{j,t}\left(1 - e^{-\frac{\bar{s}_{j,t+1}}{\bar{v}_{j,t}}}\right). \tag{3.22}$$

In this thesis, we identify that the random variable $M_{j,t+1}$ for the number of filled vacancies in occupation j follows the Occupancy distribution, with details given in Section 5.2.1.

Flow of workers We can see from (3.19) and (3.20) that $F_{ij,t+1}$ depends on $S_{ij,t+1}$, $M_{j,t+1}$, and $S_{j,t+1}$. These variables are not independent however, in the limit for large $\mathcal{L}$, we show below that

$$\bar{f}_{ij,t+1} \approx \bar{s}_{ij,t+1}\frac{\bar{v}_{j,t}\left(1 - e^{-\bar{s}_{j,t+1}/\bar{v}_{j,t}}\right)}{\bar{s}_{j,t+1}}. \tag{3.23}$$

Define

$$S_{j\setminus i,t+1} = \sum_{k\neq i} S_{kj,t+1} \tag{3.24}$$

as the number of applications received by occupation j from unemployed workers in all occupations k except occupation i. Then the total number of applications sent to j is given by

$$S_{j,t+1} = S_{ij,t+1} + S_{j\setminus i,t+1}. \tag{3.25}$$

We define the following multivariate form for the flow of workers,

$$F_{ij,t+1} = g(S_{ij,t+1}, S_{j\backslash i,t+1}) = S_{ij,t+1} \frac{\bar{v}_{j,t}(1 - e^{-(S_{ij,t+1}+S_{j\backslash i,t+1})/\bar{v}_{j,t}})}{S_{ij,t+1} + S_{j\backslash i,t+1}}. \tag{3.26}$$

This enables us to perform a multivariate Taylor expansion of g around the expected values of $S_{ij,t+1}$ and $S_{j\backslash i,t+1}$. Recall that, for fixed i, $S_{ij,t+1}$ follows a multinomial distribution with $U_{i,t}$ trials and success probability $q_{ij,t+1}$, for $j \in I$, with

$$q_{ij,t+1}|\mathbf{V}_t = \frac{V_{j,t}\mathcal{A}_{ij}}{\sum\limits_{l} V_{l,t}\mathcal{A}_{il}}. \tag{3.27}$$

So $S_{ij,t+1}$ and $S_{il,t+1}$ are drawn from the same multinomial distribution and are therefore not independent but $S_{ij,t+1}$ and $S_{kj,t+1}$ are drawn from different distributions and are therefore independent. One can think of this as the number of applications from occupation i to occupation j, $S_{ij,t+1}$, affects $S_{il,t+1}$ because workers in i can only apply to one destination occupation per time step. If a worker, in i, applies to occupation j, they cannot then also apply to occupation l. In the extreme, if all workers from occupation i apply to occupation j then none can apply to occupation l. On the other hand, workers in different occupations apply for jobs completely independently of each other. A worker in occupation i applying to occupation j does not affect a worker in occupation k also applying to occupation j.

It follows that $S_{ij,t+1}$ and $S_{j\backslash i,t+1}$ are independent. We now expand g around the expected values of $S_{ij,t+1}$ and $S_{j\backslash i,t+1}$ using the Taylor series. Therefore

$$\begin{aligned}
&\bar{g}(S_{ij,t+1}, S_{j\backslash i,t+1}) \\
&= g(\bar{s}_{ij,t+1}, \bar{s}_{j\backslash i,t+1}) \\
&\quad + \frac{1}{2}\frac{\partial^2}{(\partial S_{ij,t+1})^2}\big(g(\bar{s}_{ij,t+1}, \bar{s}_{j\backslash i,t+1})\big)\mathrm{Var}[S_{ij,t+1}] \\
&\quad + \frac{1}{2}\frac{\partial^2}{(\partial S_{j\backslash i,t+1})^2}\big(g(\bar{s}_{ij,t+1}, \bar{s}_{j\backslash i,t+1})\big)\mathrm{Var}[S_{j\backslash i,t+1}] + \cdots .
\end{aligned} \tag{3.28}$$

We introduce the following notation for the next expression only. Let $v = \bar{v}_{j,t}$, $x = S_{ij,t+1}$, $y = S_{j\backslash i,t+1}$ and $h(x, y) = g(S_{ij,t+1}, S_{j\backslash i,t+1})$. Let the expectation of x be denoted by μ_x and the variance denoted by σ_x^2, similar for y. Calculating the partial differentials of g in (3.26), we find that

$$\begin{aligned}\bar{h}(x,y) = h(\mu_x,\mu_y) + \sigma_x^2 \Bigg[& \left(\frac{v\mu_x}{(\mu_x+\mu_y)^3} - \frac{v}{(\mu_x+\mu_y)^2}\right)(1-e^{-(\mu_x+\mu_y)/v}) \\ & + \left(\frac{1}{(\mu_x+\mu_y)} - \frac{\mu_x}{(\mu_x+\mu_y)^2} - \frac{1}{2}\frac{\mu_x}{v(\mu_x+\mu_y)}\right)e^{-(\mu_x+\mu_y)/v}\Bigg] \\ & + \sigma_y^2\Bigg[\frac{v\mu_x}{(\mu_x+\mu_y)^3}(1-e^{-(\mu_x+\mu_y)/v}) \\ & - \left(\frac{\mu_x}{(\mu_x+\mu_y)^2} - \frac{1}{2}\frac{\mu_x}{v(\mu_x+\mu_y)}\right)e^{-(\mu_x+\mu_y)/v}\Bigg] + o(\mathcal{L}^{-1}).\end{aligned} \tag{3.29}$$

Since μ_x, μ_y, σ_x^2, σ_y^2, and v scale linearly with $\mathcal{L}$, in the limit of a large number of agents, the second-order derivative terms are of the order of a constant c. The remaining terms are at most order $\mathcal{L}^{-1}$. The first term, $h(\mu_x,\mu_y)$, scales with $\mathcal{L}$. Therefore, in the limit when $\mathcal{L}$ is large,

$$\bar{f}_{ij,t+1} = \bar{s}_{ij,t+1}\frac{\bar{v}_{j,t}\left(1-e^{-\bar{s}_{j,t+1}/\bar{v}_{j,t}}\right)}{\bar{s}_{j,t+1}}. \tag{3.30}$$

That is, using (3.17) and (3.27),

$$\bar{f}_{ij,t+1} = \bar{u}_{i,t}\frac{\bar{v}_{j,t}\mathcal{A}_{ij}}{\sum_l \bar{v}_{l,t}\mathcal{A}_{il}}\frac{\bar{v}_{j,t}\left(1-e^{-\bar{s}_{j,t+1}/\bar{v}_{j,t}}\right)}{\bar{s}_{j,t+1}}, \tag{3.31}$$

where

$$\bar{s}_{j,t+1} = \sum_k \bar{s}_{kj,t+1} = \sum_k \left[\bar{u}_{k,t}\frac{\bar{v}_{j,t}\mathcal{A}_{kj}}{\sum_l \bar{v}_{l,t}\mathcal{A}_{k,l}}\right]. \tag{3.32}$$

Hence, if we bring together (3.1)–(3.3), (3.13), and (3.14), we have the approximation

$$\bar{e}_{i,t+1} = \bar{e}_{i,t} - (\delta_u\bar{e}_{i,t} + (1-\delta_u)\gamma\max\{0, \bar{d}_{i,t} - \mathcal{D}^\dagger_{i,t}\}) + \sum_j \bar{f}_{ji,t+1}, \tag{3.33}$$

$$\bar{u}_{i,t+1} = \bar{u}_{i,t} + (\delta_u\bar{e}_{i,t} + (1-\delta_u)\gamma\max\{0, \bar{d}_{i,t} - \mathcal{D}^\dagger_{i,t}\}) - \sum_j \bar{f}_{ij,t+1}, \tag{3.34}$$

$$\bar{v}_{i,t+1} = \bar{v}_{i,t} + (\delta_v\bar{e}_{i,t} + (1-\delta_v)\gamma\max\{0, \mathcal{D}^\dagger_{i,t} - \bar{d}_{i,t}\}) - \sum_j \bar{f}_{ji,t+1}, \tag{3.35}$$

with $\bar{f}_{ij,t+1}$ defined in (3.31) and (3.32).

This completes the del Rio-Chanona model framework. A wealth of model variations and robustness checks can be found in the del Rio-Chanona et al. paper [11].

3.1.3 Limitations

As with any model, there are certain limitations of the del Rio-Chanona model that we discuss next. Some of these limitations are remedied in this thesis, and others are beyond the scope of the model, and this thesis.

One key difference between this model and the other labour market models discussed in Section 2.6 is the homogeneous agents. Currently, the agents do not have different skill levels, nor do they have different effort levels and preferences for leisure such as Axtell [96]. Further agent heterogeneity could be added through demographics such as age and gender although each of these factors would require additional data to be implemented into the del Rio-Chanona model.

Another limitation of the model is the fairly simple model dynamics. This simplicity has the benefit of enabling the deterministic approximation, which speeds up model analysis, but limits the different labour market mechanisms that are modelled. For instance, by not including geographical mobility, the current model is limited to analysis of a single city, and there is no entry and exit of workers, which complemented by including age heterogeneity, would enable analysis of labour market issues such as early retirement, or an ageing population.

The model relies on an exogenous network structure to calculate the probabilities that a worker applies to each occupation. This reliance means that workers cannot make new transitions in reaction to the labour demand shocks that are analysed with this model at present. As discussed above, Fair and Guerrero [35] have been able to endogenously construct the mobility network for the UK, providing very promising results for this avenue of research for labour market predictions in the future.

The use of such static networks also limits the timescales that the model can be used for. Over time, mobility patterns change and therefore it should be expected that the underlying network changes. Since this model uses a static mobility network, one cannot use it for long-term analysis. In fact, the US Standard Occupational Classification (SOC) System was revised in 1990, 2010, and 2018, suggesting that occupations, and consequently the occupational mobility network, are relevant for around 10 years. This is the timeline for our analysis in Chapter 4, and in Chapter 7, we discuss moving away from this reliance on the static mobility network.

We do tackle some of these limitations in this thesis by adding geography, and model mechanics such as on-the-job search, multiple apps, and wage-driven dynamics. However, there is a great deal of scope for further development, which we discuss in Chapter 7.

3.2 Regional occupational mobility

As it has been repeatedly shown that the regional mobility affects labour market outcomes [55, 56, 57], the first extension to the del Rio-Chanona model in this thesis is to include geographical mobility. Before we extend the del Rio-Chanona model to have workers moving between both occupations and regions, we consider the methods that previous studies have used to model movement between regions.

When looking at migration between countries, gravity models have been widely used. These relatively simple models model country-to-country migration using their relative size and the distance between them [100]. There has been some strong criticism of these models [101], especially when used to influence policy without careful consideration of their limitations. However, Reina et al. [102] show that with the right treatment of data and the inclusion of socioeconomic and demographic factors, gravity models can still be useful for studying migration flows. These gravity models are often used by those modelling human migration to understand infectious disease spread [103]. In this context, radiation models are also used [104]. Like gravity models, radiation models take into account the relative populations of two places; however, they also take into account the opportunities that exist in the regions between the origin and the destination, rather than just the distance between them. They have the benefits of being parameter-free and having the flexibility to be generalised [105, 106].

These methods use characteristics of the regions to define the likelihood of migration between them. Others categorise these relationships using different data sources. In their extension of the Blanchard-Katz model, Vega and Elhorst [67] consider a variety of ways to define the spatial weights of the model, based on relationships between regions. They consider using spatial weights based on travel time, population size, and bordering regions, and find that, for their model, using a mixture of binary spatial weights for bordering regions and second-order neighbours is best.

These spatial weights can also be calculated from mobility data. As we discussed in Section 2.5, mobility networks can be constructed between regions, as well as other categories. Fair and Guerrero [35] build their agent-based model with three worker mobility networks (industries, occupations, and regions). In this case, workers move between occupations in the same way, regardless of their industry or region. Tong et al. [107] develop a two-layer firm level mobility network for high and low skilled workers which enables them to capture the different mobility of the two skill groups.

They find the mobility patterns of the two groups to be quite different, with many of the edges not coinciding. Such multilayer networks are a powerful extension to the classical network literature [108].

In their paper, del Rio-Chanona et al. [11] use a single-layer occupational mobility network, as discussed in Section 2.5, to describe the relationship between occupation pairs. Building on the work by Mealy, del Rio-Chanona, and Farmer [88], del Rio-Chanona et al. work with the occupational mobility network for the US. The network they use is calculated using data on historic transitions recorded in a US household survey. The transitions of workers transitioning between jobs are counted and a weighted, directed network is built with occupations as nodes.

As described above, a multilayer network could give the flexibility for different regions to have different occupation level mobility. In this thesis, we build on the multilayer mobility network structure first introduced by Tong et al. [107], and add geographical mobility by extending the occupational mobility network, where nodes represented occupations, to include regional mobility. In the multilayer network framework, each layer represents a region, and the network within each layer is the occupational mobility network specific to that region. Due to the many inter-layer edges that could be present between every region and every occupation, we describe this extended network as a single graph with two labels at each node for ease, so nodes now represent occupation-region pairs.

We construct the regional occupational mobility network using data from the Relação Anual de Informações Sociais[3] (RAIS). The RAIS data contains all active employer-employee contracts in each year in the formal labour force in Brazil, which accounts for about 67% [109] of the total labour force in Brazil.

The data are collected by the Brazilian Ministry of Labor and Employment yearly via employers, recorded at each location (rather than each firm). The data contains information at the firm- and worker-level, including firm location, industry, and region, and worker age, gender, race, contract dates, and wage. Many studies have analysed the RAIS data and the coverage of specific demographics. Specifically, there is the Pesquisa Nacional por Amostra de Domicílios[4] (PNAD) that can be used to understand the sample bias in the RAIS data. De Negri et al. [110] find that the RAIS data accurately reflects nationwide demographics such as gender and age, with some non-negligible differences for level of education and income which they attribute to the different philosophies of the two datasets (a census of the formal labour force

[3]Annual Social Information Survey

[4]Brazil National Household Sample Survey

and a survey of the full labour force). Cornwell, Rivera, and Scmutte [111] use the PNAD data to understand the relationship between the reported race of workers and their income. PNAD is self-reported data, while RAIS is collected from employers. Cornwell, Rivera, and Scmutte find that in RAIS, the reported race of workers is subjective and contributes as much as 40% of the racial wage gap as some workers who change job, have their race reported differently by their new employer.

We will use the RAIS data for our analysis. In Section 6.1.1, we discuss network estimation when using survey data however the full coverage of RAIS data of the formal labour force means that, with some measurement error as discussed above, the occupational mobility network we construct does capture every edge present in the true mobility network for the formal labour force in Brazil. We leave construction of the mobility network for the whole of Brazil using PNAD for further work.

We use data from 2011 to 2019 with 409 occupations and 16 regions.[5] These 409 occupations are a hybrid of the CBO2002 classification. We start with the 4-digit level of the CBO2002 classification, which has 570 occupations. In order to have a every node connected in one single network and a consistent list of occupations across all regions, we merge some of these occupations into hybrid 2- and 3-digit codes. This merging results in 409 unique occupations.

A link between two occupation-region pairs is the count of workers moving between them, observed in the data. The adjacency matrix of the network is then

$$\mathcal{T}_{i\alpha,j\beta}, \tag{3.36}$$

the number of workers that transitioned from occupation i in region α to occupation j in region β between 2011 and 2019.

The resulting network shows that both occupational and regional mobility strongly influence labour flows. As shown in Figure 3.1a, occupation-region pairs in the same region are positioned in close proximity to one another, signalling stronger links. We use the Force Atlas 2 [112] algorithm to position the nodes in these visualisations. Nodes repel each other like charged particles while an edge acts as an attraction force. The algorithm only uses the network structure with no additional information about the nodes but, despite this, the position of the regions in Figure 3.1a corresponds to their true geographical location in Brazil as shown in Figure 3.1b. This suggests that inter-regional transitions are somehow influenced by the geography of Brazil. To

[5]To avoid confusion with the five macroregions of Brazil, we use region here to refer to the 16 regions, defined by Ferreira Filho and Hanusch [20]. The regions are either states, for example Bahia, or groups of states, for example RSudeste, which is made up of Minas Gerais, Espírito Santo, and Rio de Janeiro.

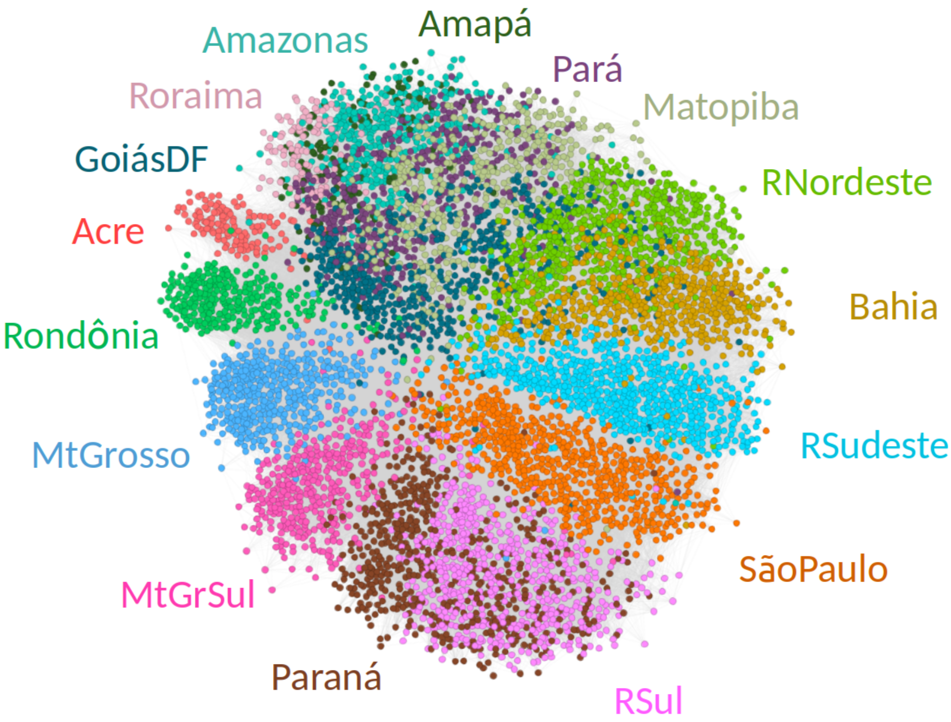
understand this further, below we compare the regional mobility network with the gravity model which uses the distance between regions to determine the inter-regional weights.

For some regions, the regional separation in the full regional occupational mobility network in Figure 3.1a is striking; Acre, Rondônia, RSudeste, and São Paulo are particularly separated. This suggests workers in these regions make more transitions within their region, than to other regions. However, this is not the case for all of the regions. Paraná and RSul appear mixed together, as do Roraima, Amazonas, Amapá, Pará, and Matopiba. This suggests that workers transition between these regions more frequently, and when we look at occupational links, we see that it is the occupations within these regions that are displayed close together.

In Figure 3.2a, we consider only occupation links. In this network, we find that nodes of the same broad category appear close together.[6] This suggests that it can be hard to transition to an occupation in a different category. In Figure 3.2b, using the same colouring as the national mobility network in Figure 3.2a, we can see that a similar occupation level structure is seen within each region's occupational mobility network. Notably, the higher wage occupations such as managers and professionals appear more towards the outside of the network, especially in the coastal regions, while lower skill occupations such as agricultural and industrial jobs are more towards the centre. This reflects that workers in those occupations transition more between regions than workers in the higher wage occupations. It appears that, in the mixed regions of Roraima, Amazonas, Amapá, Pará, and Matopiba we mentioned above, similar occupations are close together. These five regions mixing, in this projection of the network, suggests that workers transition more between these regions, while staying in the same occupational group (compared to other regions without as much inter-regional mobility). To understand the network structure further, we can measure the assortativity of the regional network across occupations and regions.

Assortativity of a network, with attribute x, was originally defined by Newman [113] and later extended to weighted networks by Yuan, Yan, and Zhang [114]. Assortativity measures the tendency of nodes to be connected with other nodes similar in attribute x. We use weighted categorical assortativity, which

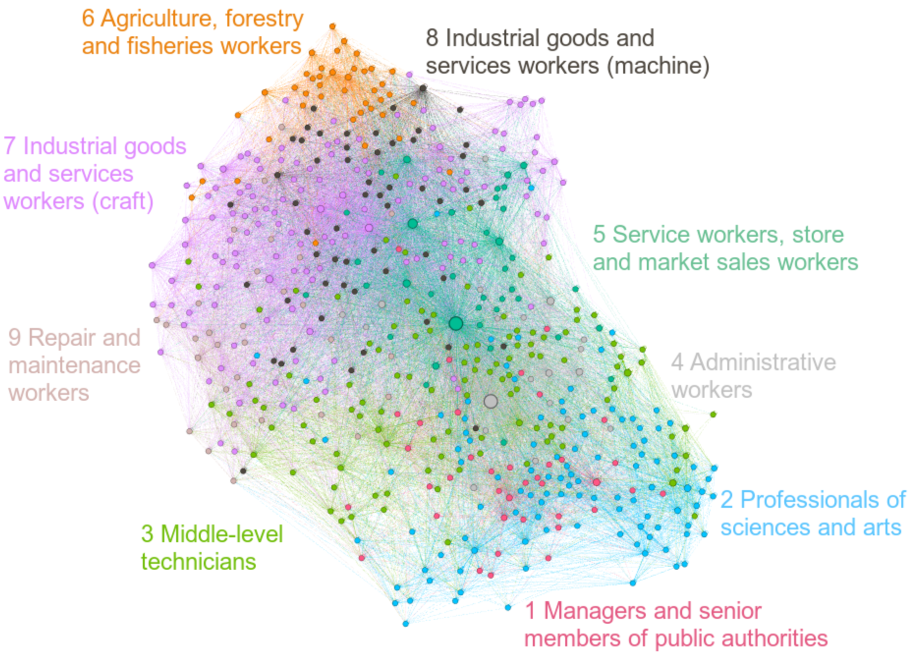
[6]These nine labelled broad occupation categories are the CBO2002 1-digit level classification, such as 'Professionals of sciences and arts', henceforth called occupational groups.



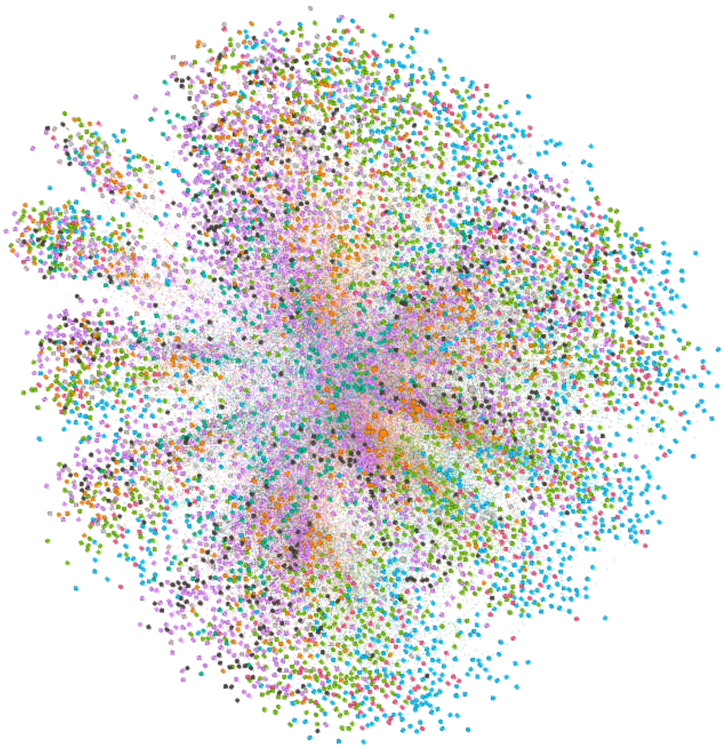
(a) Regional occupational mobility network coloured by region

(b) Brazil's states coloured by region

Figure 3.1: **Regional occupational mobility network for Brazil.** In (a) the full regional occupational mobility network, and in (b) a map of Brazil with states coloured similarly to (a).



(a) National occupational mobility network coloured by occupational group



(b) Regional occupational mobility network coloured as in (a)

Figure 3.2: **National occupational mobility network for Brazil.** In (a) the national occupational mobility network, without disaggregation by state, and in (b) the full regional occupational mobility network, coloured by occupational group.

is defined by Bücker et al. [9] as

$$r = \frac{\sum_i (e_{ii} - a_i b_i)}{1 - \sum_i a_i b_i} \tag{3.37}$$

where e_{ij} is the fraction of all edge weights that join nodes with value i to j, and a_i and b_j are the fraction of edges that start at i and end at j respectively.

In the occupational and regional network, we find that regional assortativity is 0.77, and the assortativity with occupational groups is 0.56. These results confirm that both regions and occupations influence labour mobility. The larger value for regional assortativity indicates that transitions happen mostly within, rather than between, regions.

Another network science tool we can make use of is community detection. Similar to the analysis of Nimczik [91], we can use a community detection algorithm to compare the network clusters with the administrative regions of Brazil. We use the Louvain community detection algorithm [115]. We choose the resolution parameter that maximises the Jaccard index between the community detection partition and the region partition.

As we saw in the visualisation in Figure 3.1, the regions of Brazil are strongly represented in the structure of the regional occupational mobility network. In Table 3.1, we see that the majority of occupations in each of regions are identified in their own cluster, apart from Matopiba and Pará, and RSudeste and São Paulo are joined in pairs within clusters 5 and 10, and Roraima is split into two clusters, 13 and 14. Interestingly, while Matopiba and Pará are mixed in the network visualisation, RSudeste and São Paulo are seemingly separate labour markets, aside from a few occupations that seem embedded in the opposite region. The occupation-region pairs in clusters smaller than 100 are counted in the 'Other' category, with most of these being in a community of size 1.[7]

As we discussed in above, a widely used model for migration is the gravity model. We can aggregate the regional occupational mobility network by region and compare the regional relationships from the empirical mobility network with a gravity model. The gravity model proposes, for origin region α and destination region β that

$$F_{\alpha,\beta} = k \frac{P_\alpha P_\beta}{(D_{\alpha\beta})^b} \tag{3.38}$$

[7] There are 1,251 communities of size 1, 45 communities of size 2, and there are 21 other communities of size 3–16 containing 125 occupation-region pairs.

Cluster	Acre	Amapá	Amazonas	Bahia	GoiásDF	Matopiba	MtGrSul	MtGrosso	Pará	Paraná	RNordeste	RSudeste	RSul	Rondônia	Roraima	SãoPaulo
0	353															
1		318														
2			355												7	
3				273												
4		1			298										1	
5		7				326			304		1					
6							321									
7								349	1							
8										331			1			
9											250					1
10	1	6	1	7	7	1	5		1	3	1	255	1		2	269
11					1								282			
12	1													375		
13															129	
14															180	
Other	45	56	53	129	103	82	83	60	103	75	157	154	125	32	74	139

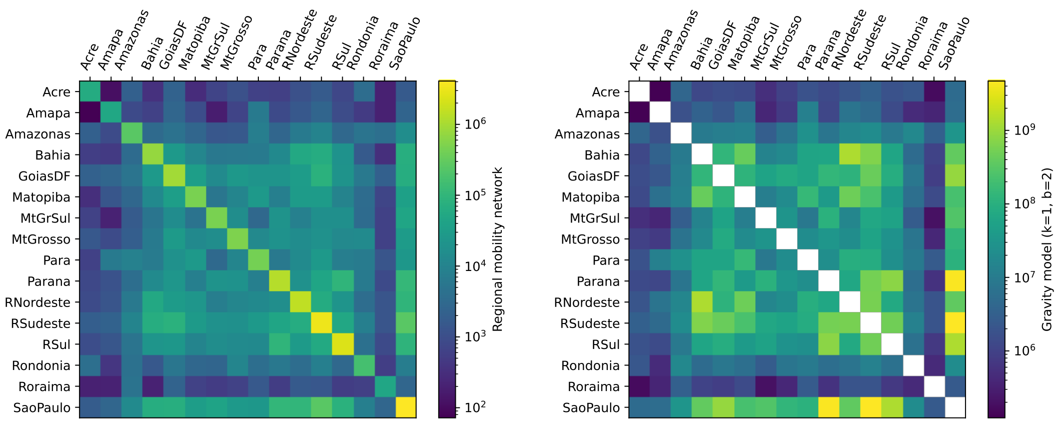
Table 3.1: **Community detection.** Allocation of regions within the clusters with size over 100, and the remaining counts in 'Other'.

where P_α is the population of region α, $D_{\alpha\beta}$ is the distance between regions α and β, and k and b are model parameters.[8]

In Figure 3.3 we show the adjacency matrix of the regional mobility network for Brazil and the matrix $F_{\alpha,\beta}$. We can see that the two matrices are similar, particularly the low mobility in Roraima, as well as Acre and Amapá. The gravity model does not give any values for the self-loops of the network, that is, how likely people are to stay within their current region. Del Rio-Chanona et al. [11] also had this issue with the US network as the census did not track people when they moved address. They used a single parameter to calibrate the weight of this self-loop. However, we can obtain this for each region from the RAIS dataset which gives some heterogeneity in the diagonal. When we look at the self-loops in the regional mobility network, we see that RSul, São Paulo, RSudeste, RNordeste, and Acre have the highest proportion of transitions within their region. This echoes what we saw in Figure 3.1a, where these regions have a clear separation from other regions.

One could study the rich structure in the regional occupational mobility network much further, comparing the mobility patterns of workers between regions with measures such as the Weighted Jaccard index [116], however, we leave this for future work and now focus on the distribution of demographics across the national

[8] We take $k = 1$ and $b = 2$ for this discussion.



(a) Regional mobility network (b) Gravity model matrix

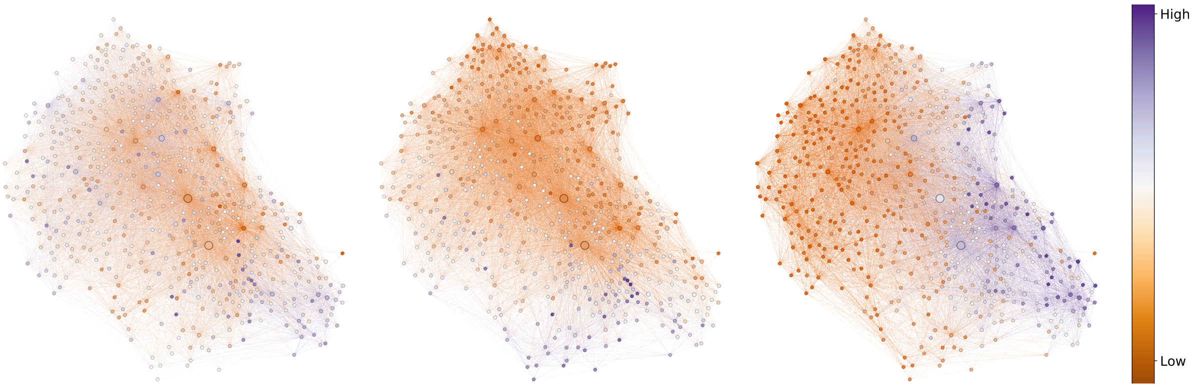
Figure 3.3: **Regional mobility network and gravity model.** In (a) the adjacency matrix for the regional mobility network of Brazil, and in (b) the matrix of relationships proposed by the gravity model.

occupation network, as shown in Figure 3.4. We see that there is a considerable separation of occupations according to the proportion of female workers and the average wage while there is more of an even distribution of occupations around the network according to average age.

In Figure 3.4a, we show the average age of workers in each occupation. Almost all occupational groups have occupations where the average worker age is both high and low, apart from 'Managers', where the average age of workers in all but three of 38 occupations is above the overall average.

In Figure 3.4b, we can see that the occupations with lower wages are concentrated in the upper and upper right section of the network. This contains occupations in the Administrative, Services, Agriculture, and Industry (craft and machine) occupational groups. Most occupations in the 'Agriculture, forestry and fisheries' occupational group earn below the median wage, all except 'forest mechanisation workers' and supervisors in 'farming', 'forestry, and aquaculture'. Similarly in Administrative and Services occupations, only supervisors and a select group of higher skilled occupations earn a higher than median wage.

In Figure 3.4c, we can see that occupations can also be clearly split by the proportion of female workers. In our data, the female workforce make up around 43.7% of the workers. These female workers are largely concentrated in the 'Professionals of sciences and arts', 'Administrative' and 'Services' occupational groups while very few female workers are employed in occupations in 'Repair and maintenance' or in 'Extraction and construction', 'Metal and composite



(a) Age (b) Wage (c) Proportion female

Figure 3.4: **Demographics on the occupational mobility network.** Each node is coloured by (a) median wage, (b) proportion of workers who are female, and (c) median age from 2018.

processing', and 'Electronics manufacturing and installing' within the 'Industrial (craft)' occupational group.

The two divisions of occupations by wage and by gender split are not the same. There are occupations with high wages with both high and low proportion of female workers, and vice versa, although, as is widely reported [117, 118, 119] in the literature, the highest 20 paid occupations, mostly managerial, judicial and engineering roles, have more men than women.

That different demographics are represented in different sections of the network highlights the importance of considering labour market structure, implemented in our model by using a mobility network, when studying labour impacts. If a labour demand shock is concentrated in one area of the network, the unemployment effects and labour frictions could also be concentrated on one demographic group within the labour force, resulting in increased inequality. Many macro-economic models trying to understand different economic transitions assume perfect mobility, or do not even consider different occupations, and so cannot capture this important structure.

To demonstrate the impact that including mobility frictions has on our model, we present the results for both the regional occupational mobility network, and the complete network. The adjacency matrix of the complete network, $\mathcal{K}_n$, with n nodes, is given by

$$\mathcal{K}_n = \frac{1}{n} J_n, \tag{3.39}$$

where J_n is the matrix of ones.

In the labour market model, we use the mobility network to represent the probability that a worker in occupation-region $i\alpha$ transitions to any of the other occupation-region pairs, therefore, the network used in the model is defined as

$$\mathcal{A}_{i\alpha,j\beta} = \frac{\mathcal{T}_{i\alpha,j\beta}}{\sum\limits_{j\beta} \mathcal{T}_{i\alpha,j\beta}}, \tag{3.40}$$

where $\mathcal{T}_{i\alpha,j\beta}$ is the count of transitions made by workers between occupation-region pair $i\alpha$ and pair $j\beta$. The probability a worker in occupation-region $i\alpha$ applies to an vacancy open in pair $j\beta$ is given by (3.6), and is proportional to the $i\alpha, j\beta$th element of this adjacency matrix, and the number of vacancies open in occupation-region $j\beta$.

3.2.1 Toy model

To illustrate the impact of including regions in our labour market model, we look at a two-occupation toy model with and without regions. As we discussed in Section 2.4, the geographical inequality of demand shocks has the chance to exacerbate economic transitions. In our toy model, total demand for the two occupations follows the same trajectory and we compare how different regional splits of this demand change affect unemployment rates. With this relatively simple model, we are able to show the benefits of adding geography into our model, especially for a transition which has the possibility of affecting regions differently.

Set up

We illustrate the need for extending the model to regions with a toy model. We run a simple scenario for two occupations with and without regions. All the occupations are identical to start with, in terms of employment, unemployment, and open vacancies. The heterogeneity comes from the network and the demand shocks. In the universal network, $\mathcal{A}$, without regions, Occupation 1 is the preferred occupation, as shown in Figure 3.5. In the regional network, $\mathcal{A}^g$, shown in Figure 3.6a, Region A is the preferred region, as we can see when we aggregate the network to regional mobility only, in Figure 3.6b. When we aggregate the regional network to occupational mobility only, we recover the same adjacency matrix as in the occupation only case, as shown in Figure 3.6c.

We discuss five scenarios, one without regions using the $\mathcal{A}$ network, and four scenarios with regions, using the $\mathcal{A}_g$ network. In all five scenarios, Occupation 1 has an increase in demand, and Occupation 2 has a decrease. The total demand for each occupation is always the same however, in the four regional scenarios, a

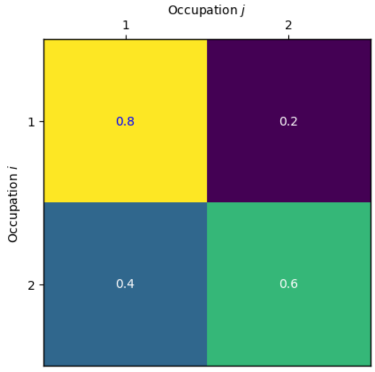
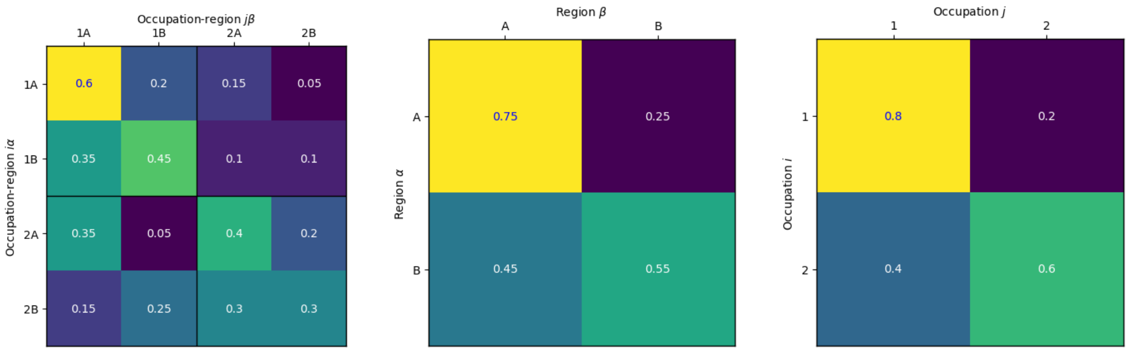


Figure 3.5: **Toy model occupational mobility network.** The occupational mobility network for the toy model with two occupations.

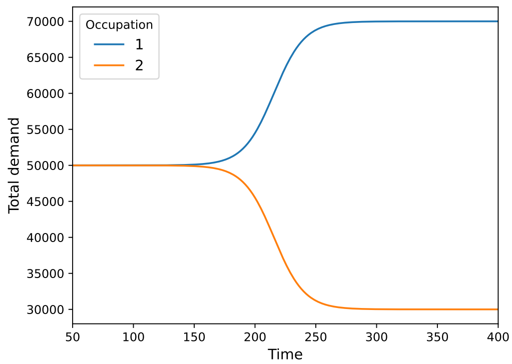


(a) $\mathcal{A}^g_{i\alpha,j\beta}$

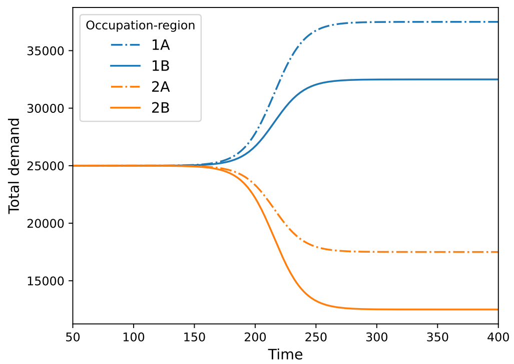
(b) $\mathcal{A}^g$ summed over occupations, $\mathcal{A}^g_{\alpha,\beta}$

(c) $\mathcal{A}^g$ summed over regions, $\mathcal{A}^g_{i,j}$

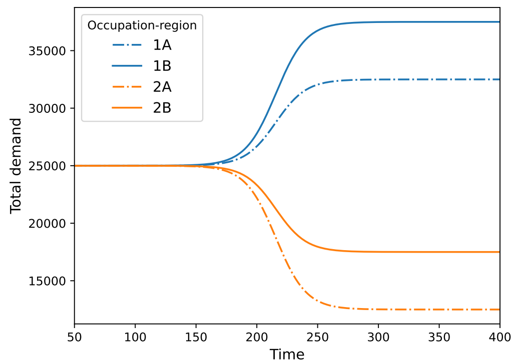
Figure 3.6: **Toy model regional occupational mobility network.** In (a) the regional occupational mobility network for the toy model with two occupations and two regions, in (b) the network summed over occupations (and row-normalised), and in (c) the network summed over regions (and row-normalised).



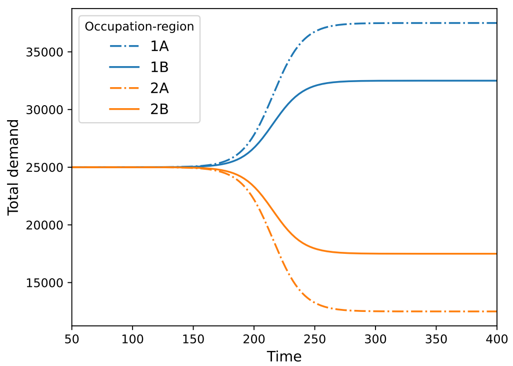
(a) Universal 1



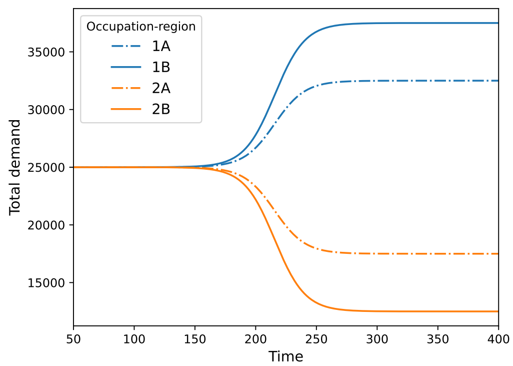
(b) Regional 1A,2A



(c) Regional 1B,2B



(d) Split 1A,2B



(e) Split 1B,2A

Figure 3.7: **Target demand profiles for toy model for the five scenarios.** In (a) Universal 1 without regions then, in (b) Regional 1A,2A with regional heterogeneity, (c) Regional 1B,2B, and in (d) Split 1A,2B has both occupational and regional heterogeneity, as does (e) Split 1B,2A.

different mixture of occupations and regions are favoured. Given that demand always increases in Occupation 1, and decreases in Occupation 2, we label the scenarios with a code that indicates which region in each occupation has the greater final demand. For example, in the Regional 1A,2A scenario, Region A has more demand in both occupations. In the Split 1A,2B scenario, while the aggregate occupation demand is kept the same, Occupation 1 has more demand in Region A but Occupation 2 has more demand in Region B. The five scenarios are

- Universal 1; demand increases in Occupation 1 and decreases in Occupation 2 (Figure 3.7a),
- Regional 1A,2A; demand increases in Occupation 1, decreases in Occupation 2, and Region A has more demand in both occupations (Figure 3.7b),
- Regional 1B,2B; demand increases in Occupation 1, decreases in Occupation 2, and Region B has more demand in both occupations (Figure 3.7c),
- Split 1A,2B; demand increases in Occupation 1, decreases in Occupation 2, and Region A has more demand in Occupation 1, Region B has more demand in Occupation 2 (Figure 3.7d), and
- Split 1B,2A; demand increases in Occupation 1, decreases in Occupation 2, and Region B has more demand in Occupation 1, Region A has more demand in Occupation 2 (Figure 3.7e).

Given the network and the scenarios, Regional 1A,2A is *with* the popularity of the two regions, since Region A has greater in mobility in the network while Regional 1B,2B is *against* the popularity, where more demand is created in the less popular region. In the split scenarios, the greater demand is split between the two regions. In Split 1A,2B, occupation-region pair, 1A, has the greatest demand, this is also the most desired occupation and region in the network, but in Occupation 2, Region B has the more demand. In Split 1B,2A, the opposite pairing have more demand. Occupation 1 has more demand in Region B while Occupation 2 has more demand in Region A. In the results below, we will see how these relationships between the scenarios and the network mobility preferences ease or hinder workers when reacting to the demand changes.

Results

To understand the impact of the demand shocks shown in Figure 3.7, we can look at the change in unemployment in each occupation, or occupation-region, during the shock. In Tables 3.2 and 3.3 we report the percentage change in the average unemployment rate, given by (3.11), from the steady state before the shock to the rate during the shock period.

In Table 3.2, we have the occupation level unemployment rate change. Despite all five scenarios having the same aggregate demand change in the two occupations, there are differences in the aggregate outcomes between them. Specifically, apart from Split 1A,2B, each of the scenarios with regional variation has a more polarised impact on the two occupations than in the Universal 1 scenario; Occupation 1 has a greater decrease in the unemployment rate and Occupation 2 has a greater increase in the unemployment rate. This is due to the increased friction introduced when we consider regional, as well as occupational frictions.

However, this is not the case in Split 1A,2B. In this scenario, the unemployment impacts are less polarised than the Universal 1 scenario, both with a smaller decrease in Occupation 1 and a smaller increase in Occupation 2. This is because the network makes it easier for unemployed workers in any occupation-region pair to take advantage of the increase in demand in pair 1A, therefore reducing the unemployment gains in Occupation 1 that are realised in the other scenarios and reducing the negative unemployment impacts in Occupation 2. What is surprising is that this scenario, Split 1A,2B, has better outcomes than Regional 1A,2A where demand increases most for occupation-region pair 1A, but also 2A. To understand this better, we can look at the disaggregated results.

At the occupation-region level, we see, in Table 3.3, that the disaggregated impacts on Occupation 1 are similar across the two regions for all four scenarios. This is notable since there is a difference in demand of 5,000 between the two regions in each scenario, however not that surprising, because both regions have an increase in demand, in Occupation 1, we just differ which region has the largest increase. We see instead that the main difference comes when comparing Regional 1A,2A and Split 1B,2A with Regional 1B,2B and Split 1A,2B. The two scenarios where unemployment in Occupation 1 decreases the most are the two scenarios when occupation-region pair 2A has the better conditions, rather than 2B. This is caused by the interplay between the target labour demand and the mobility network. Occupation-region 2B is the least popular within the network and so when the favourable demand conditions in Occupation 2 are in Region B, workers cannot react as quickly (from any

	Occupation 1	Occupation 2
Universal 1	-2.89	3.96
Regional 1A,2A	-3.08	4.48
Regional 1B,2B	-3.47	4.82
Split 1A,2B	-2.54	3.91
Split 1B,2A	-3.37	4.92

Table 3.2: **Unemployment rate change for toy model at the occupation level.** Percentage change of the average occupation level unemployment rate.

	Occupation 1		Occupation 2	
	Region A	Region B	Region A	Region B
Regional 1A,2A	-3.32	-3.24	1.87	7.10
Regional 1B,2B	-2.67	-2.87	8.70	0.55
Split 1A,2B	-2.63	-2.89	8.61	-0.19
Split 1B,2A	-3.28	-3.25	1.94	7.79

Table 3.3: **Unemployment rate change for toy model at the occupation-region level.** Percentage change of the average occupation-region level unemployment rate.

origin occupation-region), and the unemployment benefits of the demand increase in Occupation 1 are smaller.

In Occupation 2, however, the unemployment rates in the two regions are not similar. The unemployment rate increases more in the region with the biggest decrease in demand. This is expected but what is of interest is the magnitude of the unemployment increase, and especially that unemployment actually decreases in Region B in Split 1A,2B, in Occupation 2. In this scenario, workers separated by the state-dependent process in pair 2B are able to quickly find employment elsewhere (due to their position in the network).

We can clearly see the effect that including regions has on unemployment outcomes, even when the aggregate demand for the two occupations is the same. These nuanced findings from a simple two-occupation, two-region set up are motivation for the next chapter, where we apply the model to a case study in Brazil.

So far in this toy model, we have only reported the results from running the deterministic approximation. Although we will carry out the main analysis of the two versions of the model in Chapter 5, in Figure 3.8 we show the results from 10 runs of the full agent-based model, the average of these runs, and the deterministic approximation for Regional 1A,2A. Here, for 600 time steps of the model, the full agent-based model takes on average 15.61 seconds to run while the deterministic

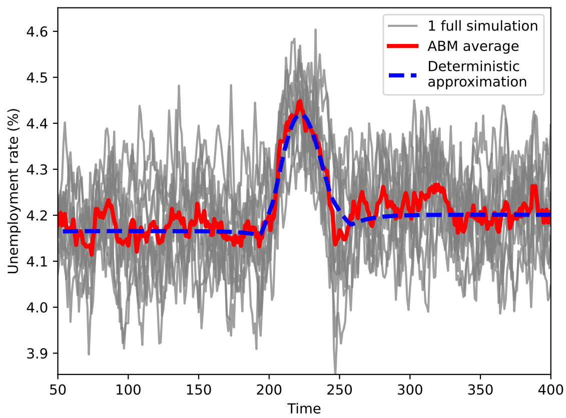


Figure 3.8: **Aggregate unemployment rate for Regional 1A,2A.** The unemployment rate for the Regional 1A,2A scenario with the full simulation and deterministic approximation.

approximation takes 0.74 seconds, more than 20 times faster. This hints at the powerful computational benefits of the deterministic approximation. This motivates the development of this particular agent-based model, with careful attention paid to the approximation for exploitation of this speed-up during future calibration.

3.3 Conclusions

In this chapter, we first introduced the del Rio-Chanona et al. labour market model. We took care to discuss the deterministic approximation which we already see gives important computational benefits, even in the small toy model of Section 3.2.1. In Chapter 5, we relax some modelling assumptions and again calculate the deterministic approximation, showing these computational benefits as well as further discussion of the large population assumption.

We then discussed different regional mobility models before constructing the regional occupational mobility network for Brazil, and using network science techniques to untangle the complex structure. Adding regional mobility into the model gives us the ability to expand our analysis further than a single city. We can now incorporate the heterogeneity of demand shocks across regions, as well as occupations. In the toy model discussed in this chapter, we begin to see the second order effects of heterogeneous demand shocks across regions and occupations. In the next chapter, we apply the model to two growth pathways in Brazil using this regional labour market model. We identify these second order effects, as well as looking at both supply and demand changes through unemployment and vacancy rates.

Chapter 4

Brazil case study

This chapter addresses the questions:

- What regions and occupations are at risk of negative labour market impacts of two growth pathways in Brazil?
- How do those growth pathways impact unemployment and vacancy rates?

This chapter features results that are now superseded in the upcoming paper it is based on titled 'Skill and spatial mismatches for sustainable development in Brazil' with authors: Anna K. Berryman, Fernanda Senra de Moura, Joris Bücker, Pete Barbrook-Johnson, Marek Hanusch, Penny Mealy, J. Doyne Farmer, and R. Maria del Rio-Chanona. While this chapter uses data provided by the World Bank, it does not represent the views of the World Bank, or the countries it represents. This work follows a case study published by the EEIST consortium where Berryman et al. [120] present the scenarios only at the occupation level.

4.1 Introduction

All countries need to undertake some degree of structural change in order to transition to a low carbon economy. This will impact the labour market in different ways. While current studies suggest that more jobs are likely to be created than destroyed, newly created jobs may require different skills [47] or arise in different locations from those that disappear [7, 121]. Such skill or spatial-related mismatches can drive unemployment in some occupations and regions, while leaving unfilled job vacancies in others. An unmanaged transition can lead to declining income and greater inequality and deprivation [48, 122], giving rise to 'green discontent' in neglected regions [123].

Consequently, examining the labour implications of transition policies at occupational and regional levels is paramount.

While Brazil's energy matrix is one of the least carbon-intensive in the world, its emissions related to deforestation and agriculture make it one of the largest greenhouse gas emitters [18]. Brazil's transition to an environmentally sustainable economy, and achieving their 2050 goal of carbon-neutrality [19], depends on shifting away from deforestation-intensive activities, as the Amazon Rainforest is one of the world's largest carbon sinks [17].

Recently, Hlongwane and Khobai found that how this shift and the increased energy demand are managed has potential employment impacts, as well as economic impacts in the BRICS nations [124]. Others have used IO models to look at the impact of specific interventions, such as the adoption of green hydrogen [125] and the Amazon green recovery [126]. In this chapter, we apply the labour market model to two distinct growth pathways with very different environmental outcomes and study their impact on unemployment and vacancies across occupations and regions in Brazil.

We apply our framework to labour demand scenarios taken from two growth pathways [20] for Brazil, which model increased productivity in the manufacturing sector (*Manufacturing growth path*), which leads to lower emissions and less deforestation, and in the agriculture sector (*Agriculture growth path*), which leads to higher emissions and only slightly less deforestation.

We introduce the regional occupational mobility network for Brazil and visualise it for the first time, which reveals strong regional structure. We detail the transition scenarios and introduce the unfilled vacancy rate. We then find that the Manufacturing growth path has lower aggregate unemployment than the Agriculture growth path as well as smaller impacts across occupations and regions. The simple toy model in Section 3.2.1 gives an insight into the possible second order effects we can model with this extended network. To illustrate these second order effects further, we also present the unemployment and vacancy rate impacts for the complete network. Our analysis identifies the skill and spacial mismatches of worker availability and demand during the two growth pathways, and we identify possible inequalities exacerbated by these mismatches, as well as regions and occupations most at risk of negative labour market impacts during the transition scenarios.

4.2 Methods

In this section we introduce the transition scenarios, which define the target demand, $\mathcal{D}^{\dagger}$, in the model. As well as studying the unemployment impacts of the transition scenarios, we study the unfilled vacancy rate. As we discussed in Section 2.3, 'hard-to-fill' vacancies are often only studied through surveys with few quantitative models [21, 22, 54] therefore, we introduce definitions for the unfilled vacancy, and unemployment rates. We can also use the extended regional model to identify both occupational and regional frictions that workers and employers might experience during the transition scenarios, which we quantify with a variance decomposition analysis introduced in this section.

For this analysis, we use the parameter values for δ_u, δ_v, and γ calibrated by del Rio-Chanona et al. [11]. The data available for Brazil did not facilitate re-calibrating the parameters due to the unavailability of job vacancy data. Calibrating these parameters to the Brazil economy would have a quantitative effect on the paper outcomes, but qualitatively the effects tell the same story. The forthcoming paper based on this chapter will have a comprehensive calibration of these three model parameters to Brazil data.

4.2.1 Transition scenarios

We use two scenarios that model growth pathways for Brazil plus a baseline scenario from Ferreira Filho and Hanusch [20]. They use a CGE model that includes a land-use module to look at different scenarios for Brazil from 2018 to 2030, to try to understand how different sectoral productivity pathways relate to Brazil's decarbonisation targets. Due to the importance of deforestation to the green transition in Brazil, the land-use module in the Ferreira Filho and Hanusch analysis gives insights into the levels of deforestation.

Ferreira Filho and Hanusch use the TERM (The Enormous Regional Model [127]) model, tailored to Brazil with a land-use module. This models each region as a separate economy in addition to the land-use change module that uses relative prices of agricultural products and data on how land-use has changed within each region to model the future land-use. The model is run from 2015 to 2018 to reproduce the observed Brazilian economy and then from 2018 to 2030 to understand the impact of the economic interventions. That is, a baseline scenario is run with no change to the economy from 2018 to 2030 and then the scenarios, which we introduce next, are run and compared to the baseline.

We look at two of these scenarios. The *Manufacturing growth path* which assumes an annual 0.5% total factor productivity (TFP) growth in the manufacturing sector across Brazil, and the *Agriculture growth path* which assumes an annual 0.5% TFP increase in the agriculture sector across Brazil. Ferreira Filho and Hanusch conclude that the Manufacturing growth path is associated with a 3.9% increase in GDP, 0.8 million hectares less deforestation, and over $67,000$ kT less CO_2 emissions in Brazil compared to the baseline scenario[1] with no change in productivity. The Agriculture growth path is projected to have an increase of 1.8% to GDP, 0.3 million hectares less deforestation but $18,000$ kT more CO_2 emissions compared to the baseline scenario. The Manufacturing growth path can be interpreted as a shift towards more sustainable development, whereas the Agriculture growth path doubles down on a sector that has historically been Brazil's competitive strength but causes deforestation and emissions.

We convert the labour demand output from the two sectoral scenarios and the baseline scenario to target demand in our model. First, we convert sector level labour demand data to occupation level labour demand for each scenario, using a breakdown of occupational employment per industry in 2018 and 2019. We assume this breakdown to be stable for the duration of the scenario. The CGE model includes an assumption on population growth, but our model can only handle a fixed population size. Therefore, we normalise all three scenarios with the total employment in the baseline scenario.

If total labour demand for occupation-region pair $i\alpha$ in year τ in scenario s is $\mathcal{D}_{i\alpha;\tau,s}$, then the normalised target demand for labour, $\widetilde{\mathcal{D}}_{i\alpha;\tau,s}$, is

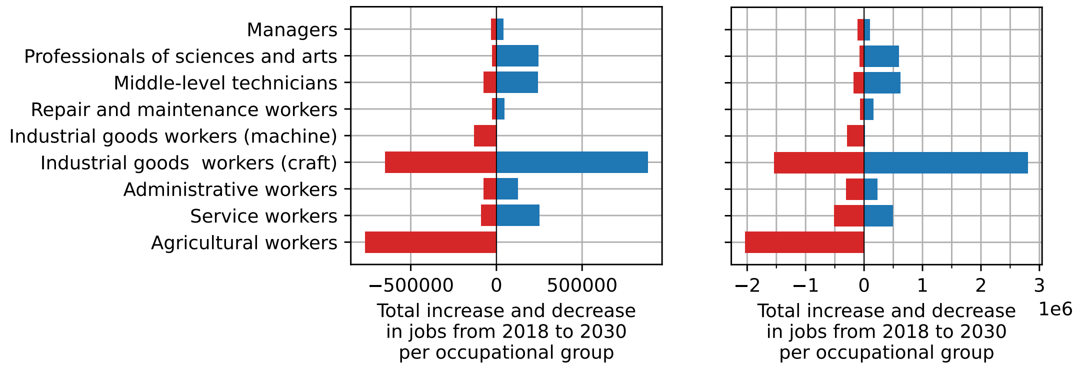
$$\widetilde{\mathcal{D}}_{i\alpha;\tau,s} = \frac{\mathcal{D}_{i\alpha;\tau,s} \cdot \sum_{i,\alpha} D_{i\alpha;2018,baseline}}{\sum_{i,\alpha} D_{i\alpha;\tau,baseline}}. \tag{4.1}$$

Within each year, we interpolate the demand per model timestep linearly to calculate target demand ($\mathcal{D}^{\dagger}_{i\alpha;t}$). In Figure 4.1, we show the number of jobs created and destroyed in each occupational group.[2] These figures show the difference in magnitude of reallocation between the two growth paths as well as the qualitative similarities.

We plot the occupation level demand change on the national occupational mobility network to understand where the occupation level demand shown in Figure 4.1 is increasing and decreasing. In Figure 4.2, we see the national network with nodes

[1]The pathways are presented in relation to a baseline scenario in which the total factor productivity does not change over time. We therefore present our results in relation to this baseline scenario. Any difference in outcomes can be interpreted as the effect of the growth pathway.

[2]That is, we count the total jobs created from 2018 to 2030 in the occupation-region pairs within each occupational group where demand increases and we count the total jobs destroyed in those occupation-region pairs where demand decreases.



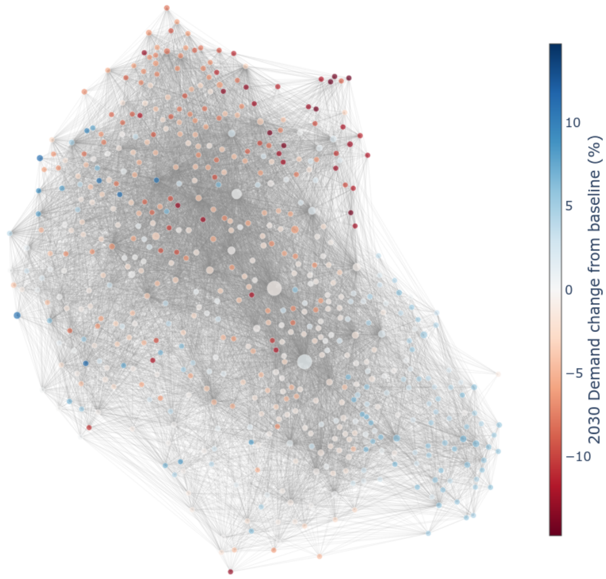
(a) Manufacturing growth path (b) Agriculture growth path

Figure 4.1: **Occupation level labour demand change.** Labour demand change per occupational group for (a) the Manufacturing growth path, and (b) the Agriculture growth path.

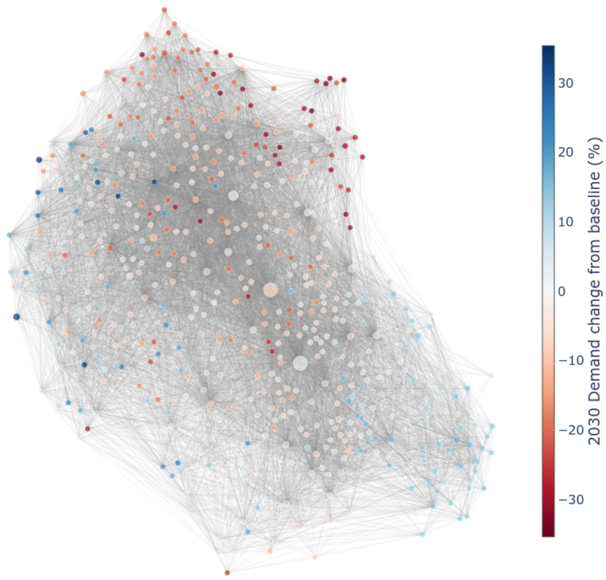
coloured with the 2030 demand for the two growth paths compared to the baseline. In the Agriculture growth path shown Figure 4.2b, we see the occupations where demand is decreasing, mostly in 'Agriculture, forestry, and fisheries' (100% of workers in this occupational group face a demand decrease), and 'Industrial goods and service (machine)' (91% of workers), compared to the baseline and also where demand is increasing, mostly 'Professionals of sciences and arts' (71% of workers see demand increase) and 'Middle-level technicians' (69% of workers). The separation within the network of these two groups of occupations indicates that there might be some mobility frictions that hinder the transition proposed by the TFP growth paths.

There are also areas with a mix of occupations with increasing and decreasing demand. This mix of fortunes for occupations close to one another in the network could offer employment opportunities for the occupations with decreasing demand in neighbouring (or nearby) occupations. This is likely to be more difficult for the occupations in agriculture, forestry and fisheries and machine industrial goods as these occupations are largely surrounded by other occupations with decreasing demand. Conversely, employers in the occupations surrounded by other occupations with increasing demand might find it difficult to fill new positions. The occupational groups represented in this mix of occupations with varying demand change in both growth paths are mainly 'Industrial goods and services (craft)' and 'Repair and maintenance'.

The scenarios and the labour market model we use in this study have different foundational methodologies. We are interested in identifying the impact of frictions workers could face during the growth pathways that are missed in the CGE model,



(a) Manufacturing growth path



(b) Agriculture growth path

Figure 4.2: **Demand change on the occupational mobility network.** 2030 demand change compared to the baseline for (a) the Manufacturing growth path, and (b) the Agriculture growth path.

where these occupation-level frictions are not modelled. This initial study takes the scenarios as exogenous and further work should aim to couple the two models together, essentially embedding the labour market network model into the CGE model so that labour mobility frictions are considered in each timestep of the CGE model.

4.2.2 Unemployment and unfilled vacancy rates

Akin to the previous chapter, the unemployment rate of occupation-region pair $i\alpha$ represents the proportion of workers in occupation-region $i\alpha$ that have not found new employment, after being separated while employed in occupation $i\alpha$. The average unemployment rate over the study period for occupation-region pair $i\alpha$ is defined as

$$u_{i\alpha;average}(T) = \frac{\sum_{t\in T} U_{i\alpha;t}}{\sum_{t\in T}\left(U_{i\alpha;t} + E_{i\alpha;t}\right)}, \tag{4.2}$$

where T is the set of time steps that correspond to the years of transition, 2018 to 2030.

The average x-month vacancy rate $v^{(\geq x)}_{i\alpha;average}$ is the fraction of vacancies over all jobs in occupation-region pair $i\alpha$ that are open for at least x months, on average over the study period, defined as

$$v^{(\geq x)}_{i\alpha;average}(T) = \frac{\sum_{t\in T} V^{(\geq x)}_{i\alpha;t}}{\sum_{t\in T}\left(V_{i\alpha;t} + E_{i\alpha;t}\right)}. \tag{4.3}$$

where $V^{(\geq x)}_{i\alpha;t}$ is the number of vacancies open in occupation-region pair $i\alpha$ at time t that have been open for at least x-months.

Throughout this chapter, we use the average 6-month vacancy rate, but the results are qualitatively unaffected by choosing the duration of an unfilled vacancy to be three, six, or twelve months.

4.2.3 Occupation-related versus region-related network effects

We use variance decomposition to understand the contribution of between-region versus between-occupation variation to the total variation of unemployment and vacancy rate outcomes. Based on Baumgarten, Felbermayr, and Lehwald [128], we decompose the variance of the unemployment rate $u_{i\alpha;average} = u_{i\alpha}$ as

$$\underbrace{\frac{1}{N}\sum_{i,\alpha}(u_{i\alpha} - \bar{u})^2}_{\text{total variance}} = \underbrace{var(\bar{u}_\alpha)}_{\text{between-region}} + \underbrace{var(\bar{u}_i)}_{\text{between-occupation}} + \underbrace{var(u_{i\alpha} - \bar{u}_\alpha - \bar{u}_i)}_{\text{residual variance}}, \tag{4.4}$$

where N is the total number of occupation-region pairs, $\bar{u}_\alpha$ is the mean unemployment rate change within region α, $\bar{u}_i$ is the mean unemployment rate change within occupation i, and $\bar{u}$ is the overall mean unemployment rate change across all occupation-region pairs. The variance decomposition for the vacancy rate follows analogously.

4.3 Results

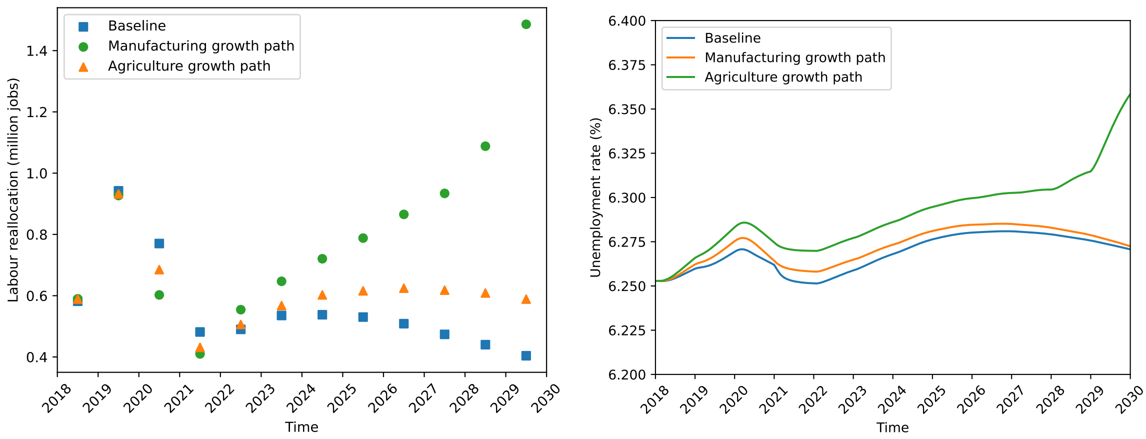
In this section, we explore the potential labour market impacts of the Manufacturing and Agriculture growth paths in Brazil. We show that both pathways have increased unemployment and unfilled vacancy rates compared to the baseline, with the Agriculture growth path facing greater reallocation and higher aggregate unemployment. We investigate how these impacts vary across the wage distribution and across regions and broad occupations. We use the regional occupational mobility network to capture the frictions that workers face when transitioning between occupations, in reaction to the growth pathways, particularly showing the impact of our model when we compare the results using the mobility network and a network with no frictions.

At the occupation level, our results indicate that the lowest-wage occupations are impacted most negatively. Unless the growth path will come with additional help for workers to transition to skills that are in demand, this might lead to growing inequality. Specifically, if the productivity increase in Manufacturing is realised, retraining will be needed for agriculture and repair/maintenance workers. While services, agriculture, and repair/maintenance workers will likely need targeted policies if the Agriculture growth path is followed.

Our results also identify where these retraining policies should be targeted at the region level, and where worker relocation or local development policies need to be considered to enable smoother development in Brazil. In order to follow the Manufacturing growth path successfully, regional policies in Mato Grosso, RSul, and the northern regions should be considered. If the Agriculture growth path is followed, policies need to be aimed at inland, agriculture focused regions such as Roraima and Rondônia.

4.3.1 Inequality of unemployment outcomes

The Manufacturing growth path has better unemployment outcomes compared to the Agriculture growth path. The aggregate average unemployment rate from



(a) Yearly labour reallocation

(b) Aggregate unemployment rate

Figure 4.3: **Brazil demand and unemployment rate.** In (a) the volume of labour reallocation (Definition 6) per year, and in (b) the aggregate unemployment rate between 2018 and 2030.

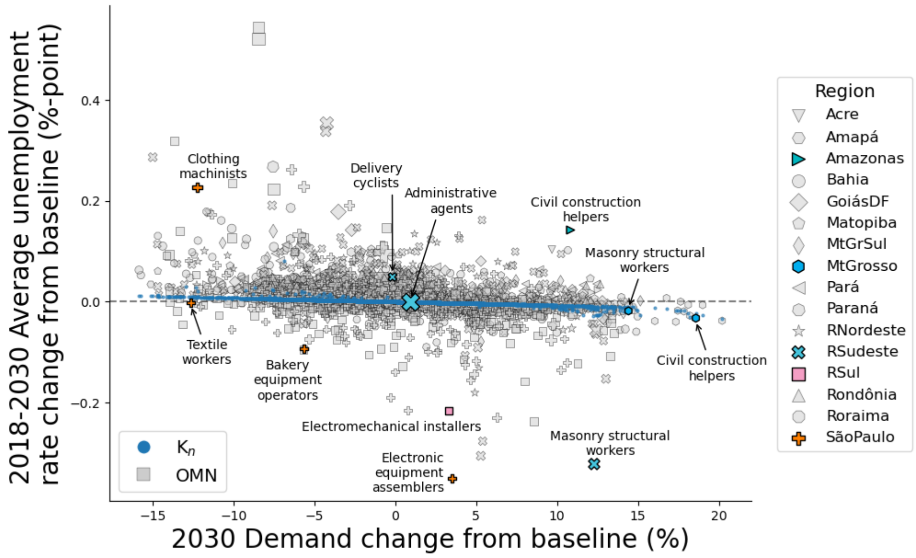
2018 to 2030 in the Manufacturing growth path is 6.273% and 6.290% in the Agriculture growth path, compared to 6.268% in the baseline scenario. The aggregate unemployment rates and yearly labour reallocation (Definition 6) are shown in Figure 4.3. The differences between the aggregate unemployment rates of the two scenarios and the baseline represent somewhat small differences.[3] However, we are interested in the disaggregated variation, at the regional and occupational level. At this level, we will see that the unemployment rate can change up to 1% in a single occupation-region, which would have significant impact to the workers of this occupation-region.

Figures 4.4a and 4.5a show the change in the unemployment rate by occupation-region in relation to the change in labour demand, compared to the baseline and, as we might expect, there is an inverse relationship. In both scenarios about half of the occupation-region pairs face a decrease in labour demand by 2030, compared to the baseline.[4] Our model shows that this translates to 53% of occupation-region pairs being negatively affected by the Manufacturing growth path (i.e. face higher unemployment rates than in the baseline) versus 59% in the Agriculture growth path.[5]

[3] Indeed if we compare the time series using a permutation test, we find that the differences between the aggregate unemployment rates for the scenarios compared to the baseline are not statistically significant.

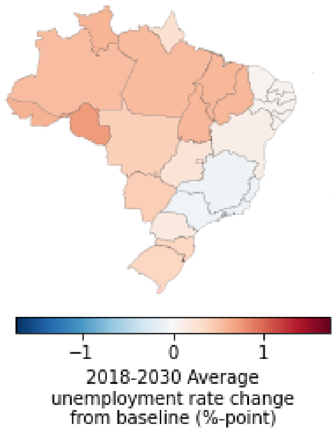
[4] 55.1% in the Manufacturing growth path and 54.8% in the Agriculture growth path.

[5] These occupation-region pairs represent 60% and 56% of workers respectively.



(a)

(b)



(c)

Figure 4.4: **Manufacturing growth path unemployment outcomes.** In (a) the percentage-point change of the 2018–2030 average unemployment rate from the baseline against percentage demand change compared to baseline in 2030, each data point represents an occupation-region pair, shaped by region and sized proportional to employment in 2018 with inset labels identifying different occupations, in (b) the percentage-point change of the 2018–2030 unemployment rate from the baseline for each occupational group and region, and in (c) the regional percentage-point change of the average unemployment rate from baseline.

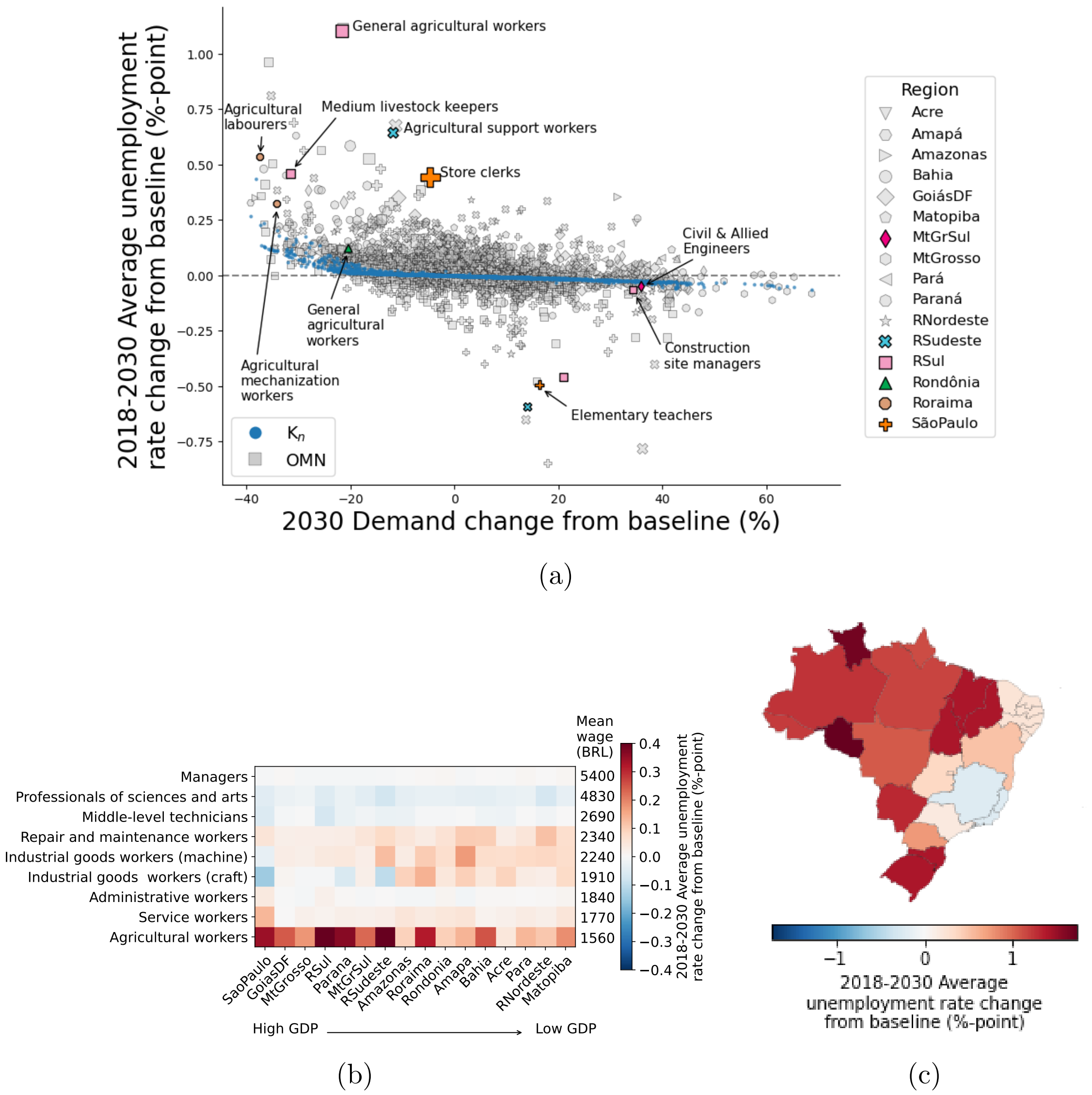


Figure 4.5: **Agriculture growth path unemployment outcomes.** In (a) the percentage-point change of the 2018–2030 average unemployment rate from the baseline against percentage demand change compared to baseline in 2030, each data point represents an occupation-region pair, shaped by region and sized proportional to employment in 2018 with inset labels identifying different occupations, in (b) the percentage-point change of the 2018–2030 unemployment rate from the baseline for each occupational group and region, and in (c) the regional percentage-point change of the average unemployment rate from baseline.

Occupations with the same change in demand face different unemployment outcomes due to their position in the network and the demand profile of their surrounding occupations. For example, in the Manufacturing growth path shown in Figure 4.4a, 'Textile workers' and 'Clothing machinists' in São Paulo have a similar change in demand but different unemployment effects – the former is almost unaffected by the scenario, while the latter is negatively affected. 'Textile workers' in São Paulo are able to mitigate the decrease in demand for workers in their occupation because they are able to find employment elsewhere. In contrast, 'Clothing machinists' in São Paulo are unable to find employment elsewhere, as many of their closest other options also have a decrease in demand. These complex network effects constitute the 'second order effects' that introduce variance in unemployment outcomes.

We also show the results for the case with no worker frictions in Figures 4.4a and 4.5a. Here, we run the model with every occupation-region pair connected to all other occupation-regions with equal weights. Unemployed workers are able to find a new job in any occupation-region pair. In this case, there is very little variation for occupation-regions with the same change in demand and very little impact on the unemployment rate compared to the baseline. This is because unemployed workers do not face any skill or spatial frictions and so can apply for a job in any occupation-region pair, greatly reducing the friction within the model.

In the Agriculture growth path, as shown in Figure 4.5a, 'Store clerks' in São Paulo see a significant increase in the unemployment rate compared to the baseline for a relatively small decrease in overall demand. Workers in this occupation-region have few options to react to this small decrease. This occupation-region has relatively low mobility and the closest options in alternative occupations and regions also experience a decrease in demand. We find similar patterns for 'Agricultural support workers' and other agricultural workers, suggesting that the productivity gains in the agriculture sector may not translate into better outcomes for agriculture workers.

Skill mismatches

High-wage occupations are better positioned to benefit from both transition scenarios. We find that the best outcomes tend to be concentrated in occupations with higher wages in both growth paths, as shown in Figures 4.4b and 4.5b. This can increase inequality in labour market outcomes and, subsequently, income and welfare. In the Manufacturing growth path, 'Middle-level technicians', 'Professionals of sciences and arts', and 'Industrial goods (craft) workers' are the occupational groups most likely to benefit while occupational groups with lower wages face increased unemployment rates. In the Agriculture growth path, the high-wage occupational groups, 'Managers', 'Professionals of sciences and arts', and 'Middle-level technicians', face lower unemployment than in the baseline while 'Agriculture, forestry and fisheries workers', the occupational group with the lowest mean wage, face a large increase in the unemployment rate. This disparity in unemployment outcomes across the wage distribution is caused by low mobility and being surrounded by other decreasing demand occupations, as we saw in Figures 3.4 and 4.2.

Spatial mismatches

At the region level, we find that most regions see an increase in the aggregate unemployment rate compared to the baseline, as shown in Figures 4.4c and 4.5c. In the Manufacturing growth path, we find an increase in the unemployment rate in all regions except São Paulo and RSudeste, and in the Agriculture growth path the unemployment rate increases in all regions except RSudeste. We can also see that unemployment decreases are focused in the coastal and metropolitan regions, while the more rural, Amazonian regions face the largest increases in unemployment. In the Agriculture growth path, Roraima and Rondônia are particularly adversely affected, with the average unemployment rate increasing by 1.66 and 1.72 percentage-points compared to the baseline.

Contribution of skill versus spatial

Unemployment rate changes are much more varied by occupation than by region. This is evidenced in Figures 4.4b and 4.5b where most broad occupational groups have similar unemployment outcomes across all regions. Decomposing the total variance of unemployment outcomes, we find that differences between regions explain little of the heterogeneity in outcomes.[6] Indeed, occupation network effects, as captured by the between-occupation component of the variance, explain about 43% and 46% of the variance in unemployment rates in the Manufacturing and Agriculture growth paths respectively, as shown in Figure 4.6.

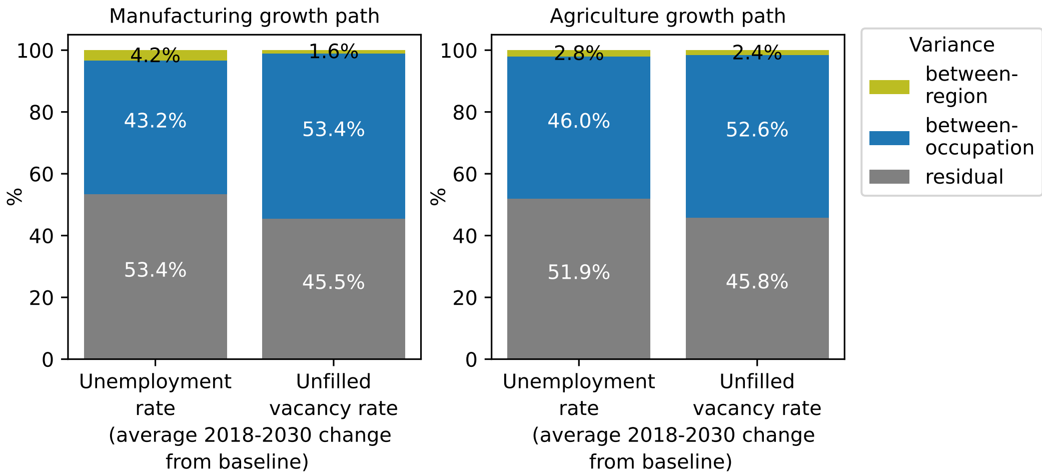


Figure 4.6: **Variance decomposition.** Variance decomposition of the unemployment and 6-month unfilled vacancy rate outcomes for the Agriculture and Manufacturing growth paths.

4.3.2 Unfilled job vacancies risk slowing down development

Higher unemployment rates affect workers whereas an abundance of unfilled vacancies affects employers and risks slowing down the pace of development. We show the average unfilled vacancy rate in Figures 4.7a and 4.8a. In both scenarios there is a positive relationship between changes in labour demand and unfilled vacancy rates. In the Manufacturing growth path we see evidence of friction, with 32% of occupation-region pairs facing higher unfilled vacancy rates than in the baseline, and 39% in the Agriculture growth path.[7]

We also observe second order network effects in the vacancy rate changes for occupation-region pairs with similar demand changes. For example, in the Agriculture growth path, vacancies for 'Systems technicians' in Rondônia remain unfilled because there are few occupation-region pairs connected to this pair in the network from where workers could apply. Whereas 'Telecommunications technicians' in São Paulo have a strong probability of staying in this occupation-region, which results in open vacancies being filled faster, keeping the unfilled vacancy rate low.

[6]It is worth noting that this result is consistent with the fact that the productivity shocks simulated in each scenario are homogeneous across regions and that the between-region components of the demand shocks are relatively low as well.

[7]These occupation-region pairs represent 23% and 20% of workers respectively.

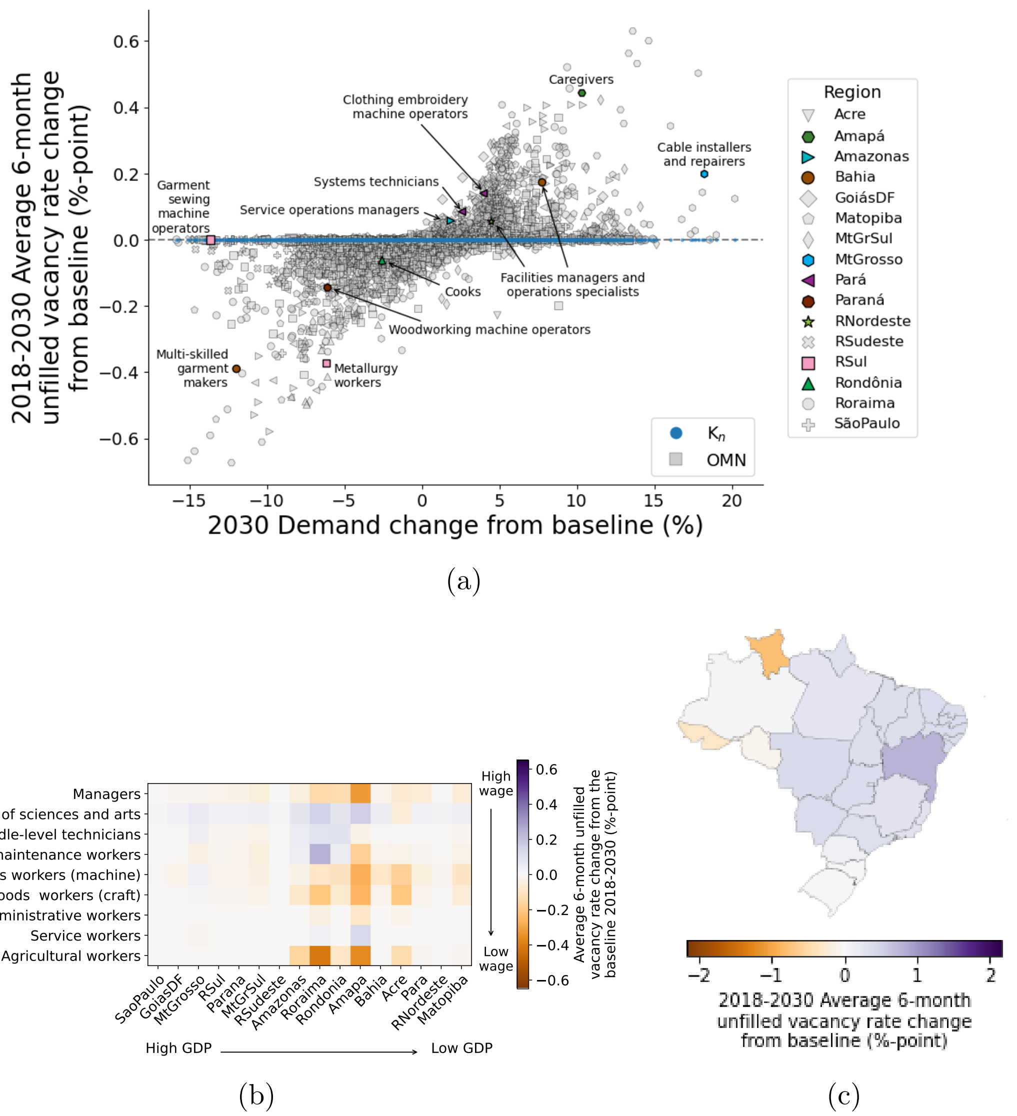
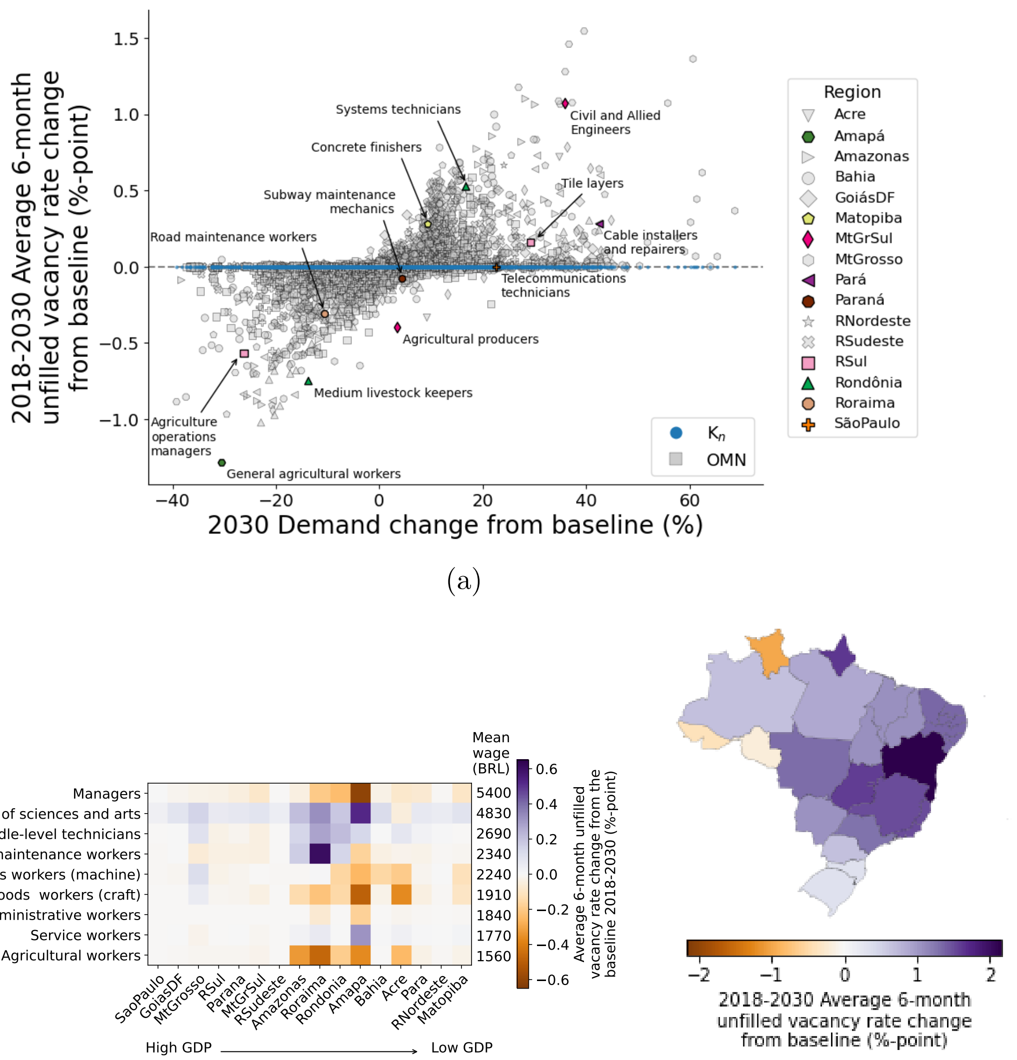


Figure 4.7: **Manufacturing growth path vacancy outcomes.** In (a) the percentage-point change of the 2018–2030 average unfilled vacancy rate from the baseline against percentage demand change compared to baseline in 2030, each data point represents an occupation-region pair, shaped by region and sized proportional to employment in 2018 with inset labels identifying different occupations, in (b) the percentage-point change of the 2018–2030 unfilled vacancy rate from the baseline for each occupational group and region, and in (c) the regional percentage-point change of the average unfilled vacancy rate from baseline.



(a)

(b) (c)

Figure 4.8: **Agriculture growth path vacancy outcomes.** In (a) the percentage-point change of the 2018–2030 average unfilled vacancy rate from the baseline against percentage demand change compared to baseline in 2030, each data point represents an occupation-region pair, shaped by region and sized proportional to employment in 2018 with inset labels identifying different occupations, in (b) the percentage-point change of the 2018–2030 unfilled vacancy rate from the baseline for each occupational group and region, and in (c) the regional percentage-point change of the average unfilled vacancy rate from baseline.

Occupational Group	Mean wage (BRL)	2018-2030 Average unemployment rate change from baseline (%-point)		2018-2030 Average 6-month unfilled vacancy rate change from baseline (%-point)	
		Manufacturing growth path	Agriculture growth path	Manufacturing growth path	Agriculture growth path
Managers	5,401	-0.01	-0.03	-0.90	-1.45
Professionals of sciences and arts	4,828	-0.71	-1.53	1.29	2.73
Middle-level technicians	2,686	-0.45	-0.97	0.18	0.60
Repair and maintenance workers	2,343	0.57	1.29	-1.27	-2.07
Industrial goods and services workers (machine)	2,237	0.03	0.58	-1.40	-1.46
Industrial goods and services workers (craft)	1,914	-0.62	-1.16	-1.32	-1.37
Administrative workers	1,841	0.07	0.13	-0.44	-0.51
Service workers	1,774	0.17	0.71	0.20	1.05
Agricultural, forestry and fisheries workers	1,557	1.94	4.07	-2.42	-4.64

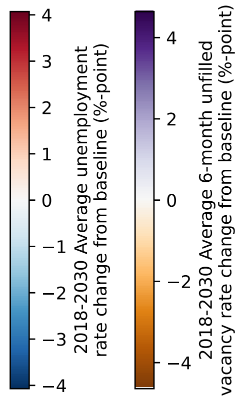


Table 4.1: **Aggregate outcomes.** Average unemployment and unfilled vacancy rate change compared to the baseline between 2018 and 2030 by occupational group.

Skill mismatches

Unfilled vacancy rates are concentrated in professional and technical occupations as shown in Figures 4.7b and 4.8b. Typically, if the unemployment rate increased during the growth path for an occupation, the unfilled vacancy rate is lower (and vice versa). This relationship aligns with expectations: if there were more job openings for unemployed workers to apply for and fill, the unemployment rate could decrease. However, for the Manufacturing growth path, we do find vacancies that are hard to fill in high wage occupations, where the unemployment rates were lower than the baseline. Management occupations face increased unfilled vacancies but saw very little impact on their unemployment rate. This indicates that employers of these highly skilled workers are affected while workers in management occupations are not. Also, service occupations have both increased unemployment and increased unfilled vacancy rates meaning workers and employers of services could be negatively impacted by the Manufacturing growth path. We show the aggregate occupation level outcomes in Table 4.1.

In the Agriculture growth path, firms hiring in the occupational groups with higher wages are likely to experience more friction and unfilled vacancies than in the baseline. It is noteworthy that in the growth path driven by productivity gains in agriculture, the vacancy rates for agricultural workers face the largest decrease compared to the baseline, in line with the higher levels of unemployment likely to be experienced by this group.

Spatial mismatches

At the region level, we see, in Figures 4.7c and 4.8c, that most regions see an increase in the unfilled vacancy rate in both scenarios compared to the baseline. Roraima, Acre, and Rondônia see a decreased rate of unfilled vacancies compared to the baseline in both scenarios, while all other regions have increased rates of unfilled vacancies. Complementary to the region level unemployment rates, the largest increase in unfilled vacancies is seen in the coastal regions, where we forecast better employment outcomes.

4.4 Discussion

Understanding the labour market impacts of different development pathways and policies is vital for policy development and public acceptance. The unemployment and vacancy impacts of development policies depend on labour mobility frictions. We introduced a regional occupational mobility network for Brazil which accounts for frictions between occupations and regions. We linked this network to an external macro-economic model and our labour market agent-based model, to investigate the labour market impacts of two growth pathways in Brazil, one in which total factor productivity increases in the manufacturing sector and another with the productivity increase in the agriculture sector.

We identified occupation-region pairs that may require the attention of policy makers due to increased unemployment or rising unfilled vacancy rates. Occupation-region pairs with high unemployment are a concern due to the obvious social and economic damage that unemployment can cause. Meanwhile, addressing vacancy rates is important to support development and ensure that the right workers with the right skills are available to fulfil demand. In Table 4.2, we identify the top five occupation-region pairs with the largest increase in their unemployment rate, and the top five pairs with the largest increase in the vacancy rate. The disaggregate outcomes of the model give policy makers power to investigate specific occupations and regions in their interest, as well as highlighting which occupations and regions could be most at risk from the two transition scenarios.

In Chapter 3, we visualised the skill and spatial frictions within the labour market of Brazil in a way that has never been done before. We find rich regional structure within the network, with inter-region transitions tightly reflecting the geography of Brazil. This network enables us to show and model key empirical labour market structures that are not captured by the usual modelling approaches. Including labour

	Agriculture	Manufacturing
Key growth path characteristics	5% increase in total factor productivity in agriculture sector	5% increase in total factor productivity in manufacturing sector
Top five occupation-regions with higher unemployment rates	1. Agricultural support workers in RSul 2. Agricultural workers in general in RSul 3. Agricultural workers in general in RSudeste 4. Agricultural support workers in RSudeste 5. Garment sewing machine operators in RSul	1. Agricultural support workers in RSul 2. Agricultural workers in general in RSul 3. Garment sewing machine operators in RSul 4. Garment sewing machine operators in RSudeste 5. Garment sewing machine operators in São Paulo
Top five occupation-regions with higher vacancy rates	1. Programmers, evaluators, and guidance counselors in Amapá 2. Teachers in the area of pedagogical training in higher education in Amapá 3. Elementary school teachers in grades five through eight in Amapá 4. High school teachers in Amapá 5. Higher education professors of architecture and urbanism, engineering, geophysics, and geology in Bahia	1. Teachers in the area of pedagogical training in higher education in Amapá 2. Programmers, evaluators, and guidance counselors in Amapá 3. Civil and Allied Engineers in MtGrosso 4. Elementary school teachers in grades five through eight in Amapá 5. Child, youth, adult and elderly caregivers in Amapá

Table 4.2: **Summary table.** Top five affected occupation-regions.

mobility frictions is important when studying the labour market impacts of economic transitions [5] and our analysis, using over 400 occupations, captures these frictions. Comparing the results of our model with and without labour frictions highlights the contribution of the network to understanding the impact of different growth paths. We identify occupation-region pairs at risk of negative impacts due to the interaction of the scenarios and their position in the regional occupational mobility network.

Our results suggest that both pathways present some risk to workers during the transition period, with better aggregate effects seen in the Manufacturing growth pathway compared to the baseline, than in the Agriculture growth pathway. The pathways also come with challenges for employers, who may struggle to find workers with the appropriate skills in their region. Therefore, training and labour mobility policies will be important to support sustained growth. The vacancy analysis could be particularly helpful to policy makers and employers as we quantify and identify the occupations and regions most at risk of unfilled vacancies, unlike much of the existing research, which largely relies on employer surveys [21, 22].

Contributing to these challenges is the difference between the regions with increased unemployment and those with increased unfilled vacancies. The regions with increased unemployment are coastal, and have historically been more attractive to workers than the interior. Carvalho and Inácio de Moraes [129] find that the Brazilian Coastal and Marine economies represented 19.0% of Brazil GDP in 2015, with much of this activity dominated by services. The regional disconnect between increased unemployment and increased unfilled vacancy rates is in line with Lim et al. [7] who find that, in the US, the new green jobs being created are not necessarily in the same location as the brown jobs. As distributional disparities such as these could fuel discontent [123], attending to these regional inequalities could be crucial to the political sustainability of any transition path.

Another possible inequality that could be exacerbated by these growth pathways is across the wage distribution. We find that workers at the bottom end of the wage distribution could be most adversely affected and therefore require assistance from policy makers if these growth pathways are realised. This possible inequality of sustainable development pathways has been identified before [8]. Our results are further evidence that retraining policies are very likely to be needed to mitigate the negative impact of transition scenarios on low-wage workers [44].

This work has presented one of the most granular labour market models available but leaves plenty of room for future work. The method could be extended by more tightly coupling our model with an external macro-economic model [130, 131] to allow for feedback between the industrial output and labour market frictions. This would create the possibility of quantifying a slowdown in growth relating to unfilled vacancies. Our model could act as a marketplace for workers and employers to fulfil the labour demand needed by the macro model, and a 'realised productivity' would be able to be calculated, with the consideration of labour frictions. The flexibility of our model means that, given the right data, we could couple it with a macro model with occupations and regions, or extend the mobility network framework to further layers such as industries, or even firms.

The worker mobility data we use covers the entire formal labour market in Brazil, about 67% [109] of the total labour force. Although the dataset we use accurately reflect nationwide demographics such as gender and age [110], the lack of informal transitions might be particularly consequential for occupations with high levels of informality (the Services, Agricultural, and Industrial craft occupational groups) which are under-represented in our data. Additional data, such as the PNAD survey, could be used to understand and correct the sample bias in the RAIS data [132].

Nonetheless, our method enables us to identify the labour market outcomes across regions and occupations with high granularity. Our results for Brazil echo those of Bergant, Mano, and Shibata [133] for the US: a green transition may not dramatically shift employment geographically but requires different skills. We contribute a new model that provides insights that policymakers need to account for when thinking about future developmental pathways; in particular, how workers at the bottom of the wage distribution could be most adversely affected and how skill shortages at the top end of the wage distribution could slow down progress. We find that skills development programs and policies should be put in place sooner rather than later.

4.5 Conclusions

In this chapter, we applied the regional labour market model to two growth paths for Brazil, identifying wage inequality in the unemployment impacts and unfilled job vacancies that could slow down the transition. This case study highlights the flexibility of applying our agent-based model to new countries and new economic transitions. We also discussed an important pathway for further work, to couple our model with a macro-economic model. This would allow for feedback between the scenarios and our labour market model to see how the frictions we model might slow down the economic transitions in the macro-economic model. Next, we extend the labour market model by relaxing some key modelling assumptions.

Chapter 5

A triad of model extensions

This chapter addresses the questions:

- How can the stochastic model framework be extended further to capture more labour market mechanisms?
- Can we still derive a deterministic approximation for this added complexity?

5.1 Introduction

As discussed in Section 2.5, such stochastic labour market models based on networks have been developed over the past few years. The seminal model of Guerrero and López [83] and López, Guerrero, and Axtell [134] was the first to explore how the structure of the mobility network affects unemployment, at the firm-level. Guerrero and López are able to derive an analytic steady-state solution which confirms that the structure of the underlying network affects unemployment outcomes for workers. They find that workers at firms surrounded by other firms with low hiring policies face longer periods of unemployment than those workers at firms with more favourable neighbours, evidence for the need to consider the structure of the underlying mobility network when studying unemployment.

Del Rio-Chanona et al. [11] took these models a step further, now accounting for spontaneous separation and state-dependent separation, as well as tracking open vacancies, in place of using hiring rates. Del Rio-Chanona et al. also find that their model is dependent on the structure of the underlying network however the further model complexity has the consequence that the authors are unable to derive the analytic steady-state solutions as Guerrero and López did. Instead, as we have discussed and leveraged in this thesis, del Rio-Chanona et al. derive a deterministic approximation to the stochastic model by taking expectations of each stochastic process.

In this chapter, we extend the del Rio-Chanona model by considering on-the-job search, wage-driven dynamics, and multiple applications. We formalise the derivation of the number of filled vacancies in the original model, recognising that this follows the so-called 'Occupancy distribution' [135]. This gives us the full probability distribution for the number of job offers at each timestep, not just the mean as given by (3.21). This enables us to calculate the approximation for multiple applications where, unlike the single application case, offers are rejected, conditional on the number of matches, and the number of job offers is no longer equal to the number of filled vacancies (if a worker receives multiple job offers, they will reject all but one, leaving the remaining vacancies open until the next time step).

We also add on-the-job search to the model, so employed workers can also apply for new jobs. Wage pressure is also added simply by adjusting the attraction factors of destination occupations to account for workers' desire for higher wages. Deriving the deterministic approximation for the model with these three extensions will ensure the del Rio-Chanona model can continue to provide fast empirical analysis.

5.2 On-the-job search

On-the-job search has been shown to account for a large proportion of worker transitions; Fallick and Fleischman [24] use the Current Population Survey (CPS) from 1994 to 2003, and find that during this time, almost two fifths of new jobs were filled by workers moving from employment to new employment and Faberman et al. [23] find that around 20% of employed workers in the Survey of Consumer Expectations reported searching for a job in the four weeks prior to the survey. More recently, Fujita, Moscarini, and Postel-Vinay [136] use a similar analysis of the CPS. Unable to repeat the exact analysis from Fallick and Fleischman due to missing responses to specific interview questions in 2007, the authors propose a method to impute the missing answers. With this imputed data, they confirm the prevalence of employer-to-employer transitions, showing that workers transition from job-to-job with a monthly probability of about 2.5%.

As well as representing a large number of transitions, on-the-job search has been shown to be a driver of the shift in the Beveridge curve. Fujita et al. [136] show that employer-to-employer transitions usually follow the business cycle, increasing during an upturn and decreasing during a downturn; and they show the beginnings of the COVID-19 impact on job-to-job transition. Joyce et al. [137] see that during the COVID-19 pandemic, the job finding rate for unemployed workers fluctuated but returned to pre-COVID levels by 2021, while the number of employer-to-employer transitions increased. They attribute the increase in vacancy postings to the increased employer-to-employer transitions, rather than a real increase in labour demand. Shimer [138] argues that the standard search and matching model [31] cannot account for the cyclical behaviour of unemployment. Martin and Pierrard [25] show how a carefully calibrated model with on-the-job search can in fact capture the cyclical volatility of unemployment that Shimer argues is missing. Eeckhout and Lindenlaub [139] also find that employed workers engaging in on-the-job search can drive fluctuations in vacancy postings, and therefore drive the Beveridge Curve, even in the absence of an external shock. With this in mind, we add on-the-job search to our labour market model.

In the context of the agent-based model, employed workers will send an application with some probability, with which they can apply for any vacancy they choose. Employed workers in occupation i apply to occupation j with the same probability as unemployed workers, as detailed in (3.6). Unemployed workers apply for jobs in the same way as before. If an employed worker in occupation i is successful, they transition straight into employment in occupation j. As Joyce et al. [137] highlight, many vacancy postings are created to fill the position left behind by an employer-to-employer transition. However, in our model, we do not directly open a vacancy for every employed worker who transitions out of occupation i. Instead, if there is still more demand for workers in occupation i at the next timestep, the probability that a vacancy is opened will increase (since realised demand has decreased due to this worker moving from occupation i to j), following (3.8).[1]

The extended model can be used with the regional occupational mobility network for occupation-region pairs however, from here on, we only use occupations to keep our notation simple (from $F_{i\alpha,j\beta;t}$, to $F_{ij,t}$). The flow of workers $F_{ij,t+1}$ now has two components; the flow of workers previously *unemployed* in occupation i to employment in occupation j, $F^{(u)}_{ij,t+1}$, and the flow of workers previously *employed* in occupation i to employment in occupation j, $F^{(e)}_{ij,t+1}$. We again liken this process to the urn problem. Now an employed worker has a ball with their name on it, with probability λ, and they place their ball in the bin representing occupation j with the

[1] This should be kept in mind when calibrating the parameter γ, the rate at which vacancies are opened due to the market adjusting towards target demand, to ensure that the delay between an employed worker leaving their job and a new vacancy being opened (if there is the demand for it) is not too long.

same probability as the unemployed workers. Once all the balls have been placed, selection follows as before. Therefore, $F^{(u)}_{ij,t+1}$ and $F^{(e)}_{ij,t+1}$ are both stochastic variables that follow a multinomial distribution, which we make explicit in the next section.

The governing equations, originally in (3.1)−(3.3), for the full stochastic agent-based model are now

$$E_{i,t+1} = E_{i,t} - \underbrace{B_{i,t+1}}_{\substack{\text{separated}\\ \text{workers}}} + \underbrace{\sum_j F_{ji,t+1}}_{\text{hired workers}} - \underbrace{\sum_j F^{(e)}_{ij,t+1}}_{\substack{\text{hired}\\ \text{employed}\\ \text{workers}}}, \tag{5.1}$$

$$U_{i,t+1} = U_{i,t} + \underbrace{B_{i,t+1}}_{\substack{\text{separated}\\ \text{workers}}} - \underbrace{\sum_j F^{(u)}_{ij,t+1}}_{\substack{\text{hired}\\ \text{unemployed}\\ \text{workers}}}, \tag{5.2}$$

$$V_{i,t+1} = V_{i,t} + \underbrace{C_{i,t+1}}_{\substack{\text{opened}\\ \text{vacancies}}} - \underbrace{\sum_j F_{ji,t+1}}_{\text{hired workers}}. \tag{5.3}$$

The change in employment for each occupation is equal to the number of workers hired into the occupation (who were previously unemployed *or* employed), minus the number of *employed* workers moving to a new job, and minus the number of workers separated. The change in unemployment is equal to the number of workers separated minus the number of *unemployed* workers hired from this occupation. The change in the number of vacancies is equal to the number of opened vacancies minus the number of vacancies filled (by unemployed *and* employed workers).

5.2.1 Deterministic approximation for large populations

As we detailed in Section 3.1.2, we can use the law of large numbers and Taylor expansion to condense the full agent-based model into a set of deterministic equations that we can run quickly. In this section, we show that the approximation still holds when on-the-job search is added.

Let the probability that an employed worker is looking for a job be λ. Let $S^{(u)}_{ij,t+1}$ be the random variable for the number of applications submitted by unemployed workers in occupation i to occupation j, and $S^{(e)}_{ij,t+1}$ be the number submitted by employed workers. Recall $q_{ij,t+1}$ is the conditional probability that a worker unemployed in occupation i applies to occupation j, given by (3.6). This is the same for unemployed workers and employed workers, given they are looking for a job while in occupation i.

Then, as in (3.17), the expected number of applications sent by unemployed workers is

$$\bar{s}^{(u)}_{ij,t+1} = q_{ij,t+1}\bar{u}_{i,t}, \tag{5.4}$$

and the expected number of applications sent by employed workers is

$$\bar{s}^{(e)}_{ij,t+1} = \lambda q_{ij,t+1}\bar{e}_{i,t}. \tag{5.5}$$

Therefore, for fixed i, $S_{ij,t+1}$ is the sum of two multinomial distributions; one with $U_{i,t}$ trials and success probability $q_{ij,t+1}$ for $j \in I$, and one with $\lambda E_{i,t}$ trials and the same success probabilities, and the total expected number of applications submitted from occupation i to occupation j is

$$\begin{aligned} \bar{s}_{ij,t+1} &= \bar{s}^{(u)}_{ij,t+1} + \bar{s}^{(e)}_{ij,t+1} \\ &= q_{ij,t+1}\bar{u}_{i,t} + \lambda q_{ij,t+1}\bar{e}_{i,t}. \end{aligned} \tag{5.6}$$

The flow of workers $F_{ij,t+1}$ is now equal to the sum of the flow of unemployed workers from occupation i to occupation j, $F^{(u)}_{ij,t+1}$, and the flow of employed workers from occupation i to occupation j, $F^{(e)}_{ij,t+1}$. $F^{(u)}_{ij,t+1}$ is a multinomial where the number of trials is the number of unemployed workers in occupation i applying to occupation j, $S^{(u)}_{ij,t+1}$, and the probabilities are given by the probability of any application to occupation j being successful, $P_{j,t+1}$. $F^{(e)}_{ij,t+1}$ is a multinomial with $S^{(e)}_{ij,t+1}$ trials and the same success probabilities, $P_{j,t+1}$, as they do not depend on the origin of the application. Recall $P_{j,t+1}$ defined above in (3.19), is given by

$$P_{j,t+1} = \frac{M_{j,t+1}}{S_{j,t+1}}. \tag{5.7}$$

The number of filled vacancies, $M_{j,t+1}$, is the number of job openings that receive at least one application, this is actually the so-called *occupancy distribution* [135].

Although the occupancy distribution has many applications, it is not a widely known distribution, indeed it still does not have its own Wikipedia page.[2] The occupancy distribution, formalised by O'Neill in 2021, tells us how many bins will be occupied when we randomly place a certain number of balls. This is somewhat similar to the birthday problem, where one wants to calculate the probability that at least two people in a room share a birthday, and the coupon collector problem, where one wants to calculate the number of coupons needed to have collected them all. The occupancy distribution has been used for many different practical applications in

[2]Search undertaken on 7th January, 2025.

biology, vocabulary, and in a particular case of the lottery when, in 1982, a contestant won two out of 500 cars with a single ticket, in a pool of 2.4 million tickets [140].

Following O'Neill [135], consider the case where we have $y \in \mathbb{N}$ bins and we allocate $x \in \mathbb{N}$ balls to these bins with probability vector $\mathbf{p} \equiv (p_1, \ldots, p_y)$. The occupancy number $1 \leq K \leq min(x, y)$ is the number of bins occupied by the balls after random allocations, $K \equiv K(x, \mathbf{p})$. For the discrete uniform case, $\mathbf{p} \equiv \left[\frac{1}{y}, \ldots \frac{1}{y}\right]$ then denote $K \equiv K(x, y)$. The probability mass function is defined by

$$\mathrm{Occ}(k|x, y) = \frac{(y)_k \cdot S(x, k)}{y^x} \tag{5.8}$$

for all integers $1 \leq k \leq min(x, y)$, where $(y)_k = \prod_{\omega=0}^{k-1}(y - \omega)$ are the falling factorials and $S(x, k)$ are the Stirling numbers of the second kind. This can be rewritten as

$$\mathrm{Occ}(k|x, y) = \frac{1}{y^x}\binom{y}{k}\sum_{\omega=0}^{k}\binom{k}{\omega}(-1)^{\omega}(k - \omega)^x, \tag{5.9}$$

which we define as $\phi(k|x, y) \coloneqq \mathrm{Occ}(k|x, y)$, and the expectation is given by

$$\mathbb{E}(K) = y\left[1 - \left(1 - \frac{1}{y}\right)^x\right]. \tag{5.10}$$

Returning to our case, we have $V_{j\beta;t}$ bins (vacancies) and $S_{j\beta;t+1}$ balls (applications) to be placed, for occupation-region $j\beta$. The number of vacancies in occupation-region $j\beta$ that receive at least one application follows the distribution

$$M_{j\beta;t+1}|(\mathbf{U}_t, \mathbf{V}_t m\mathbf{E}_t) \sim \mathrm{Occ}\left(k|x = S_{j\beta;t+1}, y = V_{j\beta;t}\right). \tag{5.11}$$

Therefore, the expected value for the flow of unemployed workers, for fixed i, is

$$\bar{f}^{(u)}_{ij,t+1} = \mathbb{E}[S^{(u)}_{ij,t+1}P_{j,t+1}|\mathbf{u}_t, \mathbf{v}_t, \mathbf{e}_t], \tag{5.12}$$

and the expected value for the flow of employed workers is

$$\bar{f}^{(e)}_{ij,t+1} = \mathbb{E}[S^{(e)}_{ij,t+1}P_{j,t+1}|\mathbf{u}_t, \mathbf{v}_t, \mathbf{e}_t]. \tag{5.13}$$

Without loss of generality, we prove that

$$\bar{f}^{(u)}_{ij,t+1} \approx \bar{s}^{(u)}_{ij,t+1}\frac{\bar{v}_j\left(1 - e^{-\bar{s}_{j,t+1}/\bar{v}_{j,t}}\right)}{\bar{s}_{j,t+1}}, \tag{5.14}$$

for the flow of unemployed workers and it follows that

$$\bar{f}^{(e)}_{ij,t+1} \approx \bar{s}^{(e)}_{ij,t+1}\frac{\bar{v}_j\left(1 - e^{-\bar{s}_{j,t+1}/\bar{v}_{j,t}}\right)}{\bar{s}_{j,t+1}}, \tag{5.15}$$

for the flow of employed workers.

Proof. As before, we define

$$S^{(u)}_{j\setminus i,t+1} = \sum_{k\neq i} S^{(u)}_{kj,t+1}. \tag{5.16}$$

Then, the multivariate form for the flow of unemployed workers is defined as

$$F^{(u)}_{ij,t+1} = g(S^{(u)}_{ij,t+1}, S^{(u)}_{j\setminus i,t+1}) = S^{(u)}_{ij,t+1} \frac{\bar{v}_{j,t}(1 - e^{-(S^{(u)}_{ij,t+1}+S^{(u)}_{j\setminus i,t+1}+S^{(e)}_{j,t+1})/\bar{v}_{j,t}})}{S^{(u)}_{ij,t+1} + S^{(u)}_{j\setminus i,t+1} + S^{(e)}_{j,t+1}}. \tag{5.17}$$

Now, again, $S^{(u)}_{ij,t+1}$ and $S^{(u)}_{j\setminus i,t+1}$ are independent. The applications sent from unemployed workers in occupation i, $S^{(u)}_{ij,t+1}$, are sent independently of the applications sent from unemployed workers in all other occupations, $S^{(u)}_{kj,t+1}$ for $k \neq i$.

Since $S^{(u)}_{ij,t+1}$ and $S^{(u)}_{j\setminus i,t+1}$ are independent, we can perform a Taylor expansion of g around the expected values for $S^{(u)}_{ij,t+1}$ and $S^{(u)}_{j\setminus i,t+1}$, analogous to (3.28). Let $x = S^{(u)}_{ij,t+1}$, $y = S^{(u)}_{j\setminus i,t+1}$, $v = \bar{v}_{j,t}$ and $s = S^{(e)}_{j,t+1}$. Let $h(x,y) = g(S^{(u)}_{ij,t+1}, S^{(u)}_{j\setminus i,t+1})$, then using the independence of x and y,

$$\begin{aligned}
&\bar{h}(x,y)\\
&= h(\mu_x,\mu_y) + \sigma_x^2\Bigg[\left(\frac{v\mu_x}{(\mu_x+\mu_y+s)^3} - \frac{v}{(\mu_x+\mu_y+s)^2}\right)(1-e^{-(\mu_x+\mu_y+s)/v})\\
&\quad + \left(\frac{1}{(\mu_x+\mu_y+s)} - \frac{\mu_x}{(\mu_x+\mu_y+s)^2} - \frac{1}{2}\frac{\mu_x}{v(\mu_x+\mu_y+s)}\right)e^{-(\mu_x+\mu_y+s)/v}\Bigg]\\
&\quad + \sigma_y^2\Bigg[\frac{v\mu_x}{(\mu_x+\mu_y+s)^3}(1-e^{-(\mu_x+\mu_y+s)/v})\\
&\quad - \left(\frac{\mu_x}{(\mu_x+\mu_y+s)^2} + \frac{1}{2}\frac{\mu_x}{v(\mu_x+\mu_y+s)}\right)e^{-(\mu_x+\mu_y+s)/v}\Bigg] + \mathcal{O}(\mathcal{L}^{-1}).
\end{aligned} \tag{5.18}$$

The same law of large numbers holds as (3.29). Since s scales linearly with $\mathcal{L}$, as do μ_x, μ_y, σ_x^2, σ_y^2, and v, again the second-order derivative terms are of the order of a constant and the first term scales with $\mathcal{L}$. Therefore, in the limit when $\mathcal{L}$ is large, $\bar{h}(x,y) = h(\mu_x,\mu_y)$ and so

$$\bar{f}^{(u)}_{ij,t+1} = \bar{s}^{(u)}_{ij,t+1} \frac{\bar{v}_{j,t}\left(1 - e^{-\bar{s}_{j,t+1}/\bar{v}_{j,t}}\right)}{\bar{s}_{j,t+1}}, \tag{5.19}$$

and it follows that

$$\bar{f}^{(e)}_{ij,t+1} = \bar{s}^{(e)}_{ij,t+1} \frac{\bar{v}_{j}\left(1 - e^{-\bar{s}_{j,t+1}/\bar{v}_{j,t}}\right)}{\bar{s}_{j,t+1}}. \tag{5.20}$$

□

Therefore the deterministic approximation is given by

$$
\begin{aligned}
\bar{f}_{ij,t+1} &= \bar{f}^{(u)}_{ij,t+1} + \bar{f}^{(e)}_{ij,t+1} \\
&= \bar{s}^{(u)}_{ij,t+1} \frac{\bar{v}_{j,t}(1 - e^{-\frac{\bar{s}_{j,t+1}}{\bar{v}_{j,t}}})}{\bar{s}_{j,t+1}} + \bar{s}^{(e)}_{ij,t+1} \frac{\bar{v}_{j,t}(1 - e^{-\frac{\bar{s}_{j,t+1}}{\bar{v}_{j,t}}})}{\bar{s}_{j,t+1}},
\end{aligned}
\tag{5.21}
$$

and (3.33)–(3.35) become

$$
\bar{e}_{i,t+1} = \bar{e}_{i,t} - (\delta_u \bar{e}_{i,t} + (1 - \delta_u)\gamma \max\{0, \bar{d}_{i,t} - \mathcal{D}^{\dagger}_{i,t}\}) + \sum_j \bar{f}_{ji,t+1} - \sum_j \bar{f}^{(e)}_{ij,t+1}, \tag{5.22}
$$

$$
\bar{u}_{i,t+1} = \bar{u}_{i,t} + (\delta_u \bar{e}_{i,t} + (1 - \delta_u)\gamma \max\{0, \bar{d}_{i,t} - \mathcal{D}^{\dagger}_{i,t}\}) - \sum_j \bar{f}^{(u)}_{ij,t+1}, \tag{5.23}
$$

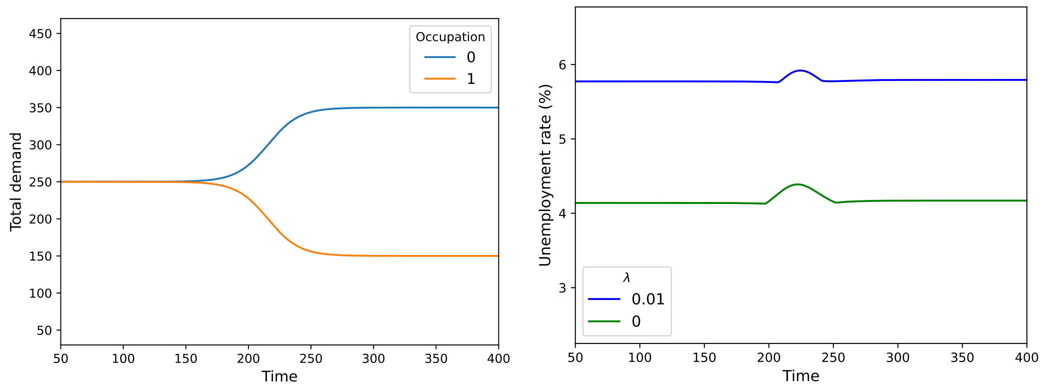
$$
\bar{v}_{i,t+1} = \bar{v}_{i,t} + (\delta_v \bar{e}_{i,t} + (1 - \delta_v)\gamma \max\{0, \mathcal{D}^{\dagger}_{i,t} - \bar{d}_{i,t}\}) - \sum_j \bar{f}_{ji,t+1}. \tag{5.24}
$$

Separation and vacancy opening is unchanged, therefore, $\bar{b}_{i,t+1}$ and $\bar{c}_{i,t+1}$ are as before in (3.13) and (3.14). This concludes the full model with the on-the-job search extension which we now investigate with a toy model.

5.2.2 Toy model

To illustrate the extended model and the deterministic approximation, we run the toy model scenario from Chapter 3 for different values of $\mathcal{L}$. We use the Universal 1 scenario with a total population of 500, 50,000, and 5 million and compare the deterministic approximation with 10 runs of the full agent-based model. As in the previous toy model, we initially split the workers evenly between the two occupations, keeping the set-up the same and only changing the total number of workers. We set our new parameter for the probability that an employed worker applies for a new job, λ, to 0.01. In Section 6.3.1, we investigate the impact of changing this parameter.

We use the Universal 1 scenario as in Section 3.2.1, with the target demand profile for $\mathcal{L} = 500$ shown in Figure 5.1a. In Figure 5.1b, we show the aggregate unemployment rate for $\lambda = 0$, corresponding to the original model, and the extended model with on-the-job search where employed workers look for a new job with probability $\lambda = 0.01$. We see that the unemployment rate during the steady state, before the demand change, is greater when employed workers also engage in job search. This is expected. Even though the probability that an employed worker engages in a job search is only 1%, this equates to roughly 47 employed workers in the steady state. The 30 unemployed workers are then outnumbered in the application process by employed workers, reducing the employment opportunities for unemployed workers



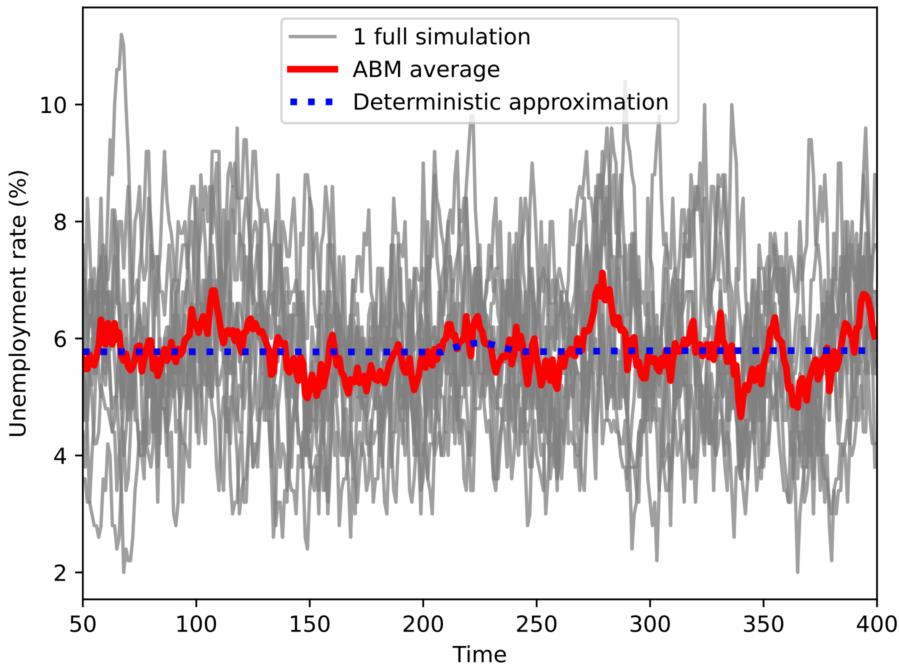
(a) Target demand, $\mathcal{D}^\dagger$ (b) Aggregate unemployment rate

Figure 5.1: **On-the-job search toy model shock demand and unemployment rate.** In (a) the occupation level demand profiles for $\mathcal{L} = 500$, and in (b) the aggregate unemployment rate with, and without, on-the-job search.

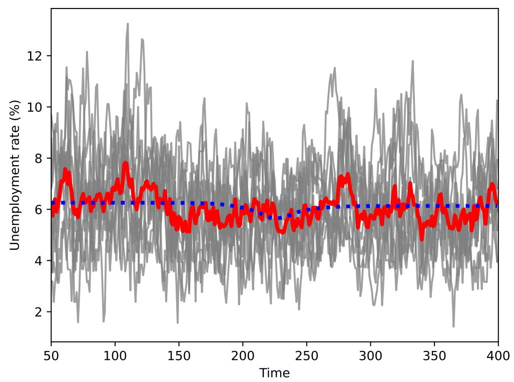
and therefore raising the unemployment rate. Del Rio-Chanona et al. [11] show that the fixed point equations for constant demand depend on the parameters δ_u and δ_v, the target demand, and the network structure. While we omit the derivation of the fixed point equations for the extended deterministic model, our simulations indicate that they also depend on the parameter λ.

Now we run the deterministic approximation, and 10 runs of the full agent-based model for different values of $\mathcal{L}$. We show the results for $\mathcal{L} = 500$, $\mathcal{L} = 50{,}000$, and $\mathcal{L} = 5{,}000{,}000$ in Figures 5.2, 5.3, and 5.4, respectively. We can see, as $\mathcal{L}$ increases, each run of the agent-based model follows the deterministic approximation more closely. In Table 5.1, we show the average standard error between the 10 agent-based model simulations. We also report the average difference between the deterministic approximation and the mean across the 10 agent-based model runs. We can see that the error decreases as $\mathcal{L}$ increases.

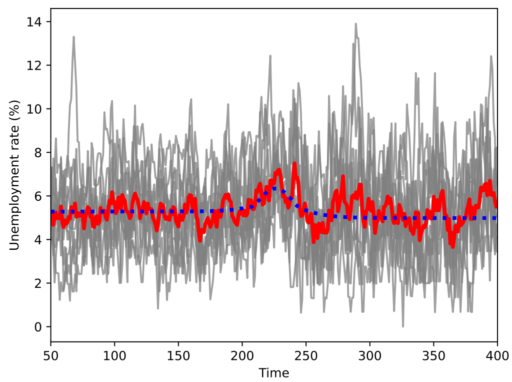
We record the time taken for each approximation and the average agent-based model simulation time, reported in Table 5.2. We can see the huge benefit of having the deterministic approximation over needing to run the full agent-based model. For $\mathcal{L} = 50{,}000$, the approximation is about 17 times faster than one run of the agent-based model, while for $\mathcal{L} = 5{,}000{,}000$, the approximation is nearly 3,000 times faster. Because of the vectorisation in the code, the deterministic approximation time only depends on the number of occupations and does not rise with $\mathcal{L}$.



(a) Aggregate

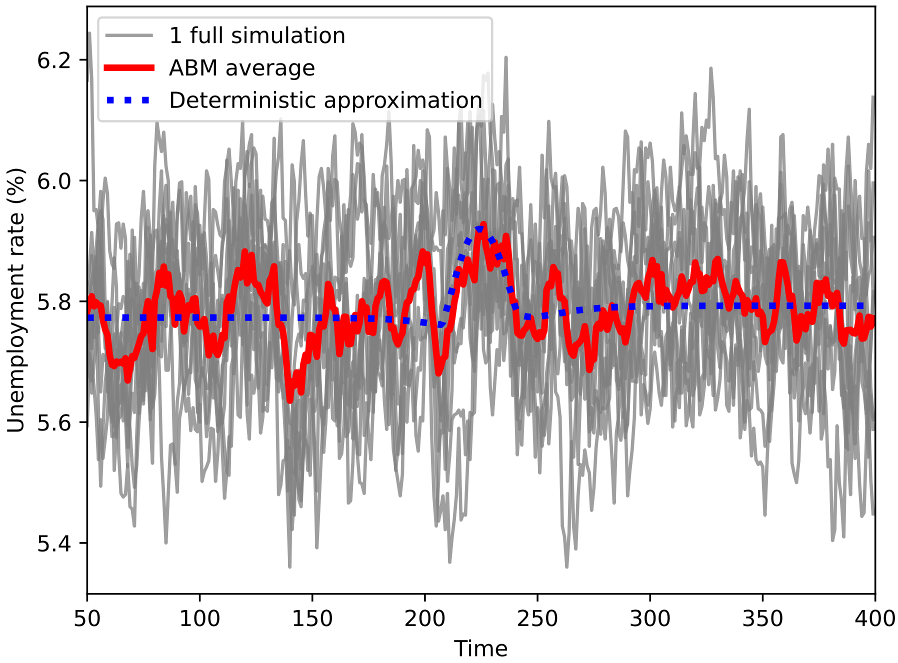


(b) Occupation 1

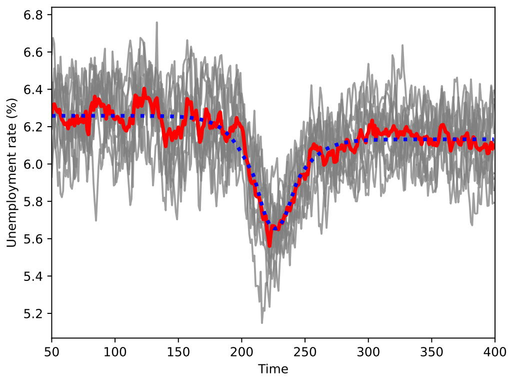


(c) Occupation 2

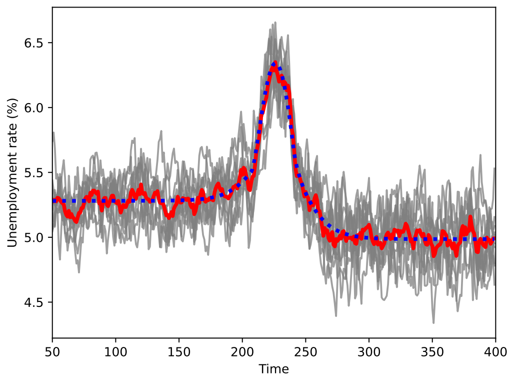
Figure 5.2: **Simulations and approximation for the toy model with on-the-job search, $\mathcal{L} = 500$.** Unemployment rates of the deterministic approximation and for 10 runs of the full model.



(a) Aggregate

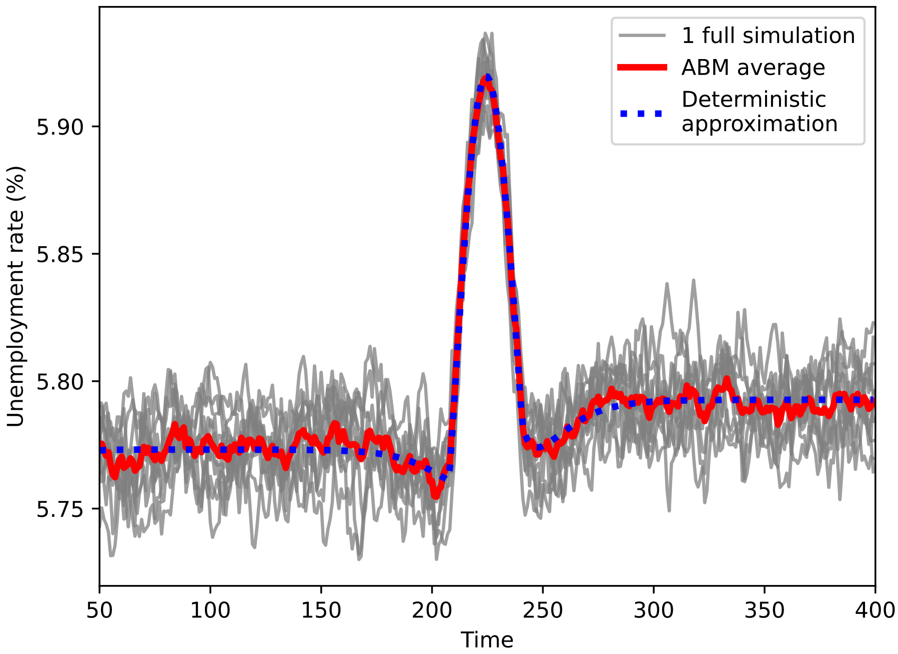


(b) Occupation 1

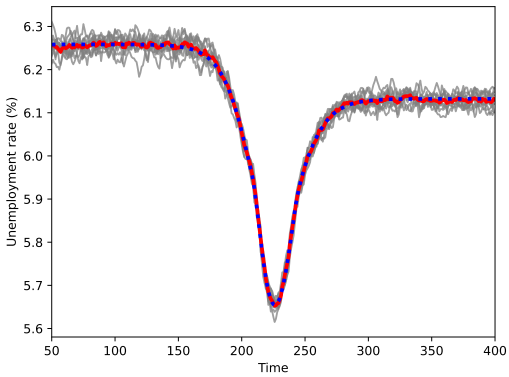


(c) Occupation 2

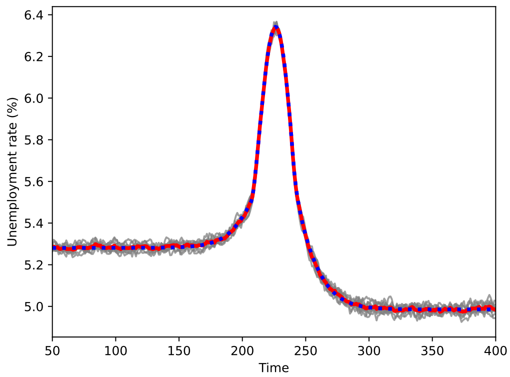
Figure 5.3: **Simulations and approximation for the toy model with on-the-job search, $\mathcal{L}$ = 50,000.** Unemployment rates of the deterministic approximation and for 10 runs of the full model.



(a) Aggregate



(b) Occupation 1



(c) Occupation 2

Figure 5.4: **Simulations and approximation for the toy model with on-the-job search, $\mathcal{L}$ = 5,000,000.** Unemployment rates of the deterministic approximation and for 10 runs of the full model.

$\mathcal{L}$	Average standard error	RMS difference of ABM average from DA	RMS difference of ABM from DA
500	0.397	0.409	1.38
50,000	0.0380	0.100	0.159
5,000,000	0.00394	0.0897	0.0903

Table 5.1: **Comparison of average unemployment rate for on-the-job search.** Decreasing standard error and root mean squared (RMS) difference between the agent-based model (ABM) mean and the deterministic approximation (DA), for different values of $\mathcal{L}$.

$\mathcal{L}$	ABM average time	Approximation time
500	0.24 s	0.33 s
50,000	5.04 s	0.30 s
5,000,000	843.49 s	0.30 s

Table 5.2: **On-the-job search toy model time taken.** Time taken for one deterministic approximation and the average for 10 runs of the agent-based model (ABM), for different values of $\mathcal{L}$.

5.3 Multiple job applications

It is evidenced in the literature that unemployed and employed workers search for job vacancies differently [26, 27]. Mukoyama, Petterson, and Şahin [26] use the American Time Use Survey, available from US Bureau of Labour Statistics, and find that, on average, employed workers spend 0.5 minutes per day searching for a job while unemployed workers spend 30.4 minutes. Therefore, with employed workers now able to apply for open vacancies in the model, we would like to be able to reflect the intensity of unemployed and employed workers' search effort. To do this, we introduce *two* new parameters, for the number of applications workers in each group send at each time step.

In addition to the discussion above on the Beveridge Curve, we find evidence of cyclical job search efforts among employed and unemployed workers. For example, Mukoyama et al. [26] find that, on average, workers spend more time searching during a recession period. Balgova et al. [27] find that, contrary to Mukoyama et al., employed workers sent more and unemployed workers sent fewer applications during the pandemic-induced recession in 2020. They find that the uncertainty around the duration of the COVID-19 pandemic contributed to this unexpected job search behaviour and that health concerns did not appear to influence worker's behaviour. However, for now, in our model, we assume that workers who are searching for a new job send a constant number of applications at each time step, independent of the economy and the business cycle.

Previously, we assumed that unemployed workers sent one application to one vacancy per time step, accepting a job offer if successful and simply sending another application to a vacancy in the next time step if unsuccessful. If a vacancy received one or more applications, one applicant was chosen uniformly at random, said worker was hired, and the vacancy closed. If a vacancy didn't receive any applications, the vacancy remained open until the next time step. Now, we suppose that unemployed workers send β_u applications per time step; and employed workers send β_e applications (if they are looking for a new job). Figure 5.5 is a flow chart that illustrates the transitions from the perspective of a worker and a job vacancy with on-the-job search and multiple applications.

If a specific vacancy receives at least one application, as before, a job offer is sent to a worker who applied to said vacancy uniformly at random. However, before the workers only sent one job application so if their application was chosen, they took up the position. Now we can be in a situation where an applicant receives more than one job offer. In this case, they will choose to take up one of the positions, uniformly at random, leaving a vacancy that received at least one job application without an accepted candidate. If a job offer is rejected, the vacancy simply remains open until the next time step. Calculating the deterministic approximation of this offer rejection represents a key contribution of this thesis. Previous labour market models have not solved this probabilistic and combinatorial problem before. The calculation is presented in the next section.

Unemployed workers and employed workers in occupation i searching for a new job have β_u and β_e balls respectively which they place in the bin of a vacancy in their chosen occupation j. As before, a ball is selected at random for each vacancy open in occupation j, represented by a bin. However, if multiple balls from one worker are selected, the worker randomly chooses an offer to accept. This results in the flow of workers, $F_{ij,t+1}$, again being a stochastic variable. We will see that, conditional on the number of offers sent, $F^{(u)}_{ij,t+1}$ follows a multinomial distribution with $U_{i,t}$ trials, $F^{(e)}_{ij,t+1}$ follows a multinomial distribution with $\lambda E_{i,t}$ trials, and success probabilities which we make explicit in the next section.

The governing equations of the full stochastic agent-based model with multiple applications are the same as given above in (5.1)−(5.3).

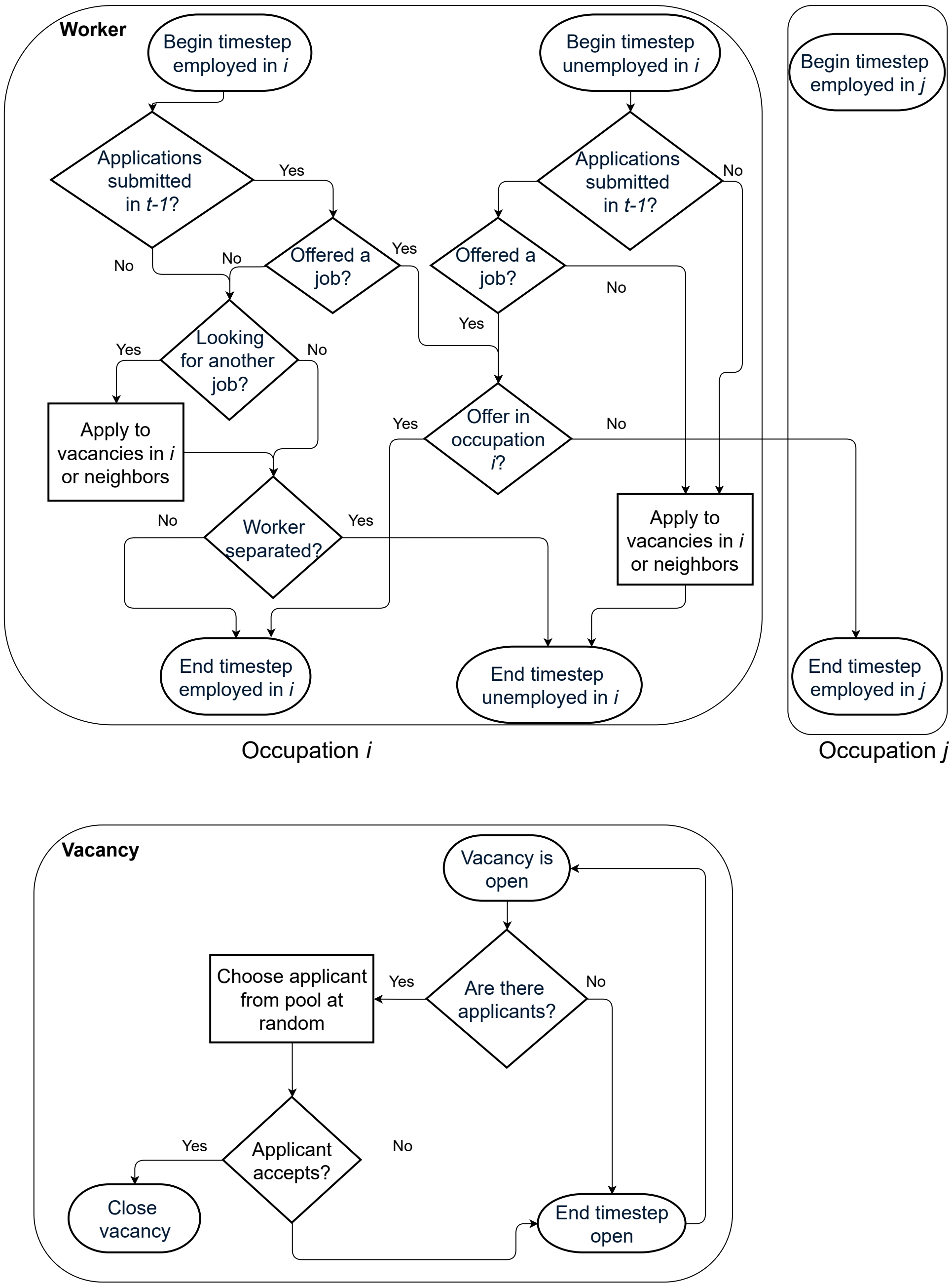


Figure 5.5: **Flow chart of the full, extended agent-based model.** *Top*: transition pathway of a worker. *Bottom*: transition pathway of a job vacancy.

5.3.1 Deterministic approximation for large populations

As before, we compute a deterministic approximation for the full agent-based model, now with workers able to send multiple applications. This adds more complexity to the stochastic process, $F_{ij,t+1}$, as offers will be rejected, if a worker receives more than one offer.

We assume that each worker sends their multiple applications to the same occupation. Therefore, the number of *applicants* from occupation i to occupation j at time t is the same as the number of *applications* in Section 5.2.1, when workers only sent one application. That is, for fixed i, $A^{(u)}_{ij,t+1}$ follows a multinomial with $U_{i,t}$ trials and $A^{(e)}_{ij,t+1}$ follows a multinomial with $\lambda E_{i,t}$ trials both with success probabilities $q_{ij,t+1}$ for $j \in I$, and their expectations are given by

$$\bar{a}^{(u)}_{ij,t+1} = q_{ij,t+1}\bar{u}_{i,t}, \tag{5.25}$$

and

$$\bar{a}^{(e)}_{ij,t+1} = \lambda q_{ij,t+1}\bar{e}_{i,t}. \tag{5.26}$$

An unemployed worker sends β_u applications and employed workers send β_e applications. Each worker sends their applications uniformly at random to the $V_{j,t}$ vacancies, therefore, more than one application can be sent to the same vacancy (for example if fewer than β_u or β_e vacancies are open in the target occupation).[3] Therefore, the expected number of applications sent to occupation j is

$$\bar{s}_{j,t+1} = \beta_u \sum_i \bar{a}^{(u)}_{ij,t+1} + \beta_e \sum_i \bar{a}^{(e)}_{ij,t+1}. \tag{5.27}$$

The number of vacancies that receive at least one application in occupation j, which depends on the number of vacancies open in occupation j and the total number of applications sent to occupation j again follows the same occupancy distribution, that is,

$$M_{j,t+1}|(\mathbf{U}_t, \mathbf{V}_t, \mathbf{E}_t) \sim \mathrm{Occ}\left(k|x = S_{j,t+1}, y = V_{j,t}\right). \tag{5.28}$$

[3]While workers typically do not apply for the same vacancy multiple times, we can interpret this as the effort a worker invests into an application to increase their chances of being accepted.

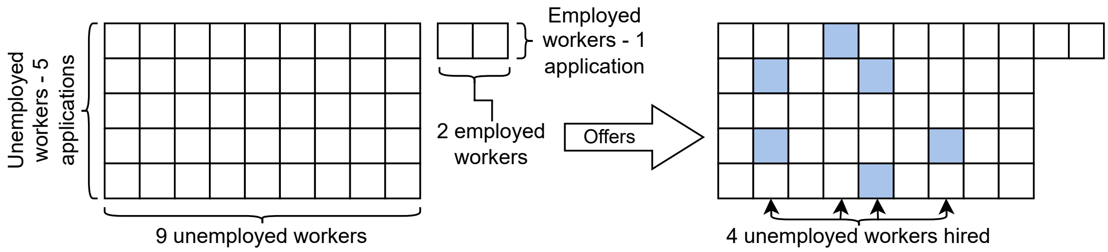


Figure 5.6: **Offer sending process schematic.** Each square represents an application sent to occupation j at time $t+1$, and each column corresponds to a worker. Blue squares in the right hand grid are successful applications.

Flow of workers

Unlike in the previous approximation where we use the Taylor expansion, for this version of the model, using the law of large numbers, we take $A^{(u)}_{ij,t+1}$ and $A^{(e)}_{ij,t+1}$ to be constant from now on. We omit the full Taylor expansion, instead showing in Section 5.5.3 that our approximation holds when we drop this assumption on $A^{(u)}$ and $A^{(e)}$. So, from here, we assume that $\bar{a}^{(u)}_{ij,t+1}$ and $\bar{a}^{(e)}_{ij,t+1}$, given by (5.25) and (5.26), applications are sent by unemployed and employed workers, respectively, and that $\bar{s}_{j,t+1}$ applications are sent to occupation j at time $t+1$, given by (5.27).

Given that we send $M_{j,t+1} \sim \mathrm{Occ}(x = \bar{s}_{j,t+1}, y = \bar{v}_{j,t})$ offers, we need to know how many unique workers receive at least one offer. Consider a grid as shown in Figure 5.6. Each square represents an application in $\bar{s}_{j,t+1}$. Sending $M_{j,t+1} = m$ offers corresponds to choosing m squares without replacement. The number of workers who receive at least one offer corresponds to the number of non-empty columns. If $\bar{a}^{(u)}_{j,t+1} = 9, \bar{a}^{(e)}_{j,t+1} = 2, \beta_u = 5, \beta_e = 1$ and we send $m = 6$ offers, then one possibility, as shown in Figure 5.6, is that $F^{(u)}_{ij,t+1} = 4$ and $F^{(u)}_{ij,t+1} = 0$. Two offers are rejected as workers represented by columns 2 and 5 randomly accept just one of the two offers they each receive.

To calculate the expected number of workers that receive at least one offer, consider that, for each unemployed worker, the probability of receiving at least one offer, is one minus the probability of not receiving any offers. The probability that an unemployed worker (applying to occupation j) does not receive the first offer is

$$1 - \frac{\beta_u}{\bar{s}_{j,t+1}}, \tag{5.29}$$

and the probability that this unemployed worker also does not receive the second offer is

$$1 - \frac{\beta_u}{\bar{s}_{j,t+1} - 1}. \tag{5.30}$$

Therefore, the probability that an unemployed worker applying to occupation j does not receive any of the m offers sent is

$$\prod_{k=0}^{m-1} \left(1 - \frac{\beta_u}{\bar{s}_{j,t+1} - k}\right), \tag{5.31}$$

and so, the probability that an unemployed worker applying to occupation j receives at least one offer is

$$1 - \prod_{k=0}^{m-1} \left(1 - \frac{\beta_u}{\bar{s}_{j,t+1} - k}\right). \tag{5.32}$$

Similarly, the probability that an employed worker applying to occupation j receives at least one offer is

$$1 - \prod_{k=0}^{m-1} \left(1 - \frac{\beta_e}{\bar{s}_{j,t+1} - k}\right). \tag{5.33}$$

So, with constant $A^{(u)}_{ij,t+1}$ and $A^{(e)}_{ij,t+1}$, the expected number of unemployed workers hired from occupation i into occupation j, if $M_{j,t+1} = m$ offers are sent, is

$$\mathbb{E}(F^{(u)}_{ij,t+1} | M_{j,t+1} = m) = \bar{a}^{(u)}_{ij,t+1} \left(1 - \prod_{k=0}^{m-1} \left(1 - \frac{\beta_u}{\bar{s}_{j,t+1} - k}\right)\right), \tag{5.34}$$

and for employed workers,

$$\mathbb{E}(F^{(e)}_{ij,t+1} | M_{j,t+1} = m) = \bar{a}^{(e)}_{ij,t+1} \left(1 - \prod_{k=0}^{m-1} \left(1 - \frac{\beta_e}{\bar{s}_{j,t+1} - k}\right)\right), \tag{5.35}$$

where $\bar{s}_{j,t+1} = \beta_u \sum_i \bar{a}^{(u)}_{ij,t+1} + \beta_e \sum_i \bar{a}^{(e)}_{ij,t+1}$.

Finally, since $M_{j,t+1}$ follows the occupancy distribution, the total number of workers hired from occupation i into occupation j from unemployment and from employment are

$$\begin{aligned}
\bar{f}^{(u)}_{ij,t+1} &= \sum_{m=1}^{\bar{v}_{j,t+1}} \mathbb{E}[F^{(u)}_{ij,t+1}|M_{j,t+1}=m]P(M_{j,t+1}=m) \\
&= \sum_{m=1}^{\bar{v}_{j,t+1}} \left[\bar{a}^{(u)}_{ij,t+1} \left(1 - \prod_{k=0}^{m-1} \left(1 - \frac{\beta_u}{\bar{s}_{j,t+1}-k}\right)\right) \phi(m|\bar{s}_{j,t+1}, \bar{v}_{j,t+1}) \right],
\end{aligned} \tag{5.36}$$

$$\begin{aligned}
\bar{f}^{(e)}_{ij,t+1} &= \sum_{m=1}^{\bar{v}_{j,t+1}} \mathbb{E}[F^{(e)}_{ij,t+1}|M_{j,t+1}=m]P(M_{j,t+1}=m) \\
&= \sum_{m=1}^{\bar{v}_{j,t+1}} \left[\bar{a}^{(e)}_{ij,t+1} \left(1 - \prod_{k=0}^{m-1} \left(1 - \frac{\beta_e}{\bar{s}_{j,t+1}-k}\right)\right) \phi(m|\bar{s}_{j,t+1}, \bar{v}_{j,t+1}) \right],
\end{aligned} \tag{5.37}$$

with $\phi(k|x,y)$ given by (5.9).

The three governing equations are the same as in the previous section (5.22)−(5.24) with the new definitions for the expected number of workers flowing from unemployment and employment in occupation i into employment in occupation j.

While the probability mass function for the occupancy distribution is computationally simple, the coefficients can quickly get too large to compute. Therefore, O'Neill [135] introduces an approximation to the occupancy distribution to eliminate the need for computing large Sterling's numbers. However, even after these are computed, the calculation still risks reaching computational barriers when calculating the ratios of large integers and the small probabilities in the tails. O'Neill presents two approximations to overcome these barriers. One is a recursive algorithm, when the number of bins and balls are known, that uses a reduced occupancy distribution with one parameter (Occ(x,x)) and a correction term. The log-probabilities of the reduced occupancy distribution are stored, up to a known maximum $x = N$ balls and bins, and fetched during the recursion when needed. The second leverages the normal distribution, to approximate the occupancy distribution and is most useful when N is too large to practically store the necessary log-probabilities, or the largest N needed is unknown.

The *scaled normal density approximation* to the occupancy distribution is given by

$$\mathrm{Occ}(k|x,y) \sim \frac{\mathrm{N}(k|\mu_{x,y}, \sigma^2_{x,y})}{\sum_k \mathrm{N}(k|\mu_{x,y}, \sigma^2_{x,y})}, \tag{5.38}$$

where $\mu_{x,y} = \mathbb{E}(K)$ and $\sigma^2_{x,y} = \mathbb{V}(K)$ for $K \sim \mathrm{Occ}(k|x,y)$. We use this normal approximation, following Algorithm 3 from O'Neill [135], throughout our

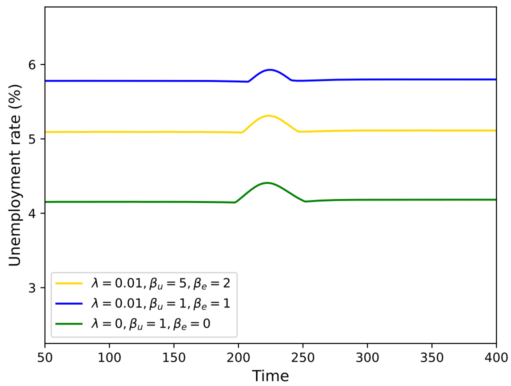


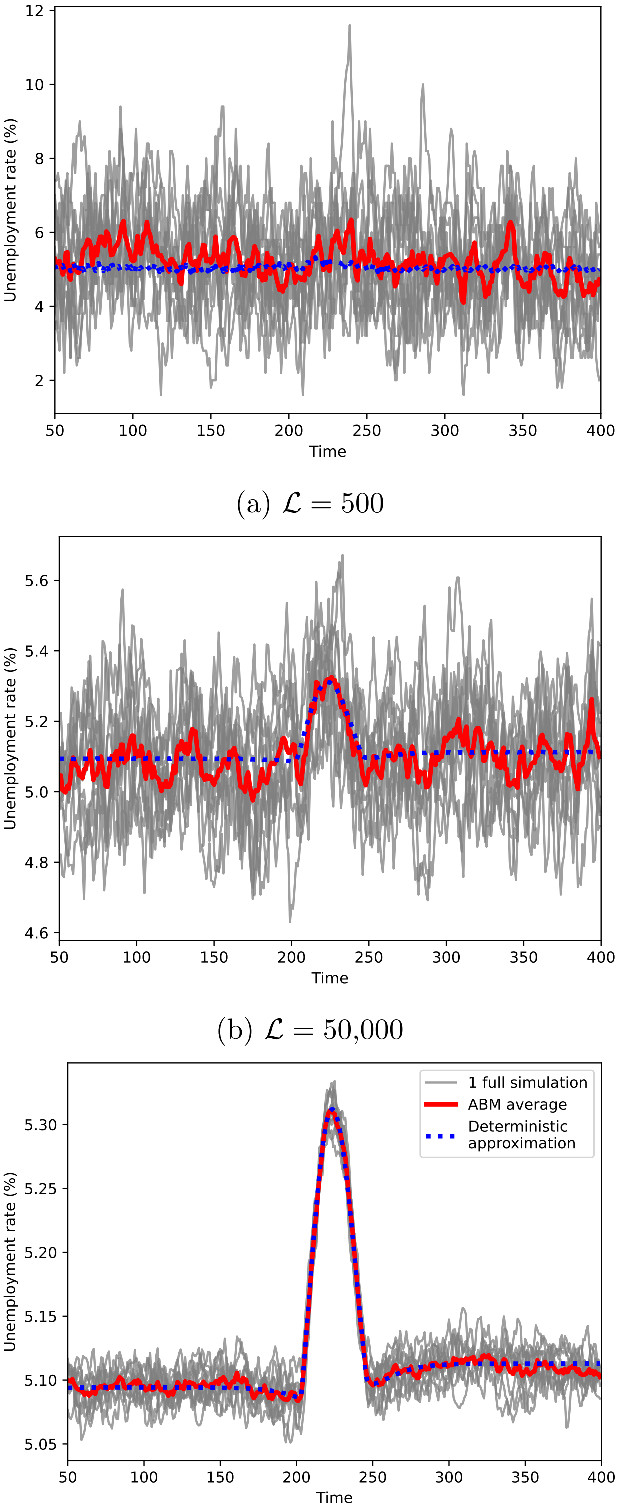
Figure 5.7: **Multiple applications toy model unemployment rate.** The aggregate unemployment rate with, and without, multiple applications and on-the-job search.

remaining analysis, and in Section 5.5.2, we investigate the bias introduced by this approximation for small values of x and y, and briefly discuss the recursive algorithm.

5.3.2 Toy model

We can set up the same toy model using the Universal 1 scenario to study multiple applications and show the approximation and full agent-based model simulations. We set the new parameters to $\beta_u = 5$ and $\beta_e = 2$ and show the effect this has on the steady state in Figure 5.7. We can see that giving unemployed workers five applications, and employed workers two applications, per time step, reduces the increase in the aggregate unemployment rate that was caused by adding on-the-job search. This suggests that the steady state is further dependent on the two new parameters, β_u and β_e. We investigate their impact further in Chapter 6.

As we saw with the on-the-job search toy model, as the total number of workers, $\mathcal{L}$, increases, each agent-based model run approaches the deterministic approximation, illustrated in Figure 5.8, and the standard error decreases, as shown in Table 5.3. We use the scaled normal approximation to the occupancy distribution probability mass function for all three deterministic approximations. In the case of $\mathcal{L} = 500$, we see an oscillation in the approximated unemployment rate, this is because we hit the special case when $\min(x, y) = 1$, and the normal approximation is most biased. We do not dwell on this though, as we apply our models to much larger populations, and, in any case, our approximation relies on the law of large numbers.



(a) $\mathcal{L} = 500$

(b) $\mathcal{L} = 50{,}000$

(c) $\mathcal{L} = 5{,}000{,}000$

Figure 5.8: **Simulations and approximations for the toy model with multiple applications.** Aggregate unemployment rate of the deterministic approximation and 10 runs of the full agent-based model for different values of $\mathcal{L}$.

$\mathcal{L}$	Average standard error	RMS difference of ABM average from DA	RMS difference of ABM from DA
500	0.365	0.446	1.44
50,000	0.0419	0.0719	0.140
5,000,000	0.00390	0.0553	0.0571

Table 5.3: **Comparison of average unemployment rate for multiple applications.** Decreasing standard error and root mean squared (RMS) difference between the agent-based model (ABM) mean and the deterministic approximation (DA), for different values of $\mathcal{L}$.

5.4 Wage driven dynamics

As well as on-the-job search and multiple applications, we now add wage pressure into the model. Haltiwanger et al. [28] show that employed workers move to higher wage jobs more than unemployed workers and Deutscher [29] shows that employed workers moving job-to-job can increase wage growth, motivating the need to add wage pressure into the model, especially now that we have on-the-job search.

One way to add wage dynamics would be to assign a wage to each open vacancy, and subsequently increasing this advertised wage if the vacancy is left open. This would also require advertised wages to decrease when vacancies are filled quickly. Adding dynamic wages would require more complex financial modelling to be incorporated into our model. Real time wages fluctuate with interest rates, productivity, inflation, technology, and many other mechanisms which are outside the current scope of our labour market model [141]. When the model is closely coupled to an external macro-economic model, as well as labour demand, the macro model could provide dynamic wages, or more economic information so that we can add these real time wage dynamics.

For now, we include wage driven dynamics in the probability that a worker applies to a specific vacancy, which now includes a function of the target occupation's wage, capturing the propensity for workers to transition to higher wage occupations. The probability that a worker in occupation i applies to occupation j, previously (3.6), is now, for some function ψ_{ij},

$$q_{ij,t+1} = \frac{\bar{v}_{j,t}\psi_{ij}(\mathbf{w})\mathcal{A}_{ij}}{\sum_l \bar{v}_{l,t}\psi_{il}(\mathbf{w})\mathcal{A}_{il}}, \tag{5.39}$$

where w_i is the median wage of occupation i. Determining $\psi_{ij}(\mathbf{w})$ is a point for future work. Some possible forms include, for a worker in occupation i applying to occupation j,

$$\psi_{ij}(\mathbf{w}) = w_j, \quad \psi_{ij}(\mathbf{w}) = (w_j - w_i)^2, \quad \psi_{ij}(\mathbf{w}) = \ln(w_j). \tag{5.40}$$

Adding wages can be interpreted as using a different network with the edges simply weighted by the $\psi(\mathbf{w})$ function, therefore, the deterministic approximation holds for this extension.

In summary, the stochastic agent-based model can be approximated using the deterministic system of equations below. In Table 5.4, we present a summary of all the variables and parameters. The three governing equations for employment, unemployment, and vacancies, previously (3.33)$-$(3.35), are now

$$\bar{e}_{i,t+1} = \bar{e}_{i,t} - (\delta_u \bar{e}_{i,t} + (1-\delta_u)\gamma \max\{0, \bar{d}_{i,t} - \mathcal{D}^{\dagger}_{i,t}\}) + \sum_j \bar{f}_{ji,t+1} - \sum_j \bar{f}^{(e)}_{ij,t+1}, \tag{5.41}$$

$$\bar{u}_{i,t+1} = \bar{u}_{i,t} + (\delta_u \bar{e}_{i,t} + (1-\delta_u)\gamma \max\{0, \bar{d}_{i,t} - \mathcal{D}^{\dagger}_{i,t}\}) - \sum_j \bar{f}^{(u)}_{ij,t+1}, \tag{5.42}$$

$$\bar{v}_{i,t+1} = \bar{v}_{i,t} + (\delta_v \bar{e}_{i,t} + (1-\delta_v)\gamma \max\{0, \mathcal{D}^{\dagger}_{i,t} - \bar{d}_{i,t}\}) - \sum_j \bar{f}_{ji,t+1}. \tag{5.43}$$

where $\bar{f}_{ij,t+1} = \bar{f}^{(u)}_{ij,t+1} + \bar{f}^{(e)}_{ij,t+1}$, and $\bar{f}^{(u)}_{ij,t+1}$ and $\bar{f}^{(e)}_{ij,t+1}$ are the expected flow of workers from unemployment and employment, respectively, in occupation i at time t to employment in occupation j at time $t+1$ given by

$$\bar{f}^{(u)}_{ij,t+1} = \underbrace{\sum_{m=1}^{\bar{v}_{j,t+1}} \left[\bar{a}^{(u)}_{ij,t+1} \left(1 - \prod_{k=0}^{m-1} \left(1 - \frac{\beta_u}{\bar{s}_{j,t+1} - k}\right)\right) \phi(m|\bar{s}_{j,t+1}, \bar{v}_{j,t+1}) \right]}_{\bar{f}^{(u)}_{ij,t+1}}, \tag{5.44}$$

$$\bar{f}^{(e)}_{ij,t+1} = \underbrace{\sum_{m=1}^{\bar{v}_{j,t+1}} \left[\bar{a}^{(e)}_{ij,t+1} \left(1 - \prod_{k=0}^{m-1} \left(1 - \frac{\beta_e}{\bar{s}_{j,t+1} - k}\right)\right) \phi(m|\bar{s}_{j,t+1}, \bar{v}_{j,t+1}) \right]}_{\bar{f}^{(e)}_{ij,t+1}}, \tag{5.45}$$

with $\phi(k|x,y)$ given by (5.9) and where,

$$\bar{a}^{(u)}_{ij,t+1} = q_{ij,t+1}\bar{u}_{i,t}, \tag{5.46}$$

$$\bar{a}^{(e)}_{ij,t+1} = \lambda q_{ij,t+1}\bar{e}_{i,t}, \tag{5.47}$$

$$\bar{s}_{j,t+1} = \beta_u \sum_i \bar{a}^{(u)}_{ij,t+1} + \beta_e \sum_i \bar{a}^{(e)}_{ij,t+1}, \tag{5.48}$$

$$q_{ij,t+1} = \frac{\bar{v}_{j,t}\psi(\mathbf{w})\mathcal{A}_{ij}}{\sum_l \bar{v}_{l,t}\psi(\mathbf{w})\mathcal{A}_{il}}, \tag{5.49}$$

for some function ψ.

Main Variables	Description
$E_{i,t}$	Number of workers employed in occupation i at time t.
$U_{i,t}$	Number of workers unemployed at time t which were last employed in occupation i.
$V_{i,t}$	Number of job vacancies in occupation i at time t.
Other Variables	
$B_{i,t}$	Number of workers in occupation i separated from their jobs at time t.
$C_{i,t}$	Number of vacancies opened in occupation i at time t.
$D_{i,t}$	Model's realised demand in occupation i at time t
$S_{ij,t}$	Number of applications sent by workers in occupation i to j
$A_{ij,t}$	Number of applicants from occupation i to j at time t
$P_{j,t}$	Probability of any application to occupation j being successful
$M_{j,t}$	Number of non-empty vacancies in occupation j at time t
$F_{ij,t}$	Flow of workers in occupation i to employment in occupation j at time t
$q_{ij,t}$	Conditional probability that a worker in occupation i applies to occ j
$p_{u,i,t}$	Occupation-specific conditional probability that a worker of occupation i is separated.
$p_{v,i,t}$	Occupation-specific conditional probability that a vacancy of occupation i opens.
Parameters	
δ_u	Rate at which employed workers are separated due to the spontaneous process
δ_v	Rate at which vacancies are opened due to the 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
$\mathcal{A}$	Adjacency matrix of the input network
$\mathcal{D}^{\dagger}_{i,t}$	Target labour demand in occupation i at time t
Δt	Duration of time step in units of weeks
λ	Probability that an employed worker is looking for a job
β_u	Number of applications sent by each unemployed worker per time step
β_e	Number of applications sent by each employed worker per time step (if they are looking for a job)

Table 5.4: **Variables and parameters.**

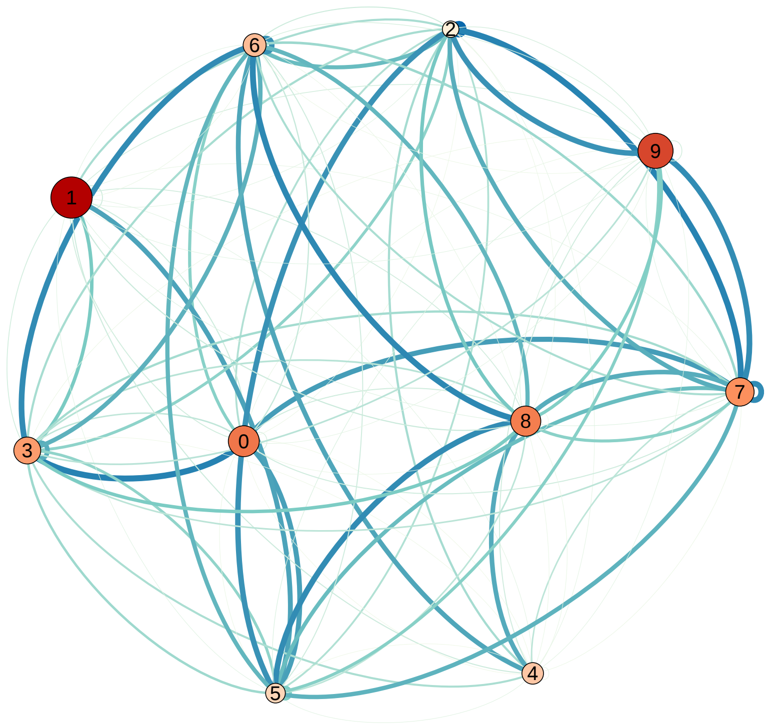
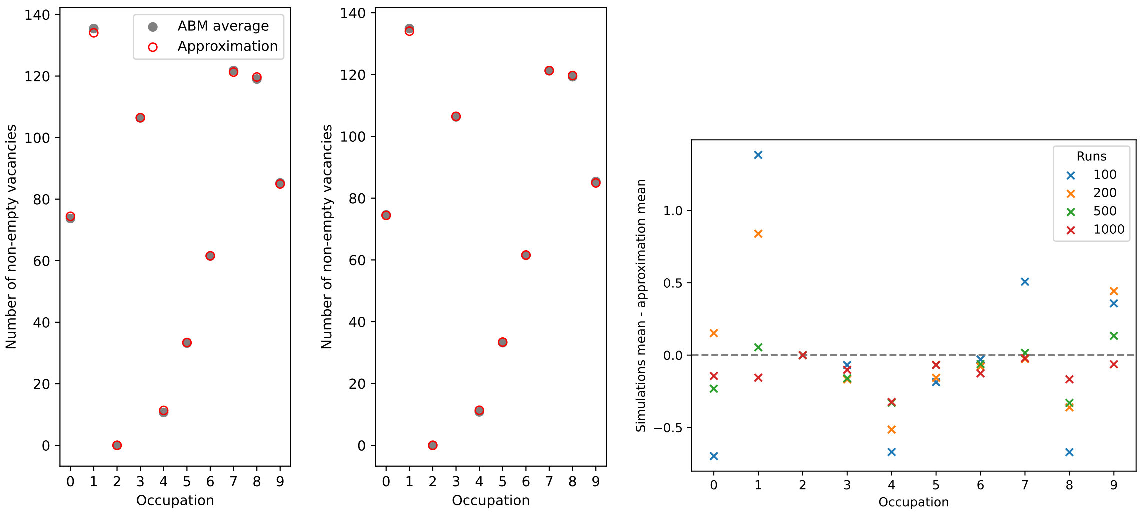


Figure 5.9: **Random occupational mobility network for 10 occupations.** The size of the nodes is proportional to the size of initial employment.

5.5 Approximation tests

Having the deterministic approximation of the full, stochastic agent-based model enables us to quickly analyse the model, investigate parameters, and gain insights into the effects of many different shocks. In this section, we investigate the relationship between the full agent-based model and the deterministic approximation. The toy models above show that the error is linear in $\mathcal{L}$ and that the approximate aggregate unemployment rate is accurate. To get a better understanding of the error introduced at individual stages in the model, we compare the full model with the approximation for some intermediary processes, such as the number of non-empty vacancies, $M_{j,t+1}$. We show that the approximations are reasonable for typical values; showing qualitatively that the approximation holds, while not formally proving so.

The following tests are for $N = 10$ occupations and $\mathcal{L} = 100{,}000$ workers, with a random occupational mobility network shown in Figure 5.9. Note that this is the complete network with non-equal weights. Initial employment is also randomised, unemployment and vacancies are set to $\mathbf{U}_0 \approx \delta_u \mathbf{E}_0 = 0.016\mathbf{E}_0$ and $\mathbf{V}_0 \approx \delta_v \mathbf{E}_0 = 0.012\mathbf{E}_0$, and wages are all set to 1.



(a) 100 simulations (b) 200 simulations (c) Error

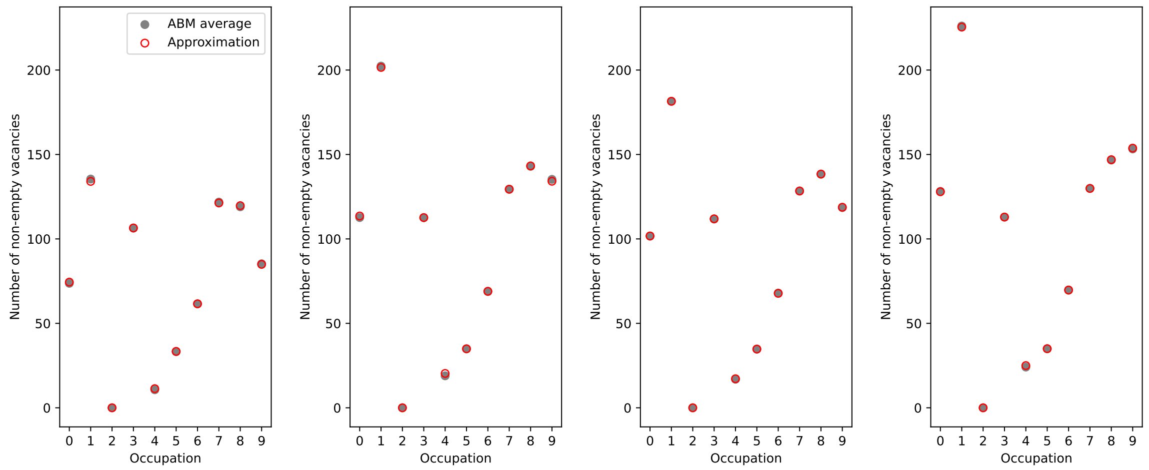
Figure 5.10: **Simulation average for non-empty vacancies.** The number of non-empty vacancies, $M_{j,t+1}$ for (a) 100 runs, and (b) 200 runs of the full agent-based model with $\mathcal{L} = 100{,}000$ and $N = 10$, and in (c), the approximation error.

5.5.1 Non-empty vacancies

First, we investigate the consistency of the number of non-empty vacancies in each occupation j with the occupancy distribution $M_{j,t+1} \sim \text{Occ}(n = \bar{s}_{j,t+1}, m = \bar{v}_{j,t})$.

In Figure 5.10, we see the mean of the full agent-based model and the expected value for the number of non-empty vacancies, $\bar{m}_{j,t+1}$, for 100 and 200 runs of the agent-based model with $\lambda = 0$ and $\beta_u = 1$. The expectation of the occupancy distribution, shown with the red circles, closely follows the mean of the full agent-based model for all 10 occupations. We plot the difference between the two means in Figure 5.10c for 100, 200, 500, and 1,000 runs, and see that our approximation error decreases when we increase the number of simulations.

In Figure 5.11, we vary our new model parameters, λ, β_u, and β_e to see how they interact with the mobility network. For Occupations 2−7, the number of non-empty vacancies does not increase much with the number of applications. This is due to the network structure; despite having fewer vacancies open, these occupations have a relatively strong in-degree and so even in the case with the fewest application, when $\lambda = 0$ and $\beta_u = 1$ shown in Figure 5.11a, nearly all open vacancies in these occupations are filled. Alternatively, in Occupations 0, 1, 8, and 9, the number of non-empty vacancies increases as more applications are sent.



(a) $\lambda = 0$, $\beta_u = 1$, $\beta_e = 0$. (b) $\lambda = 0$, $\beta_u = 2$, $\beta_e = 0$. (c) $\lambda = 0.01$, $\beta_u = 1$, $\beta_e = 1$. (d) $\lambda = 0.01$, $\beta_u = 2$, $\beta_e = 1$.

Figure 5.11: **Effect of λ, β_u, and β_e on non-empty vacancies.** The number of non-empty vacancies, $M_{j,t+1}$.

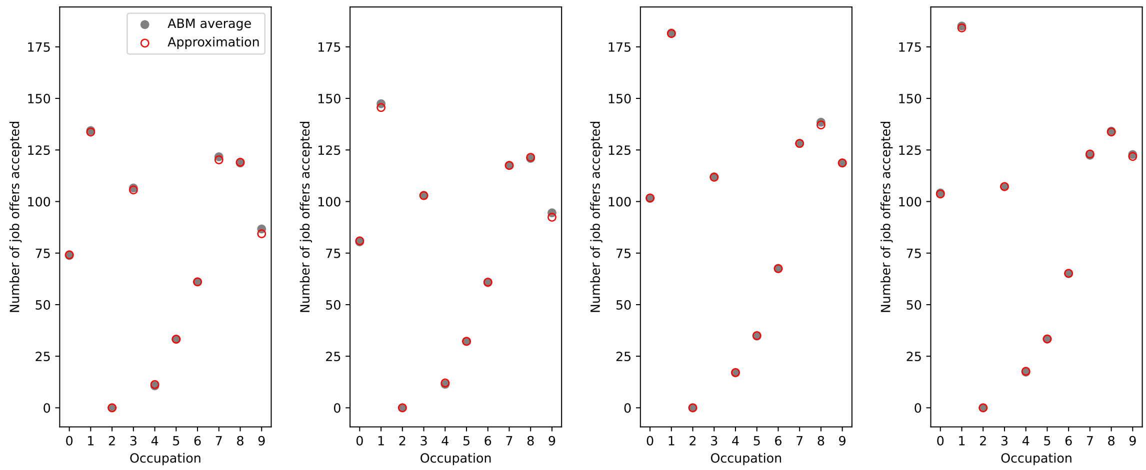
5.5.2 Accepted job offers

Next, we check that the expected number of job offers that are accepted is indeed given by the hired workers term in (5.41). The total number of hired workers into occupation j is

$$\bar{f}_{j,t+1} = \sum_i \bar{f}_{ij,t+1} = \sum_i \bar{f}^{(u)}_{ij,t+1} + \sum_i \bar{f}^{(e)}_{ij,t+1}. \tag{5.50}$$

In Figure 5.12, we see that the mean over 100 simulations for the number of job offers accepted in each occupation (and therefore the flow of workers hired into each occupation) is largely in line with the expected number of job offers accepted given by $\bar{f}_{j,t+1}$. There are no rejected offers when each worker only sends one application, so the number of job offers accepted when $\beta_u = 1$ and $\beta_e = 1$, shown in Figures 5.12a and 5.12c are the same as the number of non-empty vacancies, shown in Figures 5.11a and 5.11c. For most occupations, the number of non-empty vacancies increases as the number of applications increases, as shown in Figure 5.11. However, this increase is not realised in the number of accepted job offers. Specifically, there is little increase in the number of accepted job offers for $\lambda = 0$ and β_u increasing from one to two. This indicates that in this scenario, most unemployed workers receive an offer when only sending one application, and so do not much benefit from sending more.

We can see that the error in the approximation for the number of accepted job offers is greater than the error for non-empty vacancies, particularly for smaller values



(a) $\lambda = 0$, $\beta_u = 1$, $\beta_e = 0$.

(b) $\lambda = 0$, $\beta_u = 2$, $\beta_e = 0$.

(c) $\lambda = 0.01$, $\beta_u = 1$, $\beta_e = 1$.

(d) $\lambda = 0.01$, $\beta_u = 2$, $\beta_e = 1$.

Figure 5.12: **Effect of λ, β_u, and β_e on accepted job offers.** The number of accepted job offers, $F_{j,t+1}$.

of the three new parameters. While the error is still minimal, as we saw in the toy models in both Section 5.2.2 and 5.3.2, we do introduce a further approximation at this step, the scaled normal density approximation to the occupancy distribution. Even though using the full occupancy distribution is still impractical, due to computational restrictions, we can use the recursive algorithm proposed by O'Neill [135] to compute the exact occupancy distribution to gain insight into the error that this step of the approximation introduces.

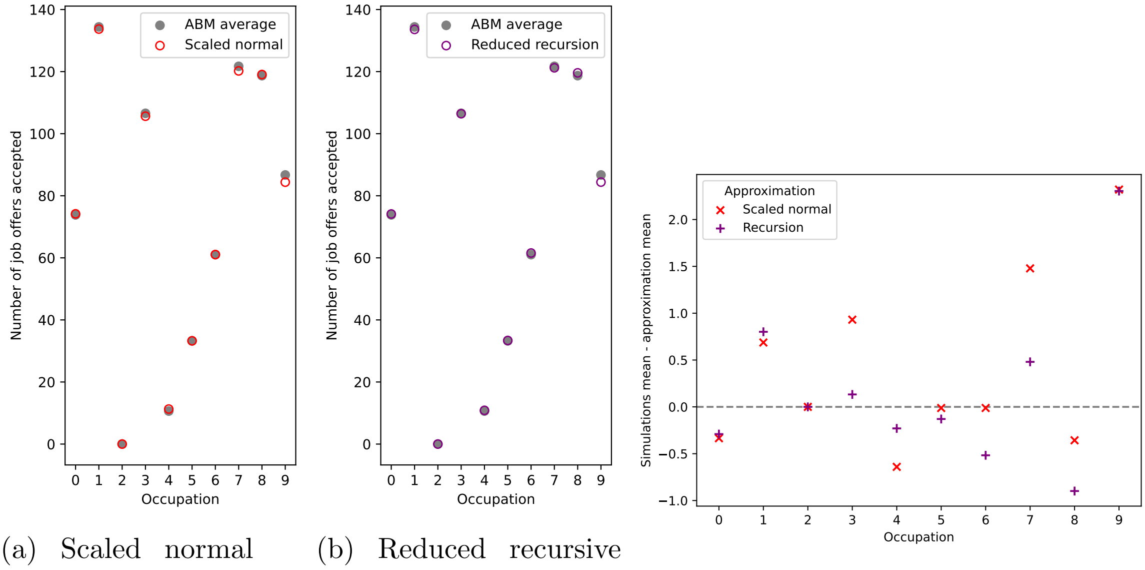
The recursion algorithm given in O'Neill [135] uses the reduced occupancy distribution where the number of bins and balls are identical. With a single parameter, the reduced occupancy distribution is

$$\mathrm{Occ}(k|x) \equiv \mathrm{Occ}(k|x,x) = \frac{(x)_k \cdot S(x,k)}{x^x} \tag{5.51}$$

for all integers $1 \leq k \leq x$, and then this scales to get back to the two-parameter occupancy distribution

$$\mathrm{Occ}(x,y) = \mathrm{Occ}(k|x) \cdot \frac{x^x}{y^x} \cdot \frac{(y)_k}{(x)_k}. \tag{5.52}$$

The two-step algorithm first calculates the log-probabilities up to a given maximum x, and then a recursive algorithm computes the mass function. Details of the algorithms can be found in O'Neill [135].



(a) Scaled normal approximation

(b) Reduced recursive algorithm

(c) Error

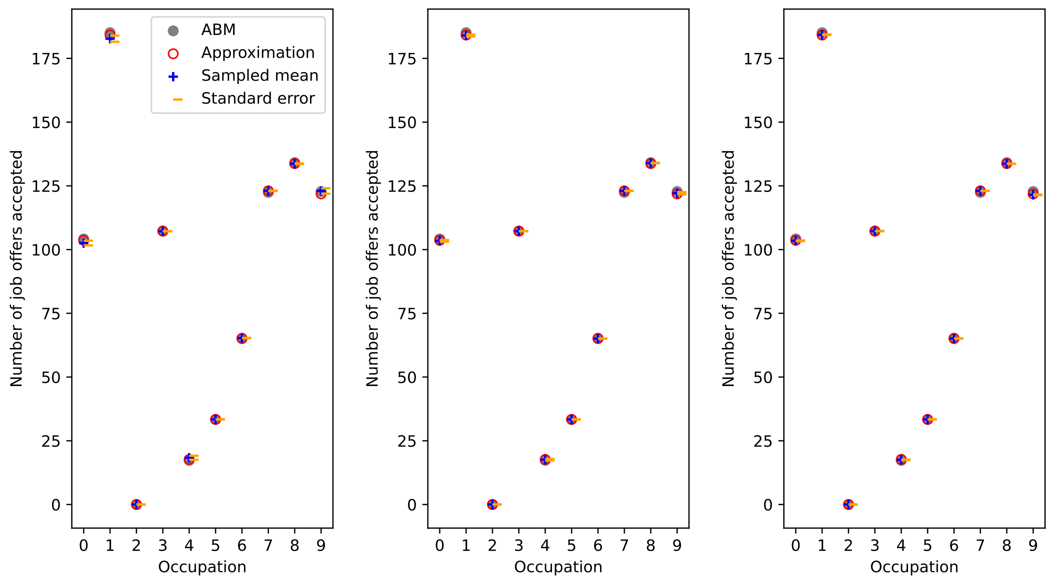
Figure 5.13: **Effect of occupancy distribution method on the number of job offers.** The number of accepted offers for (a) scaled normal density approximation to occupancy distribution (5.38), and (b) recursion algorithm using the reduced occupancy distribution (5.52) for the agent-based model with $\mathcal{L} = 100{,}000$ and $N = 10$, and in (c), the approximation error.

Indeed, when using the reduced occupancy distribution algorithm, we see in Figure 5.13 that things get better for some occupations, specifically Occupations 3, 4, and 7. However, there are some where there is almost no difference and even two where the recursive algorithm has a greater error than the normal approximation (Occupations 6 and 8). This is likely since the normal approximation to the occupancy distribution allows for non-integer values for the parameters, x and y however, we have to round $\bar{s}_j$ to use the recursive algorithm. Because of this, we use the normal approximation for all cases going forward.

5.5.3 Applicants as constant in multiple applications

The final assumption we investigate in this section is where we take as constant the number of applicants sent to occupation j in our calculations in Section 5.3, before we derive the expressions for the expected number of workers moving from occupation i to j, given in (5.36) and (5.37).

In Figure 5.14, we show the offers accepted in each occupation by unemployed and employed workers, with the standard error for 10, 100, and 1,000 samples for each multinomial $A^{(u)}_{ij,t+1}$ and $A^{(e)}_{ij,t+1}$, for $i = 0, \ldots, 9$. We can see that the standard error,



(a) 10 samples (b) 100 samples (c) 1000 samples

Figure 5.14: **Job offers accepted when applications are sampled.** The number of accepted job offers and the standard error of the approximation for 10, 100, and 1,000 samples of the multinomials $A^{(u)}_{ij,t+1}$ and $A^{(e)}_{ij,t+1}$ for $i \in 0, \ldots, 9$.

even for only 10 samples for each multinomial, is relatively small. The approximated mean, taking $A^{(u)}_{ij,t+1}$ and $A^{(e)}_{ij,t+1}$ constant before we calculate $F_{ij,t+1}$, and the sample mean both follow the mean over 10 runs of the agent-based model closely.

5.6 Conclusions

Both the extended full stochastic agent-based model and the deterministic approximation are now ready to be used to study further economic transitions. In this chapter, we introduced on-the-job search, allowed multiple job applications to be sent, and added a wage driver into the application process in the model. Calculating the deterministic approximation for such models with multiple applications is a key step forward in the development of such stochastic labour market models. We leveraged the Occupancy distribution to calculate the expectation of the conditional processes related to adding multiple applications into the model.

With these extensions, our labour market model now captures three more mechanisms, which are important for economic phenomena such as the Beveridge Curve. There are many more features of the labour market that could be added to our model, to answer different policy questions, and we discuss these in Chapter 7.

Chapter 6

Automation case study

This chapter addresses the question:

- How do different job searching intensities affect the unemployment rate?

We illustrate the effect that the three model extensions have on the automation scenario from del Rio-Chanona et al. [11], using the occupational mobility network for the US. We include a brief discussion of the survey data used to construct the US network and then investigate how the unemployment rate outcomes are affected by changing the new parameters governing the volume on-the-job search and the search intensity for workers, for both the occupational mobility network, and the complete network that removes labour mobility frictions.

6.1 US network

The US network is constructed using data from the US Current Population Survey. This is a monthly survey that identifies households by address, chosen to represent the US population. Respondents to the CPS are interviewed using 4-8-4 sampling. That is, they are interviewed for four months, not for eight months, and then interviewed for a further four months. Respondents are asked to explain their occupation, which can result in occupation mislabelling. This results in false moves being seen between occupations that are only due to a change in label of the same job, not a change in occupation [142]. Cheng and Park [92] use data from the Annual Social and Economic Supplement (ASEC) which is a yearly supplement to the CPS. Respondents are asked for their current occupation and also their primary occupation during the past year at the same interview, potentially resulting in fewer false occupation transitions due to mislabelling than with the CPS monthly responses. Another benefit of the ASEC data is that migration is recorded. In the CPS, respondents are chosen by their address, so if someone moves house, they are no longer in the survey. In the ASEC, respondents are asked about their migration status compared to one year ago.

One caveat to using survey data is that recently there has been an increase in non-response which is a problem in both the CPS and supplement ASEC data [143]. We do use weights for the respondents that take into account this non-response so this is at least partly accounted for. However, Ward and Edwards [144] also show that response issues occurred during the COVID-19 pandemic. In March 2020, interviews were all conducted via telephone instead of being in-person; Ward and Edwards propose that the increase in non-response is due, in part, to this change. This non-response is biased towards "an older sample, with fewer racial ethnic minority respondents, and a greater share of respondents with higher levels of educational attainment". They show that person weights were adjusted to "downweight" new cohorts.

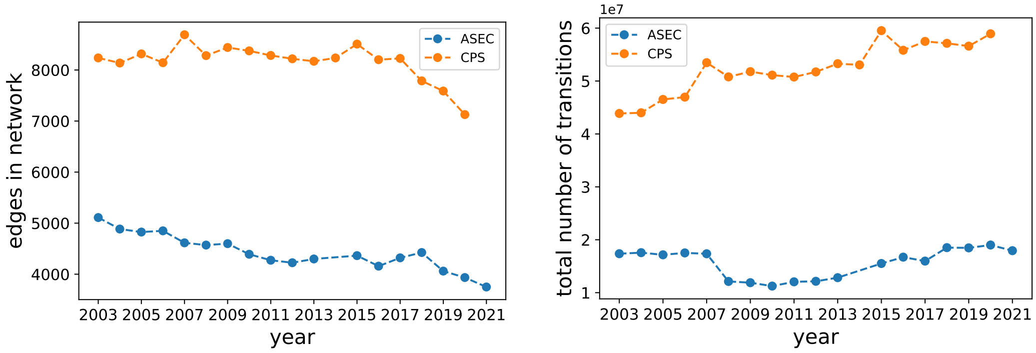
Although we do have to be aware of mislabelling, non-response bias, and the lack of migration in the CPS, these data are still informative when looking at occupational mobility. Next we will look in more detail at the two sources of data, before constructing the US occupational mobility network and investigating the new model parameters.

6.1.1 CPS v. ASEC

There are about 60,000 households in the CPS per month, and 98,000 households in the ASEC each year. With the ASEC data, we can only calculate transitions for workers that have a different OCC (current occupation) and OCCLY (primary occupation from the last 12 months). In the CPS we sample each household a number of times per year, so each worker is asked their current occupation multiple times, which can result in more than one transition. In terms of raw numbers (not using weights) for 2019 for example, in ASEC, 19,646 respondents changed occupation with 180,101 unique people interviewed. In CPS, for the whole year of 2019, 47,009 respondents changed occupation in consecutive months that they were in the sample with 197,638 people interviewed at least twice in 2019.

We generate occupational mobility networks, using 2010 harmonised occupation codes, for both CPS and ASEC data from 2003 to 2020.[1] In Figure 6.1, we can see the number of edges and total number of transitions for the weighted networks. Both datasets show a slight decrease in the number of edges (Figure 6.1a) and an

[1] 2014 is missing for ASEC due to a one-off survey redesign.



(a) Number of edges

(b) Number of transitions

Figure 6.1: **Comparison of occupational mobility networks using CPS and ASEC data.** (a) The number of edges in the networks, and (b) The number of transitions recorded in each year, i.e. the sum of all the weights in the network for networks from 2003 to 2020 for CPS and 2003 to 2021, excluding 2014, for ASEC.

increase in the total number of transitions (Figure 6.1b) since 2008. The difference in magnitude is due to the difference in sampling of occupation moves. In March 2019, the total weights of respondents in the CPS comes to around 166 billion while the total weights of respondents in ASEC comes to about 164 billion; the difference in the networks is due to the sampling of the number of occupation moves.[2]

Unfortunately, if a respondent in the CPS moves address, they are dropped from the survey. However, we can use ASEC data to look at migration. Figure 6.2 shows the variable MIGRATE1 which captures the migration status of each respondent, compared to one year ago. On average, about 90% of respondents live in the same home as they did one year ago. Of the roughly 10% who do move house, on average, 15% of these people also change occupation. Therefore, by using the CPS data, we lose about 1.5% of occupational transitions, made by people who change occupation but also migrate (and therefore leave the survey).

Given that both the CPS and ASEC data sets have similar limitations, not having migration in the CPS only loses 1.5% of transitions, and the CPS sample size is always larger than ASEC, we will use CPS data for the rest of this chapter. The survey limitations also mean that we cannot construct a regional occupational mobility network for the US, therefore this chapter is at the occupation-level only. To construct the US occupational mobility network, we count job transitions that CPS survey responders made between January 2010 and May 2017. We use the 4-digit level, so $N = 464$. Let T_{ij} be the number of workers that transitioned from

[2] The US-BLS states employment in March 2019 was about 157 billion (seasonally adjusted).

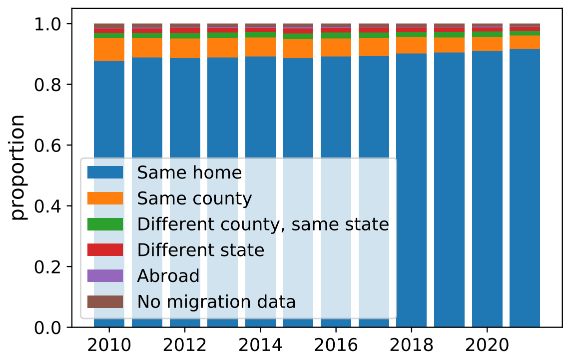


Figure 6.2: **Stacked bar plot for ASEC migration status.** Proportion of respondents with different migration status answers for 2010 to 2021 in ASEC.

occupation i to occupation j. Then, with $T_i = \sum_j T_{ij}$, the probability of a worker transitioning from occupation i to occupation j is

$$P_{ij} = \frac{T_{ij}}{T_i}. \tag{6.1}$$

As we are using survey data, we cannot know if the edges present are all of the transitions made between pairs of occupations by all workers during this time. Analysis of this type of network estimation is an active area of research for many network applications including social networks and many physiological networks. The methods include using additional information to construction social networks [145], using a node's second-order neighbourhood to quantify edge significance [146], and a rewiring method that can be used to test the significance of network metrics [147]. Statistical testing of the occupational mobility network calculated from incomplete data would be important to carry out if using the US network in future but is beyond the scope of this thesis.

Workers finding a new job can also stay in the same occupation. As we discussed in Section 3.2, the RAIS data gave us data on these self-loops however, when the occupation code is the same from month-to-month, we cannot know if the place of employment has changed, or not. Therefore, without any information available to calculate the weight of the diagonal entries of the US occupational mobility network, we assume that the probability that a worker stays within their occupation is the same for all occupations. Let α be the probability that a worker finds a new job in the same occupation, then the adjacency matrix for the occupational mobility network is

$$F_{ij} = \begin{cases} \alpha & \text{if } i = j, \\ (1-\alpha)P_{ij} & \text{if } i \neq j. \end{cases} \tag{6.2}$$

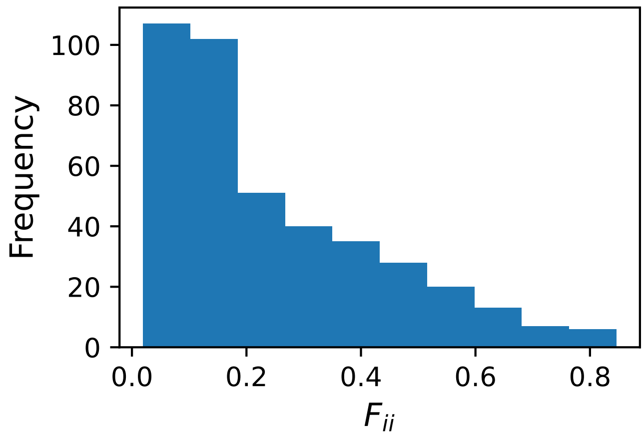


Figure 6.3: **Histogram of diagonal entries of Brazil's National Occupational Mobility Network.**

with $\alpha = 0.55$, calibrated using the annual occupational mobility rate for the US; full details can be found in [11].

With the RAIS data, we have access to information on workers' job location and can therefore count the transitions where a worker changes job but not occupation. In Figure 6.3, we can see there is variation in the diagonal of the Brazil National occupational mobility network (visualised in Figure 3.2a). This variation is present in all occupation groups and among the occupations with the strongest diagonal are Nutritionists, Professional athletes and sports referees, and Reinforced concrete frame erectors. These are all specialised occupations where we might expect a strong diagonal in the network. Given the variation in the diagonal entries of the Brazil network, in the future, we should endeavour to find data at the occupation level to calibrate α_i individually for each occupation i.

6.2 Demand shock

For illustration of the model extensions, we use the automation shock from the del Rio-Chanona et al. paper and the sigmoid function explained below.

We assume that the aggregate demand for labour is constant, that is $\mathcal{D}_t^\dagger = D_0 = \mathcal{L}$, where $\mathcal{L}$ is the total number of workers in the model. It has been shown that unemployment rates are historically quite steady while the number of hours worked per week is declining. This assumption results in occupations that feel a high level of demand shock seeing a reduction in target demand and occupations with a low level of demand shock seeing an increase in target demand.

Frey and Osborne [148] find a probability of automation for each of the 464 ACS 4-digit occupations. Del Rio-Chanona et al. use this probability as the level of automation shock felt by each occupation. Practically, the level of shock felt by each occupation is the proportion of hours worked that are no longer needed post-shock.

We use a sigmoid function to describe labour demand through time. It has been shown that the adoption of technologies tends to follow a sigmoid curve. We assume that the shock duration is 30 years, with the majority of the demand changes happening in the middle 10 years.

Assume that the target demand initial value, $\mathcal{D}^{\dagger}_0$, is equal to the steady-state realised demand of the model and that the realised demand does reach the target demand, $\mathcal{D}^{\dagger}$. Then, with a half-life of 15 years, del Rio-Chanona et al. impose the following sigmoid function on demand

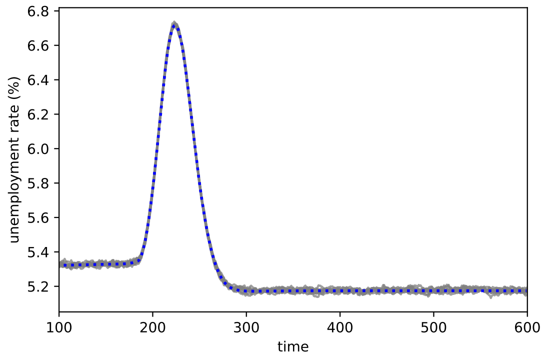
$$\mathcal{D}^{\dagger}_{i\alpha;t} = \begin{cases} d_{i\alpha;0} & \text{if } t < t_s, \\ d_{i\alpha;0} + \frac{\mathcal{D}^{\dagger}_i - d_{i\alpha;0}}{1+e^{-k(t-t_0)}} & \text{if } t \geq t_s, \end{cases} \tag{6.3}$$

where t_s is the time when the shock starts and t_0 is 15 years later.

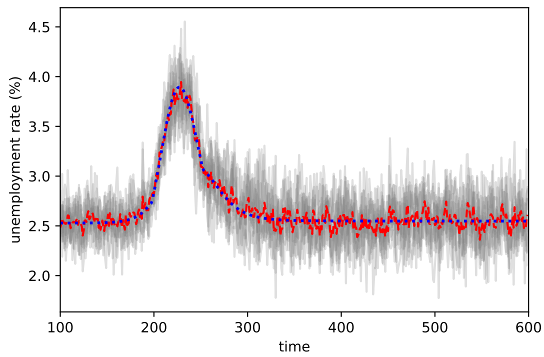
It is well studied that technological adoption follows an S-curve. [149] Since del Rio-Chanona et al. model the US automation shock, they assume a sigmoid curve for the shape of the demand shock in (6.3). This chapter seeks to investigate the impact of the new model parameters on the aggregate unemployment rate and therefore uses this sigmoid shock as an example.

6.3 Results

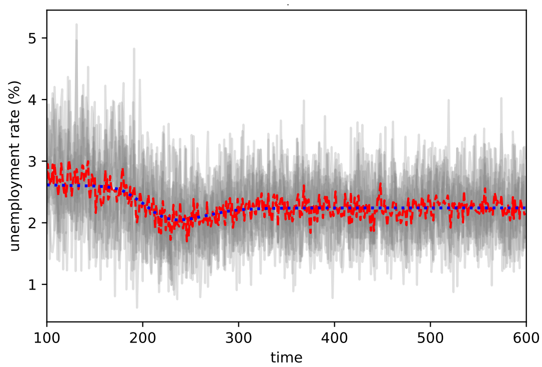
First, we run the full model for 10 simulations and compare the results with the deterministic approximation. For this illustration, we run the model with the parameter values for δ_u, δ_v, and γ taken from del Rio-Chanona et al. and $\lambda = 0.001$, $\beta_u = 5$ and $\beta_e = 1$. We see in Figure 6.4 that, although there is variation among simulation runs, as we expect, the average for the 10 runs is close to the approximation. The large fluctuations we see for fire inspectors is due to the relatively small size of this occupation. We start with 901 fire inspectors employed, compared to 24,266 medical assistants. The shape of the unemployment rate is due to the demand shock we impose at the occupation level. Aircraft mechanics have a high probability of automation (71%) which results in a spike in the unemployment rate but, after the shock, the unemployment rate returns to almost the same level. Both the fire inspector and medical assistant occupations see a decrease in their unemployment



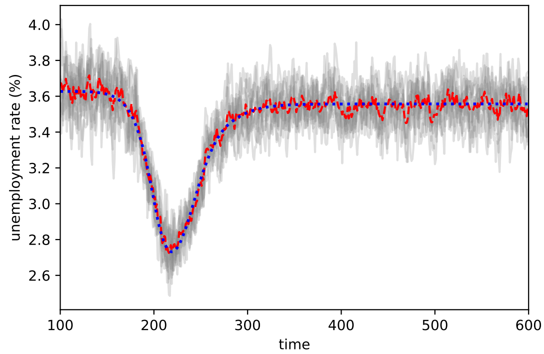
(a) Aggregate



(b) Aircraft mechanics



(c) Fire inspectors



(d) Medical assistants

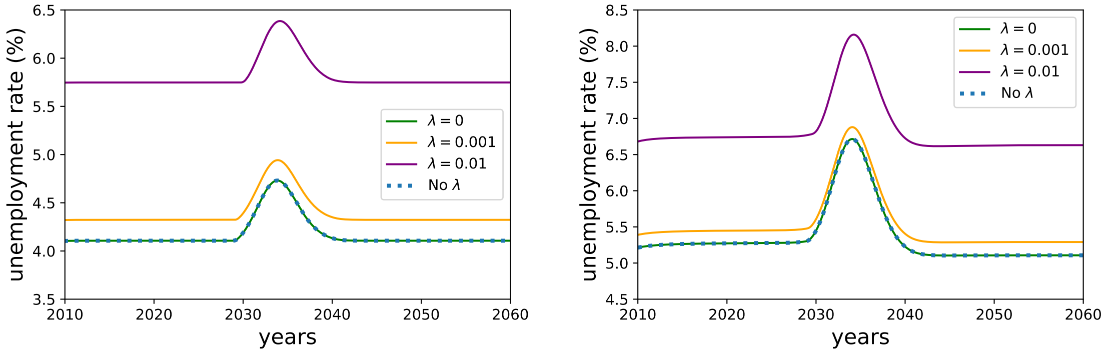
Figure 6.4: **Unemployment rate with the approximation and 10 runs of the full agent-based model.** The unemployment rate as a function of time for the occupational mobility network and the Frey and Osborne automation shock for $\lambda = 0.001$, $\beta_u = 5$ and $\beta_e = 1$.

rate during the shock, due to their relatively low probability of automation (26% and 30% respectively). However, the unemployment rate for medical assistants returns to almost the level seen before the shock and the unemployment rate for fire inspectors actually decreases.

6.3.1 On-the-job search

As we discussed above, on-the-job search plays an important role in the labour market; it directly affects the job search of unemployed workers and plays an important role in the Beveridge Curve. We use the model approximation to efficiently investigate the parameter λ.

In Figure 6.5, we show how the unemployment rate changes when we increase λ. As we expect, setting $\lambda = 0$ yields the previous results, when only unemployed



(a) Complete network

(b) Occupational mobility network

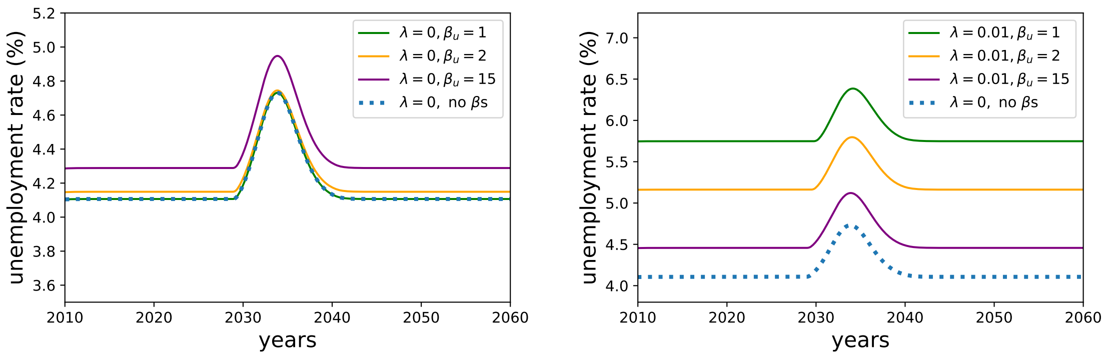
Figure 6.5: **Unemployment rate for on-the-job search.** The unemployment rate as a function of time for the Frey and Osborne automation shock for varying values of λ, the probability that an employed worker is looking for a job.

workers could apply for jobs. For the complete network in Figure 6.5a, as we increase λ, the aggregate unemployment rate increases. The shape of the unemployment rate does not vary much across the different parameters. One explanation for this increase is the huge disparity between the number of unemployed and employed workers. As more employed workers enter the job market, there are more applicants for each vacancy and so unemployed workers face more competition for jobs.

In Figure 6.5b, we see a similar relationship with the occupational mobility network as in the complete network but the unemployment rate changes shape slightly. As λ increases, the change in the unemployment rate before and after the reallocation shock decreases. In the original model, the aggregate unemployment rate changed by -0.16% (before the shock compared to after the reallocation) and for $\lambda = 0.01$, this change is only -0.11%.

6.3.2 Multiple job applications

Now, we investigate the effect of multiple applications, and the parameters β_u and β_e. In Figure 6.6, we see the effect of varying the number of applications each unemployed worker can send at each time step (with $\beta_e = 1$) for $\lambda = 0$ and $\lambda = 0.01$ with the complete network. In Figure 6.6a we see a slight increase in the unemployment rate when unemployed workers send 15 applications per time step. This is a similar effect to when we increased λ above, the applicant pool increases causing more friction in the labour market. In Figure 6.6b we see that when employed workers are also able to apply for a job (with $\beta_e = 1$), allowing unemployed workers to send more applications decreases the unemployment rate. In this case, more unemployed workers are being



(a) $\lambda = 0$

(b) $\lambda = 0.01$

Figure 6.6: **Unemployment rate for multiple applications without mobility frictions.** The unemployment rate as a function of time for the complete network and Frey and Osborne automation shock for varying values of β_u, the number of applications unemployed workers send.

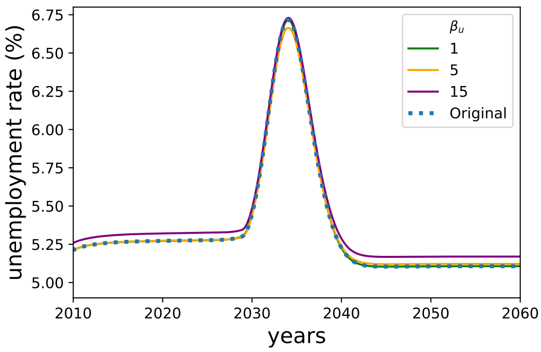
hired as we increase β_u and keep λ and β_e constant, as there are more applications from unemployed workers competing with those from employed workers.

In Figure 6.7, we see the same parameter combinations for the occupational mobility network. When only unemployed workers send applications, there is no increase in the unemployment rate as β_u increases, as we can see in Figure 6.7a. In Figure 6.7b, we can see diminishing returns. There is a large decrease in the unemployment rate when β_u increases from 1 to 2 but as we double the number of applications that unemployed workers send, the effect on the unemployment rate decreases.

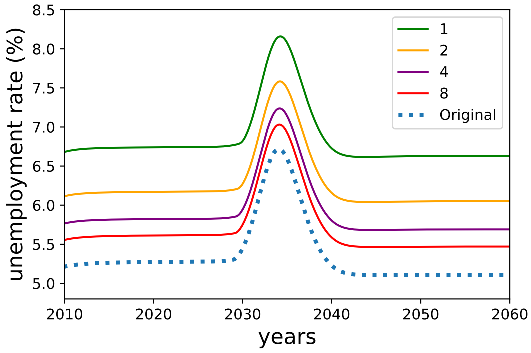
We show here that more applications usually means a lower unemployment rate. However, at some point, more applications no longer lowers the unemployment rate. This is especially true in the case of the complete network where the unemployment rate actually increases as unemployed workers send more applications when $\lambda = 0$. This has strong implications for the labour market as some unemployment benefits are dependent upon workers engaging in job search activities. If taken to the extreme, this may actually be counter-productive for the aggregate unemployment rate.

6.3.3 Wage driven dynamics

Finally, we can see the effect that adding wage driven dynamics has on the unemployment rate. We use $\psi_{ij}(\mathbf{w}) = \log(w_j)$ in (5.39). At the aggregate level, there is not much difference from the original model when we add wage pressure,



(a) $\lambda = 0$



(b) $\lambda = 0.01$

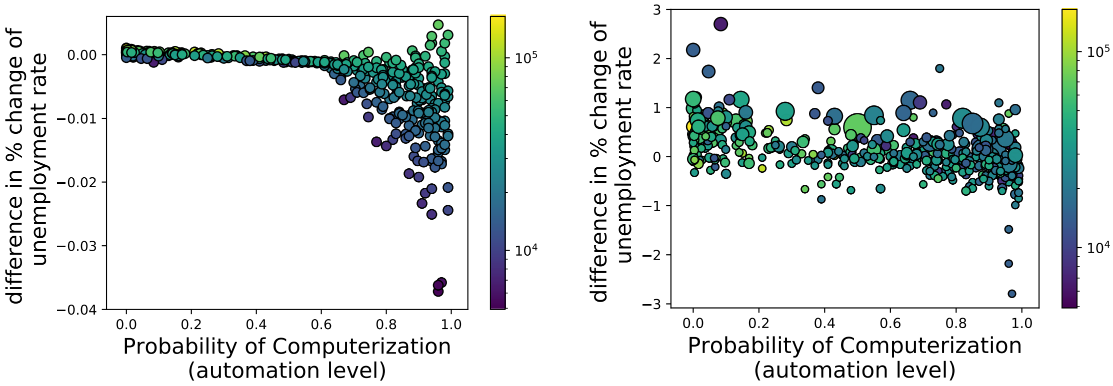
Figure 6.7: **Unemployment rate for multiple applications with mobility frictions.** The unemployment rate as a function of time for the occupational mobility network and the Frey and Osborne automation shock for varying values of β_u, number of applications unemployed workers send.

with $\beta_u = \beta_e = 1$ and $\lambda = 0$. But if we look at the effect of the automation demand shock on individual occupations, different patterns emerge.

For the complete network, we can clearly see the effect that adding wage dynamics has on the individual occupations in Figure 6.8a. Although the magnitude of these changes is small, they are clearly ordered. As every node has the same degree, only the wage differentiates the occupations. Adding wages to the model with the complete network results in occupations with lower wage (the darker dots) feeling a larger change in unemployment during the demand shock. In Figure 6.8b, the magnitude of the difference with and without wage dynamics is larger, but so are the raw percentage changes in the unemployment rate (as shown in del Rio-Chanona et al. [11]). Here, the pattern that we saw in the complete network has not materialised. We can see that the occupations with a higher in-degree (the larger dots) manage to counteract the negative effects of having a low wage, and the occupation with lower in-degree tend to be negatively affected by the introduction of wage pressure into the model.

6.4 Conclusions

With the application in this chapter, we show that the deterministic approximation to the agent-based model is accurate with a full occupational mobility network, and we explore how the new parameters shape model dynamics and influence the unemployment rate. We investigated the impact of the new parameters with both the complete network, and the occupational mobility network, providing further qualitative insights into how they affect the steady state of the model. The focus



(a) Complete network

(b) Occupational mobility network

Figure 6.8: **Effect of adding wage dynamics on the percentage change in the unemployment rate per occupation.** The difference between the percentage change in the unemployment rate without and with wage pressure. The size of the dot is proportional to the in-degree in each network and the colour represents the log of the median wage.

is on understanding the role and interaction of the mew parameters, and as such, this chapter does not calibrate the new parameters to data, nor seek to provide quantitative interpretations on the US labour market. Future work could include network estimation analysis, combining of different shocks such as the digital and green transitions, and calibrating the new parameters in order to gain quantitative insights.

Chapter 7

Conclusions

In this thesis, we have extended one of the most granular agent-based labour market models and used it to study two growth pathways for Brazil. The aim of this work is to understand how labour frictions might interact with the post-carbon transition. Using a network to capture occupational and regional mobility, we quantify the unemployment and vacancy impacts of different transition scenarios for Brazil. With the extensions, the model now better reflects the reality of the labour market, and with the extended deterministic approximation, the model can still be used practically with a large number of occupations, and regions.

The main research questions answered in this thesis are the following:

- How do workers transition between regions and occupations in Brazil?
- What modelling approaches can we use to capture these transitions?
- What regions and occupations are at risk of negative labour market impacts of two growth pathways in Brazil?
- How do those growth pathways impact unemployment and vacancy rates?
- How can the stochastic model framework be extended further to capture more labour market mechanisms?
- Can we still derive a deterministic approximation for this added complexity?
- How do different job searching intensities affect the unemployment rate?

To answer these, we first discussed the current literature on skills, geography, and labour networks, as well as the recent success of economic agent-based models in predicting emergent behaviour. We considered how skills and especially skill mismatch might affect workers and the economy in a green transition [9, 48]. We discussed the many different methods for incorporating geography into a labour market model, through gravity models, spatial relationships, and worker mobility networks [100, 66, 35], including work by Fair and Guerrero [35] that generates these mobility network endogenously for the first time. In sum, these studies point to a continued need for labour market models that consider mobility frictions and regional variation for studying the labour market impact of the post-carbon transition. With this motivation, we discussed both technical and empirical contributions.

The first technical contribution was extending the del Rio-Chanona model in Chapter 3 to include regional mobility. Using an empirical network to capture the mobility frictions, the model describes how workers may react to different demand changes. We discussed the deterministic approximation which allows us to run the model quickly, for substantial populations and large networks, before showing the second order effects on unemployment that a regional demand shock could have.

In Chapter 5, we relaxed three simplifying assumptions in the model. Employed workers can now apply for jobs, workers applying for a new job can send multiple applications, and wages now influence where workers apply. We took care to discuss the deterministic approximation of the full stochastic agent-based model with these new assumptions, to keep the model practical for use with highly granular occupations and regions. We discussed the error introduced by certain assumptions and the occupancy distribution approximation, and studied the approximation for different parameter values.

As we leveraged in this thesis, one of the benefits of agent-based modelling is the flexibility to add further complexity. Adding regional mobility and relaxing three modelling assumptions takes our model closer to reality, however, there are still many more mechanisms that could be added, to answer different research questions. For example, adding age to workers, and subsequently entry and exit of agents in the model; these could answer questions around early retirement, or the impact of an ageing population. Other worker heterogeneity such as gender, education, or skills could be added, as could dynamic wages as we discussed in Section 5.4.

One limitation of our model is that it relies on the assumption that worker mobility patterns will be as they have been in recent years. The recent history of worker transitions is captured by the occupational mobility network and then workers in the model apply for vacancies according to the probability based on these historic transitions. These historical transitions are motivated by factors such as the economy, wages, and vacancies so in order to make predictions of future occupational mobility, one could calibrate the model to a network that does not capture these historic transitions, but instead captures capabilities.

For the empirical contribution, in Chapter 4, we presented a case study for Brazil. First, we constructed and visualised the regional occupational mobility network for Brazil's labour market. Using a network enabled us to find patterns and structure in worker transitions. We found that most transitions happen within regions, particularly for Acre, Rondônia, RSudeste, and São Paulo, and that within regions, transitions happen more between similar occupations. We showed that occupational mobility is clustered between occupations with similar wages and gender split, as well as the labour demand change of the scenarios we study.

Our analysis provided quantitative predictions for the unemployment and vacancy rate impacts of two growth pathways in Brazil. We were able to identify possible inequalities that an unmanaged productivity increase in Brazil could exacerbate, and we accounted for labour mobility frictions by using the empirical mobility network. In future, our model could be used to identify patterns in the labour market impacts that an economic transition could have on both workers and employers and the framework in Chapter 4 could be used for any economic scenario with any country's mobility network to investigate these impacts. The analysis could be taken a step further by coupling our model with an external macro model, as we discussed in Section 4.4. This would allow us to estimate the effect labour frictions could have on different transition scenarios, by feeding back the labour market model's realised demand to the macro model for the next time step.

Finally, in Chapter 6, we presented a case study of an automation demand shock scenario for a city in the US. We investigated how the new parameters affect the aggregate unemployment rate, as well as comparing the unemployment outcomes of this automation scenario with and without mobility frictions. With this, and the Brazil case study, we illustrate how our model can be used to gain insight into how labour mobility frictions could interact with different demand shocks, and to identify occupations and regions where policy makers should focus efforts to mitigate the negative impact of possible demand shocks.

One limitation of the empirical analysis in this thesis is that data limitations required us to use the US calibrated parameters throughout. This is appropriate in Chapter 6 where we study the US, however this should be considered when discussing the results in Chapter 4. While the qualitative outcomes are robust to the choice of the model parameters, to provide quality, quantitative estimates would require further calibration of the parameters.[1]

[1]Forthcoming in the paper, on which Chapter 4 is based.

In sum, we presented a labour market model that captures regional and mobility frictions, often missed by existing models, with a deterministic approximation which enables us to model highly granular occupations and large populations. We discussed different avenues for future work in three main strands; adding further complexity to the model, coupling with a macroeconomic model, and calibration. We hope that this model will be used to study the labour market impacts of future economic transitions, as well as leveraging the approximation for model calibration.

Bibliography

[1] Mengdi Liu, Ronald Shadbegian, and Bing Zhang. Does environmental regulation affect labor demand in China? Evidence from the textile printing and dyeing industry. *Journal of Environmental Economics and Management*, 86:277–294, 2017.

[2] Francesco Vona. Job losses and the political acceptability of climate policies: An amplified collective action problem. *IDEAS Working Paper Series from RePEc*, 2018.

[3] W. Reed Walker. Environmental regulation and labor reallocation: Evidence from the Clean Air Act. *American Economic Review*, 101(3):442–447, 2011.

[4] Eli Berman and Linda T.M. Bui. Environmental regulation and labor demand: Evidence from the South Coast Air Basin. *Journal of Public Economics*, 79(2):265–295, 2001.

[5] Kenneth Castellanos and Garth Heutel. Unemployment, labor mobility, and climate policy. *National Bureau of Economic Research*, 2019.

[6] Marc A.C. Hafstead and Roberton C. Williams. Unemployment and environmental regulation in general equilibrium. *Journal of Public Economics*, 160:50 – 65, 2018.

[7] Junghyun Lim, Michaël Aklin, and Morgan R. Frank. Location is a major barrier for transferring us fossil fuel employment to green jobs. *Nature Communications*, 14, 2023.

[8] Misato Sato, Leanne Cass, Aurélien Saussay, Francesco Vona, and Leo Mercer. Skills and wage gaps in the low-carbon transition: comparing job vacancy data from the us and uk. Technical report, Grantham Research Institute on Climate Change and the Environment and Centre for Climate Change Economics and Policy, London School of Economics and Political Science, 2023.

[9] Joris Bücker, R. Maria del Rio-Chanona, Anton Pichler, Matthew C. Ives, and J. Doyne Farmer. Employment dynamics in a rapid decarbonization of the us power sector. *Joule*, 2025.

[10] W. Brian Arthur. Foundations of complexity economics. *Nature reviews physics*, 3(2):136–145, 2021.

[11] R Maria Del Rio-Chanona, Penny Mealy, Mariano Beguerisse-Díaz, François Lafond, and J Doyne Farmer. Occupational mobility and automation: A data-driven network model. *Journal of the Royal Society interface*, 18(174):20200898–20200898, 2021.

[12] Michael Wooldridge and Nicholas R. Jennings. Intelligent agents: theory and practice. *The Knowledge Engineering Review*, 10(2):115–152, 1995.

[13] Joshua M. Epstein and Robert L. Axtell. *Growing Artificial Societies: Social Science from the Bottom Up*. The MIT Press, 10 1996.

[14] Joshua M. Epstein. Chapter 34 remarks on the foundations of agent-based generative social science. In L. Tesfatsion and K.L. Judd, editors, *Handbook of Computational Economics*, volume 2 of *Handbook of Computational Economics*, pages 1585–1604. Elsevier, 2006.

[15] Federico Bianchi and Flaminio Squazzoni. Agent-based models in sociology. *WIREs Computational Statistics*, 7(4):284–306, 2015.

[16] Enrico Moretti. *The new geography of jobs*. Houghton Mifflin Harcourt, 2012.

[17] Ralph Trancoso. Changing amazon deforestation patterns: Urgent need to restore command and control policies and market interventions. *Environmental Research Letters*, 16(4):041004, 2021.

[18] UNEP-CCC. Emissions gap report. *UNEP*, 2021.

[19] Brazil. Brazil First NDC 2023 adjustment, 2023.

[20] J. Ferreira Filho and M. Hanusch. A macroeconomic perspective of deforestation in Brazil's Legal Amazon. Policy Research Working Papers 10162, World Bank, 2022.

[21] Peter Gladoic Hakansson and Predrag Bejakovic. Labour market resilience, bottlenecks and spatial mobility in Croatia. *Eastern Journal of European Studies*, 11:5–25, December 2020.

[22] K. Attström, S. Niedlich, K. Sandvliet, H.M. Kuhn, and E. Beavor. Mapping and analysing bottleneck vacancies in EU labour markets. *Overview report Final, Brussels: The European Commission*, 2014.

[23] R. Jason Faberman, Andreas I Mueller, Ayşegül Şahin, and Giorgio Topa. *Job Search Behavior among the Employed and Non-Employed [electronic resource].* Working paper series (National Bureau of Economic Research : Online) ; working paper no.23731. National Bureau of Economic Research, Cambridge, Mass., 2017.

[24] Bruce Fallick and Charles A. Fleischman. Employer-to-employer flows in the U.S. labor market: The complete picture of gross worker flows. Finance and Economics Discussion Series 2004-34, Board of Governors of the Federal Reserve System (U.S.), 2004.

[25] Daniel Martin and Olivier Pierrard. On-the-job search and cyclical unemployment: Crowding out vs. vacancy effects. *Journal of Economic Dynamics and Control*, 44:235–250, 2014.

[26] Toshihiko Mukoyama, Christina Patterson, and Ayşegül Şahin. Job search behavior over the business cycle. *American Economic Journal: Macroeconomics*, 10(1):190–215, 2018.

[27] Maria Balgova, Simon Trenkle, Christian Zimpelmann, and Nico Pestel. Job search during a pandemic recession: Survey evidence from the Netherlands. *IZA Discussion Paper No. 14180*, 3 2021.

[28] John C Haltiwanger, Henry R Hyatt, Lisa B Kahn, and Erika McEntarfer. Cyclical job ladders by firm size and firm wage. *American Economic Journal: Macroeconomics*, 10(2):52–85, 2018.

[29] Nathan Deutscher. Job-to-job transitions and the wages of Australian workers. Treasury Working Paper 2019-07, The Australian Government, The Treasury, Canberra, 2019.

[30] Wayne B Gray, Ronald J Shadbegian, Chunbei Wang, and Merve Meral. Do EPA regulations affect labor demand? Evidence from the pulp and paper industry. *Journal of Environmental Economics and Management*, 68(1):188–202, 2014.

[31] Richard D Morgenstern, William A Pizer, and Jhih-Shyang Shih. Jobs versus the environment: An industry-level perspective. *Journal of Environmental Economics and Management*, 43(3):412–436, 2002.

[32] Michael Greenstone. The impacts of environmental regulations on industrial activity: Evidence from the 1970 and 1977 clean air act amendments and the census of manufactures. *Journal of Political Economy*, 110(6):1175–1219, 2002.

[33] Heidi Garrett-Peltier. 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, 2017.

[34] Guillermo Montt, Kirsten S Wiebe, Marek Harsdorff, Moana Simas, Antoine Bonnet, and Richard Wood. Does climate action destroy jobs? An assessment of the employment implications of the 2-degree goal. *International labour review*, 157(4):519–556, 2018.

[35] Kathyrn Fair and Omar Guerrero. Endogenous labour flow networks. *EPJ Data Science*, 2025.

[36] Panagiotis Fragkos and Leonidas Paroussos. Employment creation in EU related to renewables expansion. *Applied energy*, 230:935–945, 2018.

[37] European Commission. A policy framework for climate and energy in the period from 2020 to 2030. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*, COM/2014/015.

[38] Manish Ram, Arman Aghahosseini, and Christian Breyer. Job creation during the global energy transition towards 100% renewable power system by 2050. *Technological forecasting & social change*, 151:119682, 2020.

[39] Omar Hesham AlShehabi. Energy and labour reform: Evidence from iran. *Journal of Policy Modeling*, 34(3):441–459, 2012.

[40] Omar Hesham AlShehabi. Modelling energy and labour linkages: A cge approach with an application to iran. *Economic Modelling*, 35:88–98, 2013.

[41] Tilak S. Liyanaarachchi, Athula Naranpanawa, and Jayatilleke S. Bandara. Impact of trade liberalisation on labour market and poverty in sri lanka. an integrated macro-micro modelling approach. *Economic Modelling*, 59:102–115, 2016.

[42] Kevin Donovan and Todd Schoellman. The role of labour market frictions in structural transformation. *Pathfinding Paper for the CEPR-led research programme on Structural Transformation and Economic Growth (STEG)*, 2021.

[43] Stéphane Moyen and Jean-Guillaume Sahuc. Incorporating labour market frictions into an optimising-based monetary policy model. *Economic Modelling*, 22(1):159–186, 2005.

[44] Anneleen Vandeplas, Istvan Vanyolos, Mauro Vigani, and Lukas Vogel. The possible implications of the green transition for the eu labour market. *Discussion Paper 176, European Commission*, 2022.

[45] OECD. Employment implications of green growth: Linking jobs, growth and green policies. *OECD Report for the G7 environment ministers*, 2017.

[46] A Bowen, K Kuralbayeva, and Eileen Tipoe. Characterising green employment: The impacts of 'greening' on workforce composition. *Energy Economics*, 72, 2018.

[47] Aurélien Saussay, Misato Sato, Francesco Vona, and O'Kane Layla. Who's fit for the low-carbon transition? Emerging skills and wage gaps in job and data. FEEM Working Paper 329079, Fondazione Eni Enrico Mattei (FEEM), October 2022.

[48] Frank M.H. Neffke, Ljubica Nedelkoska, and Simon Wiederhold. Skill mismatch and the costs of job displacement. *CESifo Working Paper No. 9703*, 2022.

[49] Fatih Guvenen, Burhan Kuruscu, Satoshi Tanaka, and David Wiczer. Multidimensional skill mismatch. *American Economic Journal: Macroeconomics*, 12(1):210–44, January 2020.

[50] Kai Liu, Kjell G. Salvanes, and Erik Ø. Sørensen. Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, 84:3–17, 2016. European Labor Market Issues.

[51] Seamus McGuinness, Konstantinos Pouliakas, and Paul Redmond. Skills mismatch: Concepts, measurement and policy approaches. *Journal of Economic Surveys*, 32(4):985–1015, 2018.

[52] Müge Adalet McGowan and Dan Andrews. Skill mismatch and public policy in oecd countries. *OECD Economics Department Working Papers*, 2015.

[53] Olga Kupets. Skill mismatch and overeducation in transition economies. *IZA World of Labor*, 12 2015.

[54] Maureen Lankhuizen, Dario Diodato, Anet Weterings, Olga Ivanova, and Mark Thissen. Identifying labour market bottlenecks in the energy transition: a combined io-matching analysis. *Economic Systems Research*, 35(2):157–182, 2023.

[55] Adrien Bilal. The geography of unemployment. Working Paper 29269, National Bureau of Economic Research, September 2021.

[56] C. Jara-Figueroa, Bogang Jun, Edward L. Glaeser, and Cesar A. Hidalgo. The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences*, 115(50):12646–12653, 2018.

[57] Henry G Overman and Xiaowei Xu. Spatial disparities across labour markets. *Oxford Open Economics*, 3(Supplement 1):i585–i610, 07 2024.

[58] Heather Dickey and Maire Carroline Magante. Exploring the drivers of internal labour migration for the regions of great britain. *Regional Studies*, 58(7):1390–1409, 2024.

[59] Kyle Butts, Taylor Jaworski, and Carl Kitchens. The urban wage premium in historical perspective. Working Paper 31387, National Bureau of Economic Research, June 2023.

[60] Sabine D'Costa and Henry G. Overman. The urban wage growth premium: Sorting or learning? *Regional Science and Urban Economics*, 48:168–179, 2014.

[61] Boris Hirsch, Elke J. Jahn, Alan Manning, and Michael Oberfichtner. The urban wage premium in imperfect labor markets. *Journal of Human Resources*, 57(S):S111–S136, 2022.

[62] Andries Brandsma, d'Artis Kancs, and Damiaan Persyn. Modelling migration and regional labour markets: an application of the new economic geography model rhomolo. *Journal of Economic Integration*, 29, 2014.

[63] James Andrew Giesecke and John Robert Madden. *Regional computable general equilibrium modeling*, pages 379 – 475. North-Holland, 1st edition, 2013.

[64] Moritz Kuhn, Iourii Manovskii, and Xincheng Qiu. The geography of job creation and job destruction. Working Paper 29399, National Bureau of Economic Research, October 2021.

[65] Christopher A. Pissarides. *Equilibrium unemployment theory*. MIT Press, Cambridge, Mass, 2nd ed. edition, 2000.

[66] Olivier Blanchard and Lawrence Katz. Regional evolutions. *Brookings Papers on Economic Activity*, 23(1):1–76, 1992.

[67] Solmaria Halleck Vega and J. Paul Elhorst. Modelling regional labour market dynamics in space and time*. *Papers in Regional Science*, 93(4):819–842, 2014.

[68] Christoph Kern and Petra Stein. Modelling Decision-Making Processes of Regional Mobility in a Dyadic Framework. *European Sociological Review*, 34(4):433–451, 06 2018.

[69] R Maria del Rio-Chanona, Penny Mealy, Anton Pichler, François Lafond, and J Doyne Farmer. Supply and demand shocks in the COVID-19 pandemic: An industry and occupation perspective. *Oxford review of economic policy*, 36(Supplement 1):S94–S137, 2020.

[70] Jonathan I. Dingel and Brent Neiman. How many jobs can be done at home? *Journal of Public Economics*, 189:104235, 2020.

[71] Daniel Garrote Sánchez, Nicolas Gomez Parra, Caglar Ozden, Bob Rijkers, Mariana Viollaz, and Hernan Winkler. Who on Earth Can Work from Home? *The World Bank Research Observer*, 36(1):67–100, 03 2021.

[72] Lukas Althoff, Fabian Eckert, Sharat Ganapati, and Conor Walsh. The geography of remote work. *Regional Science and Urban Economics*, 93:103770, 2022.

[73] Cem Özgüzel Davide Luca and Zhiwu Wei. The new geography of remote jobs in europe. *Regional Studies*, 0(0):1–19, 2024.

[74] Aidan While and Will Eadson. Zero carbon as economic restructuring: spatial divisions of labour and just transition. *New Political Economy*, 27(3):385–402, 2022.

[75] M Newman. *Networks: An Introduction.* Oxford University Press, Oxford, 2010.

[76] Albert-László Barabási. *Network science.* Cambridge University Press, Cambridge, 2016.

[77] James D. Montgomery. Social networks and labor-market outcomes: Toward an economic analysis. *The American Economic Review*, 81(5):1408–1418, 1991.

[78] Antoni Calvó-Armengol and Matthew O. Jackson. The effects of social networks on employment and inequality. *American Economic Review*, 94(3):426–454, June 2004.

[79] Antoni Calvó-Armengol and Matthew O. Jackson. Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory*, 132(1):27–46, 2007.

[80] Carlo Gianelle. *Essays on labor mobility and network analysis.* PhD thesis, Siena University, 2010.

[81] Omar A. Guerrero and Robert L. Axtell. Employment growth through labor flow networks. *PLOS ONE*, 8(5):1–12, 05 2013.

[82] Omar A. Guerrero and Eduardo López. Firm-to-firm labor flows and the aggregate matching function: A network-based test using employer–employee matched records. *Economics Letters*, 136(C):9–12, 2015.

[83] Omar A. Guerrero and Eduardo López. Understanding unemployment in the era of big data: Policy informed by data-driven theory. *Policy & Internet*, 9(1):28–54, 2017.

[84] Robert L. Axtell, Omar A. Guerrero, and Eduardo López. Frictional unemployment on labor flow networks. *Journal of Economic Behavior and Organization*, 160:184–201, 2019.

[85] Dominik Hartmann, Cristian Jara Figueroa, Mary Kaltenberg, and Paulo Gala. Mapping Stratification: The Industry-Occupation Space Reveals the Network Structure of Inequality. *SSRN Electronic Journal*, jun 2019.

[86] Ian M. Schmutte. Free to move? a network analytic approach for learning the limits to job mobility. *Labour Economics*, 29:49–61, 2014.

[87] Aleksandr Christenko. Automation and occupational mobility: A task and knowledge-based approach. *Technology in Society*, 70:101976, 2022.

[88] Penny Mealy, R Maria del Rio-Chanona, and J. Doyne Farmer. What you do at work matters: New lenses on labour. *SSRN Electronic Journal*, Apr 2018.

[89] Frank M.H. Neffke, Anne Otto, and Antje Weyh. Inter-industry labor flows. *Journal of Economic Behavior and Organization*, 142:275–292, 2017.

[90] Neave O'Clery and Stephen Kinsella. Modular structure in labour networks reveals skill basins. *Research Policy*, 51:104486, 6 2022.

[91] Jan Sebastian Nimczik. Job mobility networks and endogenous labor markets. Vfs annual conference 2017 (vienna): Alternative structures for money and banking, Verein für Socialpolitik / German Economic Association, 2017.

[92] Siwei Cheng and Barum Park. Flows and boundaries: A network approach to studying occupational mobility in the labor market. *The American journal of sociology*, 126(3):577–631, 2020.

[93] Jonas Toubøl and Anton Grau Larsen. Mapping the social class structure: From occupational mobility to social class categories using network analysis. *Sociology*, 51(6):1257–1276, 2017.

[94] Anton Pichler, Marco Pangallo, R. Maria del Rio-Chanona, François Lafond, and J. Doyne Farmer. Forecasting the propagation of pandemic shocks with a dynamic input-output model. *Journal of Economic Dynamics and Control*, 144:104527, 2022.

[95] IAIN D. COUZIN, JENS KRAUSE, RICHARD JAMES, GRAEME D. RUXTON, and NIGEL R. FRANKS. Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, 218(1):1–11, 2002.

[96] Robert Axtell. Chapter 3 - Endogenous firm dynamics and labor flows via heterogeneous agents. In Cars Hommes and Blake LeBaron, editors, *Handbook of Computational Economics*, volume 4 of *Handbook of Computational Economics*, pages 157–213. Elsevier, 2018.

[97] Esteban Moro, Morgan R Frank, Alex Pentland, Alex Rutherford, Manuel Cebrian, and Iyad Rahwan. Universal resilience patterns in labor markets. *Nature communications*, 12(1):1972–1972, 2021.

[98] William Henry Beveridge. *Full Employment in a Free Society: A Report by William H. Beveridge...* G. Allen and Unwin, 1944.

[99] International Labour Organization. What is skills mismatch and why should we care? `https://www.ilo.org/resource/article/what-skills-mismatch-and-why-should-we-care`, 2024. Accessed: 18 June 2025.

[100] Raul Ramos. Gravity models: A tool for migration analysis. *IZA World of Labor*, pages 239–239, February 2016.

[101] Robert M. Beyer, Jacob Schewe, and Hermann Lotze-Campen. Gravity models do not explain, and cannot predict, international migration dynamics. *Palgrave Communications*, 9(1):1–10, December 2022.

[102] Juan Caballero Reina, Jesus Crespo Cuaresma, Katharina Fenz, Jakob Zellmann, Teodor Yankov, and Amr Taha. Gravity Models for Global Migration Flows: A Predictive Evaluation. *Population Research and Policy Review*, 43(2):1–17, April 2024.

[103] James Truscott and Neil M. Ferguson. Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. *PLOS Computational Biology*, 8(10):1–12, 10 2012.

[104] M U G Kraemer, N Golding, D Bisanzio, S Bhatt, D M Pigott, S E Ray, O J Brady, J S Brownstein, N R Faria, D A T Cummings, O G Pybus, D L Smith, A J Tatem, S I Hay, and R C Reiner. Utilizing general human movement models

to predict the spread of emerging infectious diseases in resource poor settings. *Scientific Reports*, 2019.

[105] Christian Alis, Erika Fille Legara, and Christopher Monterola. Generalized radiation model for human migration. *Scientific Reports*, 2021.

[106] Lucas Kluge, Anders Levermann, and Jacob Schewe. Radiation model for migration with directional preferences. *Phys. Rev. E*, 106:064138, Dec 2022.

[107] Catherine Tong, Omar A Guerrero, Eduardo Lopez, and Felix Reed-Tsochas. Diffusing workers in a multiplex world. Technical report, Saïd Business School WP, 2017.

[108] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 07 2014.

[109] Gabriel Ulyssea. Firms, informality, and development: Theory and evidence from brazil. *American Economic Review*, 108(8):2015–47, August 2018.

[110] Joao Alberto De Negri, Paulo Furtado de Castro, Natalia Ribeiro de Souza, and Jorge Saba Arbache. Mercado formal de trabalho: Comparação entre os microdados da rais e da pnad. *Instituto de Pesquisa Econômica Aplicada (IPEA)*, Texto Para Discussão 840, 2001.

[111] Christopher Cornwell, Jason Rivera, and Ian M. Schmutte. Wage discrimination when identity is subjective: Evidence from changes in employer-reported race. *The Journal of Human Resources*, 52(3):719–755, 2017.

[112] Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLOS ONE*, 9(6):1–12, 06 2014.

[113] M. E. J. Newman. Mixing patterns in networks. *Phys. Rev. E*, 67:026126, Feb 2003.

[114] Yelie Yuan, Jun Yan, and Panpan Zhang. Assortativity measures for weighted and directed networks. *Journal of Complex Networks*, 9(2):cnab017, 05 2021.

[115] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008.

[116] Sergey Ioffe. Improved consistent sampling, weighted minhash and l1 sketching. In *International Conference on Data Mining*, 2010.

[117] Marianne Bertrand and Kevin F. Hallock. The gender gap in top corporate jobs. *ILR Review*, 55(1):3–21, 2001.

[118] Wolfgang Keller, Teresa Molina, and William W. Olney. The gender gap among top business executives. *Journal of Economic Behavior and Organization*, 211:270–286, 2023.

[119] Congmin Peng and Po-Wen She. Are women less likely to be managers in the uk labour market? *Economic Modelling*, 85:317–324, 2020.

[120] A. K. Berryman, J. Bücker, F. Senra de Moura, P. Barbrook-Johnson, M. Hanusch, P. Mealy, M. del Rio-Chanona, and J.D. Farmer. Modelling labour market transitions: the case of productivity shifts in brazil. Case study in: New economics models of energy innovation and transition: Addressing new questions and providing better answers. Technical report, EEIST, 2023. Available at `https://eeist.co.uk/eeist-reports/new-economic-models-of-energy-innovation-and-transition/`.

[121] Judy Jingwei Xie, Melissa Martin, Joeri Rogelj, and Iain Staffell. Distributional labour challenges and opportunities for decarbonizing the us power system. *Nature Climate Change*, 13(11):1203–1212, 2023.

[122] Steinar Westin, James J. Schlesselman, and Mieko Korper. Long-term effects of a factory closure: Unemployment and disability during ten years' follow-up. *Journal of Clinical Epidemiology*, 42(5):435–441, 1989.

[123] Andrés Rodríguez-Pose and Federico Bartalucci. The green transition and its potential territorial discontents. *Cambridge Journal of Regions, Economy and Society*, 2023.

[124] Nyiko Worship Hlongwane and Hlalefang Khobai. Renewable energy transition on employment dynamics in brics nations. *Economies*, 13(2), 2025.

[125] Patricia Helena dos Santos Martins, André Luiz Marques Serrano, Gabriel Arquelau Pimenta Rodrigues, Guilherme Fay Vergara, Gabriela Mayumi Saiki, Raquel Valadares Borges, Guilherme Dantas Bispo, Maria Gabriela Mendonça Peixoto, and Vinícius Pereira Gonçalves. Brazil's new green hydrogen industry:

An assessment of its macroeconomic viability through an input–output approach. *Economies*, 12(12), 2024.

[126] Luiza Nassif Pires, Gilberto Tadeu Lima, Pedro Romero Marques, Tainari Taioka, and José Bergamin. Amazon green recovery and labor market in brazil. Working paper, Levy Economics Institute of Bard College, 2024.

[127] Mark Horridge. The TERM model and its data base. Centre of Policy Studies/IMPACT Centre Working Papers g-219, Victoria University, Centre of Policy Studies/IMPACT Centre, July 2011.

[128] Daniel Baumgarten, Gabriel Felbermayr, and Sybille Lehwald. Dissecting between-plant and within-plant wage dispersion: Evidence from germany. *Industrial Relations: A Journal of Economy and Society*, 59(1):85–122, 2020.

[129] Andréa Bento Carvalho and Gustavo Inácio de Moraes. The brazilian coastal and marine economies: Quantifying and measuring marine economic flow by input-output matrix analysis. *Ocean and Coastal Management*, 213:105885, 2021.

[130] Samuel Wiese, Jagoda Kaszowska-Mojsa, Joel Dyer, Jose Moran, Marco Pangallo, Francois Lafond, John Muellbauer, Anisoara Calinescu, and J Doyne Farmer. Forecasting macroeconomic dynamics using a calibrated data-driven agent-based model. *arXiv preprint arXiv:2409.18760*, 2024.

[131] Giovanni Dosi, Marcelo C Pereira, Andrea Roventini, and Maria Enrica Virgillito. The labour-augmented k+ s model: A laboratory for the analysis of institutional and policy regimes. *EconomiA*, 21(2):160–184, 2020.

[132] Jason M Rivera, Ian M Schmutte, and Christopher Cornwell. Endogenous racial identity: Evidence from brazilian RAIS and PNAD data. *Working paper*, 2013.

[133] Katharina Bergant, Rui C. Mano, and Ippei Shibata. From polluting to green jobs: A seamless transition in the u.s.? *IMP Working Paper WP/22/129*, 2022.

[134] Eduardo López, Omar A. Guerrero, and Robert L. Axtell. A network theory of inter-firm labor flows. *EPJ Data Science*, 2020.

[135] Ben O'Neill. The classical occupancy distribution: Computation and approximation. *The American Statistician*, 75(4):364–375, 2021.

[136] Shigeru Fujita, Giuseppe Moscarini, and Fabien Postel-Vinay. Measuring employer-to-employer reallocation. *American Economic Journal: Macroeconomics*, 16(3):1–51, July 2024.

[137] Robert Joyce, Fabien Postel-Vinay, Peter Spittal, and Xiaowei Xu. Job opportunities after the pandemic, 2022. Available online at `https://ifs.org.uk/publications/job-opportunities-after-pandemic` (Accessed: 28 December 2024).

[138] Robert Shimer. The cyclical behavior of equilibrium unemployment and vacancies. *American economic review*, 95(1):25–49, 2005.

[139] Jan Eeckhout and Ilse Lindenlaub. Unemployment cycles. *American Economic Journal: Macroeconomics*, 11(4):175–234, October 2019.

[140] James A. Hanley. Jumping to coincidences: Defying odds in the realm of the preposterous. *The American Statistician*, 46(3):197–202, 1992.

[141] Patrick Liu, Greg Nantz, Ryan Nunn, and Jay Shambaugh. Thirteen facts about wage growth, 2017. Available online at `https://www.brookings.edu/articles/thirteen-facts-about-wage-growth/` (Accessed: 28 December 2024).

[142] Gueorgui Kambourov and Iourii Manovskii. A cautionary note on using (March) current population survey and panel study of income dynamics data to study worker mobility. *Macroeconomic dynamics*, 17(1):172–194, 2013.

[143] Bruce D Meyer, Wallace K. C Mok, and James X Sullivan. Household surveys in crisis. *The Journal of economic perspectives*, 29(4):199–226, 2015.

[144] Jason M Ward and Kathryn Anne Edwards. Assessing the link between survey interview method and survey outcomes: Evidence from the CPS and the COVID-19 pandemic. *Labour economics*, 72:102060–102060, 2021.

[145] Garry Robins Dean Lusher, Johan Koskinen. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 2012.

[146] Na Zhao, Jie Li, Jian Wang, Tong Li, Yong Yu, and Tao Zhou. Identifying significant edges via neighborhood information. *Physica A: Statistical Mechanics and its Applications*, 548:123877, 2020.

[147] Nin J Salbanya B, Carrasco-Farré C. Structure matters: Assessing the statistical significance of network topologies. *PLoS One*, 2024.

[148] Carl Benedikt Frey and Michael A Osborne. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114(C):254–280, 2017.

[149] P.A Geroski. Models of technology diffusion. *Research Policy*, 29(4):603–625, 2000.