Firms in Integrated Urban Models

Agglomeration Economies and the Dynamics of Employment Size Decisions

by

He He

Bachelor of Applied Science in Engineering Science
University of Toronto (2016)

Master of Science in Transportation
Massachusetts Institute of Technology (2018)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

September 2022

Signature of Author: ______________________________
Department of Civil and Environmental Engineering
August 16, 2022

Certified by: ______________________________
P. Christopher Zegras
Professor of Mobility and Urban Planning
Thesis Advisor

Accepted by: ______________________________
Colette L. Heald
Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

Firms in Integrated Urban Models

Agglomeration Economies and the Dynamics of Employment Size Decisions

by He He

Submitted to the Department of Civil and Environmental Engineering on August 16, 2022 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Transportation.

ABSTRACT

Jobs are key determinants of urban phenomena ranging from daily trip patterns to urban structure. Despite their importance, the representation of jobs and firms in integrated urban models is limited. Existing approaches are exceedingly static, often lack theoretical underpinnings, and rarely account for the impact of agglomeration economies.

I propose an agent-based dynamic programming structural model of firms' job creation and lay-off decisions. It models the evolutionary trajectory of firm sizes rather than discrete jumps between presumed steady states. Firms are forward-looking rational agents, attempting to follow the employment size adjustment trajectory that maximizes their present value of all future profits in face of a stochastic adjustment process. I model firms' decision-making as a continuous-time Markov decision process, solved via dynamic programming. To estimate the model's parameters, which are firm-specific, I formulate a hierarchical Bayesian estimation procedure. I repeatedly sample from the posterior distributions of the hyperparameters using a nested Gibbs and Metropolis-Hastings sampling algorithm.

With a panel micro-dataset of businesses in the Greater Boston Area, I apply the model to explore the heterogeneous impacts of agglomeration economies for manufacturing, professional services, and food and accommodation services firms. The empirical findings broadly align with urban economic theory. However, uniquely, the dynamic structural model enables me to distinguish between benefits that increase productivity and those that reduce labour market friction. Overall, I find that employment size adjustments are more costly for more skills-intensive sectors. Finally, using the estimation results from Boston, I examine the estimated impacts of a major urban rail line investment – the Green Line extension – in terms of job creation and gross production increase, and the cost of labour market frictions in terms of firms' foregone profits.

Thesis advisor: P. Christopher Zegras
Title: Professor of Mobility and Urban Planning

ACKNOWLEDGEMENTS

To my advisor, Chris, directly or indirectly, you have had a hand in not just in my research but also every aspect of my thinking and growth over the past 6 years. You have been guiding, supportive, and lenient, all when I have needed it. Your continued trust in me and advocacy of my work have meant more than you can imagine. Thank you!

To my committee, Jinhua, Moshe, Siqi – safe to say this project of mine has come a long way since our first meeting. Your deep and diverse expertise have been invaluable. Thank you for every comment, every piece of practical and technical advice, and every insightful discussion we have had.

While I am responsible for every keystroke in this document, the ideas in my dissertation are products of the nourishing environment at MIT. To my office mates, Peyman, Jintai, Jonathan, Apaar, and Ru, my colleagues Shenhao, Rounaq, Dan, Diem, and Karina, and my class mates of which there are more than I can mention here, thank you for being inspiring, provocative, productive.

They tell you how long, difficult, and lonely a PhD can be. I did not fully understand what that meant until I had to go through it. To my friends and my family – you gave it all meaning, and I genuinely could not have done it without you. Thank you!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

CDF	cumulative distribution function
CTPS	Central Transportation Planning Staff
DPDC	dynamic programming discrete choice
GBA	Greater Boston Area
IID	independent and identically distributed
I/O	input/output
IUM	integrated urban model / integrated urban modelling
LRT	light-rail transit
MAPC	Metropolitan Area Planning Council
MCMC	Markov chain Monte Carlo
MDP	Markov decision process
PDF	probability density function
PMF	probability mass function
PUMA	public use micro area
TAZ	traffic analysis zone
TFP	total factor productivity
WACC	weighted average cost of capital

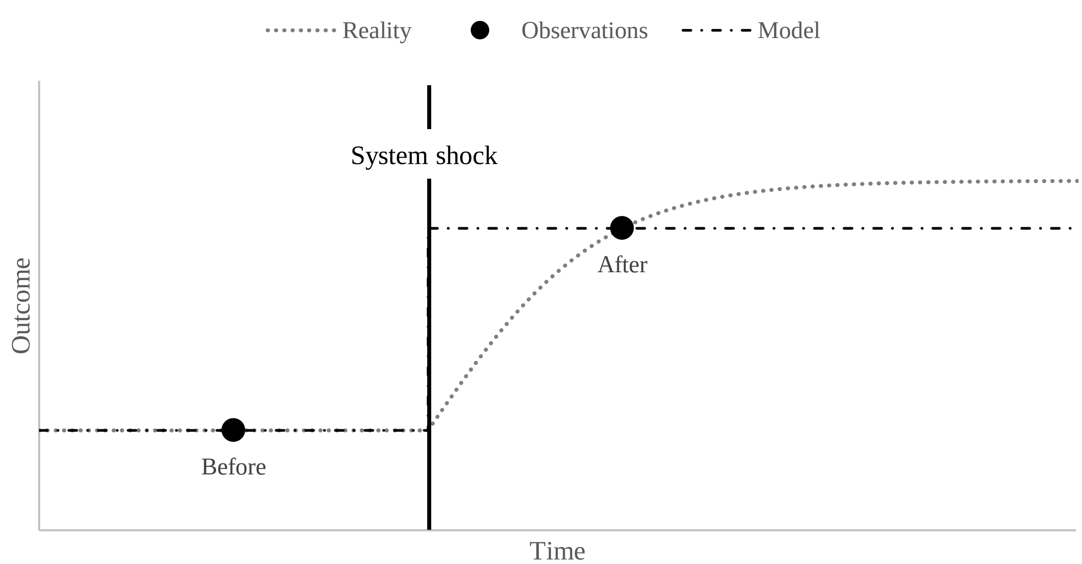
1 INTRODUCTION

1.1 Research Motivation & Objectives

The rise of human civilization is intimately tied to the first permanent settlements. These early agglomerations provided several benefits, including protection, amenities, and various geographical endowments. However, since the industrial revolution, cities' role as labour markets have become more dominant. In 1800, only an estimated 3% of the world's population was urban; that figure crossed 50% in 2007, and by 2050 a predicted 68% of people will live in cities. The industrial revolution transformed production, made it more efficient but also less tied to natural resources and more reliant on large indivisible facilities, such as factories and machinery. Agglomeration into cities enabled the scale required for this industrialized mode of production. Evidence of this relationship between the nature of work and urbanization can be seen in many developing contexts today, where the more recent industrialization has caused a rapid and, in some cases, overwhelming, wave of rural-to-urban migration that cities and planners are struggling to accommodate. In the information age, the role of cities has continued to evolve. As work has become more skill-based and knowledge-intensive, cities' role as catalysts of innovation has become more important. The frequent and diverse interactions in urban areas facilitate (tacit) knowledge transfer. How the next stages of the structural transformation of work, including automation, e-commerce, and telecommuting, change these dynamics largely remains an open question. Perhaps, the cataloguing of our collective experiences from the Covid-19 pandemic can provide some answers. Undoubtedly, role of cities as labour markets will continue to evolve.

Jobs and firms are not only drivers of urbanization but also key determinants of urban structure, intra-urban location decisions, and day-to-day trip patterns. Considering their fundamental role in the evolution of cities, they have not received too little attention in the integrated urban modelling (IUM) literature. Broadly, existing approaches fail to account for the vast heterogeneity in firm characteristics that underlie the discussed differences in urban outcomes, e.g. production technology, size, lifecycle stage, etc. and/or lack sound economic underpinnings in their modelling assumptions. Further contributing to this lack of sophistication is the static nature of current models. Cities comprise imperfect systems riddled with inertial effects and friction. Generally, they are not in steady state but instead continuously adapting to changing conditions/environments, e.g. economic up/downturns, transportation infrastructure investments, urban policy, etc. However, existing IUMs do not allow for these transitory states. That is, current modelling frameworks implicitly assume that system shocks cause jumps between steady states but abstract away the, sometimes long, paths between these steady states. This is illustrated by Figure 1-1. Error is visualized as the gap between the model and reality. Between the shock and the “after” observation, the model potentially overestimates the impact of the shock, whereas it potentially underestimates the impact following the “after” observation. Naturally, the severity of these potential errors depends on both the timing of observations and the shape of the real response curve. Systems with less inertia and friction will look more akin to the model response and thus generally be less erroneous.

Figure 1-1: Schematic illustration of potential errors when disregarding dynamics



Finally, existing IUMs are very limited in their representation of agglomeration economies even though they underlie the very existence of cities. With roots in transportation engineering, most frameworks capture various benefits associated with reduced travel costs, e.g. commuting or goods transport. However, agglomeration also increase productivity of workers and firms by facilitating knowledge transfer, sharing the costs of indivisible facilities, and increasing specialization and variety. Additionally, they reduce search costs as part of the thick market effects. In turn, these benefits could increase labour demand and lower unemployment. Relating the effects back to Figure 1-1, productivity-increasing agglomeration effects would be represented by a higher horizontal asymptote, whereas friction-reducing effects would increase the steepness of the curve, allowing it to reach steady state faster. The study of agglomeration economies is by no means straightforward. A vast theoretical literature exists on the micro-foundations of agglomeration economies. However, linking theory and empirical research remains a challenge. In almost all real-world scenarios, a multitude of underlying mechanisms of agglomeration economies work in tandem, thus making the task of disentangling and identifying the effect of each individual mechanism empirically difficult. Furthermore, the bundle of mechanisms at play for each sector and each firm can vary greatly, depending on firm characteristics and production processes. Just as the mechanisms differ, so too do the intensities and spatial scopes of the effects.

Despite these challenges, pursuing more sophisticated representations of firms and jobs is a worthwhile and necessary effort to improve the fidelity of existing IUMs. In particular, failing to account for inertial effects can lead to biased coefficient estimates, resulting in erroneous predictions; and limited modelling of agglomeration economies results in an inability to both explain and predict more nuanced spatial outcomes, such as sector-specific cluster formation and interactions between sectors. Conversely, a more complete firm-side model would enable the design of more appropriately targeted interventions, e.g. in the form of place-based policies.

In light of these deficiencies, the overarching purpose of my dissertation is to improve the modelling of businesses and jobs in IUMs. To this end, I propose a novel dynamic agent-based model of firms' hiring and firing decisions. The improvements, although intertwined, can broadly be grouped as follows:

1. **Agent-based and economically sound decision-making**: I model the decision-making processes at the level of individual businesses. This allows for an economically rational decision-making model and heterogeneity in characteristics and preferences. This is in

contrast to existing approaches, which are too aggregate and/or lacking in terms of economic underpinnings.

2. **Dynamic**: Whereas existing approaches abstract away transitory periods and effectively model discrete jumps between steady states, I make explicit the paths of development. This provides a natural way to incorporate agent history and inertial effects.
3. **Agglomeration economies:** My approach takes several steps towards improving the modelling of agglomeration economies.
    a. The proposed model separately quantifies effects acting through two channels: that increase businesses' productivity and those that reduce labour market friction. Conceptually, the productivity benefits are identified by the steady state outcomes, whereas friction reduction is identified by how quickly firms approach their steady state employment sizes.
    b. By varying the spatial decay parameter in the gravity-based accessibility formulation, I explore the spatial scope of agglomeration economies.
    c. I explore how agglomeration economies work differently for different sectors (manufacturing, professional services, and food and accommodation services), and compare and contrast the results with a priori expectation based on theory.

With the proposed modelling structure and a panel dataset of businesses in the Greater Boston Area, I estimate employment size decision models for firms in the following three sectors: manufacturing, professional services, and food and accommodation services. Using these models, I conduct two impact analyses. First, I examine the impact of transit improvements in the form of the Green Line extension on local job creation and gross production over time. In the second study, I take a closer look at the impacts of labour market frictions. Specifically, I examine the gap between actual and optimal employment size and translate that into foregone profits.

1.2 Dissertation Structure

The remainder of the dissertation is structured as follows: In Chapter 2 Background, I present reviews of two literatures: IUMs and agglomeration economies. In the section on IUMs, I begin with a brief historical overview before focusing on the modelling of firms and jobs and then the treatment of dynamics within common IUM frameworks. For the review of existing work on agglomeration economies, I draw primarily from the urban economics literature. I begin by discussing the micro-foundations underlying agglomeration economies, specifically the mechanism: sharing, matching, and knowledge spillovers. Then, taking the theory into practice, I present existing empirical evidence of the benefits, how agglomeration economies have been measured, and common empirical challenges.

In Chapter 3 The Model, I present the dynamic programming Markov model. Following an introduction of the model structure, I derive the solution, i.e. optimal policy, to the Markov decision process (MDP) that firms face. The firm's optimal policy is required to evaluate the model's likelihood function, which I present subsequently. To estimate the model, I formulate a hierarchical Bayesian model, and repeatedly sample from the posterior distributions of the hyperparameters using a joint Gibbs and Metropolis-Hastings sampling algorithm. Finally, I discuss the model's limitations, its relation to dynamic programming discrete choice models, how it could be integrated in a larger IUM framework, and other potential applications.

In Chapter 4 Application to Boston, I take the model to the data. I begin by presenting the datasets that I use for the study and discuss how I use gravity-based accessibility to measure agglomeration. Subsequently, I show various descriptive statistics to provide an overview of the study context and data. Then, I present the model specification in detail, i.e. how the Boston data are used in the model, before presenting and discussing the estimation results. I wrap up the chapter with a discussion about the empirical limitations.

In Chapter 5 Impact Analysis, I use the estimation results from the previous chapter to conduct impact analyses and, in doing so, demonstrate the potential usefulness of the proposed modelling approach. In particular, I examine the potential impacts of the Green Line extension in Boston and the foregone profits resulting from labour market frictions.

Finally, in Chapter 6 Conclusion, I summarize the key contributions and findings made in this dissertation and discuss possible avenues for future research.

2 Background

2.1 Integrated Urban Models

Overview

Integrated urban models (IUMs) are simulation tools for exploring urban development across space and over time. Generally, they comprise representations of urban land use and travel impedances and a quantitative model framework for how they interact. Urban development, at its core, captures where people live and where they conduct their out-of-home activities. However, these encompass a long list of decisions, including but not limited to individuals' mode choices and day-to-day activity schedules, households' vehicle ownership choices and residential location decisions, and firms' location and employment size decisions. Importantly, all these decisions, directly or indirectly, both depend on and affect the land use-transportation system. Capturing these feedback effects when attempting to quantify the impacts of, say, new transportation infrastructure, zoning bylaws, or place-based economic policies is a central argument for the usefulness of IUMs.

Among the first integrated models of regional land use change was Lowry's *Model of the Metropolis* (Lowry, 1964). The Lowry model, as it is known colloquially, deserves particular attention because its use of the gravity model exemplifies the practice for the first two decades of integrated modelling. Starting with exogenously determined basic (export) sector employment, the Lowry model iteratively estimates non-basic employment and spatially allocates workers to residential locations by use of the gravity model until convergence. Since their original incarnations in the largely mechanistic underpinnings represented in the four-step travel forecasting models and Lowry-type integrated models, IUMs have undergone significant improvements. These have been aided by technological (e.g., computing, data, microsimulation) and theoretical (e.g., activity-based approaches; real estate transaction-based) advances. Overall, we have seen a movement from aggregate towards disaggregate approaches, from equilibrium-to disequilibrium assumptions, and higher resolution data and greater technical sophistication (Engelberg et al., 2021).

After more than half a century of innovations and advances today's state-of-the-art IUMs look very different from Lowry's original Model of the Metropolis. The Lowry model and its contemporaries relied on empirically calibrated gravity formulations to capture the interaction between land use and travel impedance. However, the models were lacking in terms theoretical underpinnings, instead encouraging the analyst to choose the functional form for the gravity formulation that best fit the empirical colocation patterns (Lowry, 1964). Major steps have also been taken to advance the theoretical foundations of IUMs. First and foremost, the proliferation of random utility theory and its *logsum* measures provided a theoretically founded way to quantify the benefits of the land use-transport system (Ben-Akiva and Lerman, 1985). Utility-based frameworks have since become the de facto standard in the integrated modelling literature. In parallel, Anas (1983) showed the equivalence between the negative exponential gravity formulation and multinomial logit models, thus providing theoretical foundation for the whole class of gravity-based spatial interaction models.

Whereas early IUMs modelled decision makers as homogeneous masses, most modern modelling frameworks are, to some extent, disaggregated. This prevents potential aggregation biases, makes explicit heterogeneity in preferences, and enables analysis of uneven impacts across different population characteristics. The most sophisticated models are agent-based, modelling decisions of individual agents at the level of decision-makers. The data for each agent, including socio-demographic attributes, location information, social linkages, etc., are stored in persistent databases, such that heterogeneity in characteristics and preferences remain consistent across different sub-models.

Improvements in computational power and data collection methods have enabled modern IUMs to become more granular in their representation of both space and time. For example, the original application of the Lowry to Pittsburgh divided the city into 15 zones by distance to the city centre. By comparison, it is not uncommon for current IUMs to operate with zone counts in the thousands and use information from individual parcels (Engelberg et al., 2021). Advancements have also been made to model the temporal dimension more authentically since the Lowry model and equilibrium-based spatial interaction models only allowed for a static solution. Most modern IUM frameworks have broken with strict equilibrium conditions and moved towards gradual adjustments or steady state solutions. However, these processes remain ad hoc rather than theoretically founded. A more elaborate discussion of the modelling of dynamics in IUMs is found in the subsequent sections.

Considering these developments, current state-of-the-art IUM frameworks are without a doubt far more sophisticated than the early spatial interaction models. However, this has come at the very real price of increased complexity, making the models more difficult to communicate and less transparent. On the other hand, the practical benefits remain somewhat elusive and too often difficult to document. Consequently, the vast majority of real-world IUM applications, e.g. by metropolitan planning organizations, are of frameworks that are much simpler but more practical than the academic state-of-the-art (Engelberg et al., 2021).

Over the years, numerous articles have provided overviews of the field, see e.g. Wegener (2004), Hunt et al. (2005), Iacono et al. (2008), Acheampong and Silva (2015), and Engelberg et al (2021). Rather than repeat their work, I focus my attention on two aspects of particular relevance to my model: the modelling of jobs and firms, and the treatment of dynamics. These are the topics of the following two sections.

Jobs and Firms

Although the feedback effects between the spatial distributions of households and jobs has been a core component of urban models from their inception, academic efforts in modelling since have disproportionately been focused on exploring the behaviour of individuals and households. Innovations in choice modelling have yielded theoretically founded microscopic models of activity and travel behaviour, vehicle ownership decision, and residential location choices. However, firm-side models have generally lagged behind, often lacking in either granularity or sound theory. This gap is even more apparent in the context of IUMs as the majority of firm and employment modelling efforts are stand-alone rather than integrated. The relative dearth of research on firm behaviour is likely not a result of negligence but rather the additional hurdles associated with studying firms compared to individuals and households. In particular, micro-level data for firms are rarely publicly available if even collected, and there is considerable heterogeneity in firm behaviour across industrial sectors and stages of firm life cycles (Kumar and Kockelman, 2009). Importantly, birth of new firms, relocation of existing firms, and growth within existing firms all play their part in total job creation, and no model without all three captures the full picture. However, it is worth noting that almost three-quarters of jobs in the U.S. are created by existing firms (U.S. Bureau of Labor Statistics, 2022). Yet, the bulk of research that does exist on firm behavior pertains to their location decisions, whereas employment expansion and contraction, i.e. hiring and firing decisions at micro-level, has received far less attention. Nonetheless, I provide a review of the latter here.

Mirroring the IUM literature as a whole, advances in firm and employment models since the Lowry model largely follow two branches: equilibrium-based spatial input-output models and agent-based microsimulation models. Building on Leontief's (1966) seminal work modelling inter-industry economic relationships using input-output matrices, spatial input-output models introduce to this framework the spatial dimension. These models typically comprise zonal-level production functions for firms, utility functions for workers, and transport cost functions for goods and people. By making explicit the transport costs, these models determine the spatial equilibrium of production and consumption of firms and workers by balancing the inputs and outputs of each zone. Notable examples of equilibrium-based spatial input-output models include MEPLAN (Echenique et al., 1990), PECAS (Hunt and Abraham, 2005), TRANUS (de la Barra, 2011), METROSCOPE (Metro Research Center, 2016), and RELU-TRAN (Anas and Liu, 2007).

With roots in the economics tradition, this family of models has a strong theoretical foundation. However, they tend to be more aggregate than agent-based models. They generally treat zones as the unit of analysis. However, zones are neither monoliths nor decision-makers. Thus, while the approach captures the location-specific attributes that are necessary to establish spatial equilibrium, it disregards any heterogeneity within the groups of individuals, households, and firms that the zones represent. The high level of aggregation mirrors that of the input-output tables the models are built on. Furthermore, each zone is associated with a set of production, utility, and cost functions. Determining the equilibrium solution requires solving the system of equations comprising these functions for every zone. Thus, granularity is limited by the computational constraints in addition to the data sources.

At the other end of the spectrum, agent-based microsimulation models consider the development of each individual firm. Of particular interest are the so-called firmographic models, which map out the stages and transitions of firm life cycles, including formation, migration, growth, decline, and dissolution. This is analogous to a demographic model mapping out major life events of people, such as birth, death, completion of education, changes to household structure, etc. Firmographic models are by their nature more disaggregate; e.g. individual firms are typically treated as distinct entities with their own evolutionary trajectories (van Wissen, 2000; de Bok, 2009; Moeckel, 2009). Furthermore, some models also allow for firm-specific panel effects (Mostafa, 2017). Most pertinent to the work presented in this dissertation are the employment size growth/contraction models. Given existing firm and local attributes, e.g. sector, size, agglomeration, etc., these model the employment size transitions between periods – either as transition probabilities between discretized bins (Kumar and Kockelman, 2009; Mostafa, 2017) or as autoregressive continuous counts (van Wissen, 2000; de Bok, 2009; Zondag et al., 2015; Ravulaparthy et al., 2017; Zhuge and Shao, 2019).

It should be noted that few, if any, of these firmographic models have been integrated into a larger IUM framework that also models the activity and location decisions of households and workers. In part, this is because agent-based models by and large do not concern themselves with the broader market conditions and equilibrium like spatial input-output models do. As such, tight integration is not necessary for them to operate.

While agent-based models can capture firm heterogeneity by virtue of their microscopic nature, they typically fail to represent the underlying decision processes despite modelling decision-making at the level of the decision-maker. In fact, the production, utility, and cost functions of spatial input-output models are typically more theoretically informed than the transition equations of firmographic models. The latter, for the most part, capture statistical correlations rather than structural relationships.

Finally, there are of course numerous models and frameworks that lie in between the extremes of the spectrum, equilibrium-based spatial input-output models and agent-based microsimulation models, and incorporate aspects from both traditions. For example, IRPUD (Wegener, 2011) and TigrisXL (de Graaff and Zondag, 2013) operate at aggregate zonal levels, similar to most equilibrium models. However, they do not require equilibrium, and instead use employment transition models more akin to those in firmographic models. Broadly, current models of firm employment appear to make a trade-off between theoretical structure and disaggregation. However, these are not necessarily contradictory. Rather, the trade-off is an artifact of the methodological roots, and has been maintained by inadequate micro-data and computational limitations.

Dynamics

Urban development is first and foremost a spatial phenomenon. However, it is crucial to also acknowledge its temporal dimension. Cities evolve with considerable inertia and are unlikely to be in equilibrium at any given moment. Urban infrastructure investments perhaps best illustrate this; construction takes years, lifecycles range from decades to more than a century, and the impact on the surrounding land use is nearly irreversible (Wegener et al., 1986). However, even smaller scale urban processes, such as employment and job search, are riddled with dynamic phenomena that cannot be explained by traditional economic theory (Faggian, 2014). Despite these dynamics, modern modelling efforts have largely neglected the temporal dimension (Simmonds et al., 2013). The majority of IUMs, especially those that have found wide application in practice, are built on static equilibrium assumptions (Jin and Wegener, 2013). These models abstract away time and consequently represent urban development as memory-less, path-independent, and effectively instantaneous. Such discrepancies between our understanding of real-world urban development and modelling assumptions are potentially problematic when we use the models to evaluate responses to land use and transport policies (Simmonds et al., 2013). These limitations vis-à-vis the inadequate representation of time in IUMs are not revelatory; warnings of overreliance on static discrete choice models (Timmermans, 2003) and calls for better modelling of dynamics (Miller, 2018b) have been brought to attention in the literature before.

In particular, the equilibrium assumption in spatial input-output models made them intrinsically static. In other words, once the equilibrium solution has been determined, the state of the model can only change as a result of exogenous perturbations. While relatively simple and theoretically attractive, this behaviour is hardly a good representation of urban development in reality as cities appear to be in constant flux. Some models attempt to circumvent this by introducing *quasi-*

dynamics. Specifically, spatial input-output models typically arrive at the equilibrium state through a converging iterative algorithm. By only running a single iteration of such algorithms in each time step, the models appear to change over time absent exogeneous perturbations. However, the link between the rate of convergence of the algorithm and the rate of urban development is questionable. In other words, the quasi-dynamics do not model urban development over time but rather model convergence over iterations, while undermining the equilibrium assumption on which the models are built.

The relaxation of the equilibrium constraint in agent-based microsimulation models allows for somewhat more organic handling of dynamics with period-to-period transitions. However, none of these transition models capture the underlying frictions or decision processes. Thus, they merely capture averages and fail to account for potential spatial differences or temporal variation, e.g. greater or lesser inertia during times of great upheaval.

2.2 Agglomeration Economies

The benefits of proximity have been a topic of study since at least 1826, when von Thünen presented a theoretical model of agricultural land use in relation to a market located in the city (von Thünen, 1826). His model formalized the trade-off between transportation costs and rent as a function of the distance to the city centre. In turn, he showed that, all else equal, land uses would arrange themselves in concentric rings around the city with less land-intensive, and consequently, higher density uses with high freight costs closest to the centre. Conversely, the outer rings would be occupied by land intensive uses with low freight costs. Although the context of this original inquiry was agricultural, we see the same principles apply in a post-industrial context more than a century later in Alonso's bid-rent model (Alonso, 1964). He formalizes the preferences and willingness to pay for land, i.e. the properties of the "bid rent" curves, of urban firms, households, and agriculture. Using a game theoretic approach, he then derives the resulting equilibrium land uses.

Alonso's and von Thünen's models show how dense build-up, i.e. cities, might arise as a natural consequence of people wanting to take advantage of lower transportation costs and access to goods and labour markets. However, alone they cannot adequately explain the vast unevenness of density and productivity across space[1] (Ottaviano and Thisse, 2004). In addition to lower transport costs, proximity and access also give rise to external benefits (and costs), i.e. economies of agglomeration. Broadly, these are scale and network effects but also crowding and congestion. Marshall (1890) famously wrote about the technological spill-overs between firms locating in close proximity to each other that, "The mysteries of the trade become no mystery; but are as it were in the air". He describes the benefits of specialization among co-locating firms – so-called localization economies – which additionally include labour pooling and reduced transport costs between suppliers and buyers. Examples of such industrial colocation include the semiconductor industry in Silicon Valley, entertainment industry in Los Angeles, financial

[1] This follows from the spatial impossibility theorem (Starrett, 1978) as discussed by Octaviano and Thisse (2004).

services and advertising in New York City, country music production in Nashville (Carlino and Kerr, 2015). The benefits of within-industry co-location stand in contrast to urbanization economies which derive from urban diversity. Inspired by her own experiences in New York City, Jacobs (1961) popularized the idea that the melting pot of urban life is a crucial avenue for cross-fertilization and innovation. Considerable progress has been made since the writings of Marshall and Jacobs, both in terms of theory and empirics. In the following sections I review the theoretical micro-foundations underlying agglomeration externalities. Then, I discuss the challenges identifying these effects in practice and present some empirical evidence from the recent literature.

The micro-foundations of agglomeration economies

In this section I provide an overview of the micro-foundations of agglomeration economies. Although all agglomeration benefits appear as increasing returns to scale, they arise from numerous different mechanisms. An understanding of these causal channels serves as a foundation for interpreting the empirical findings. Furthermore, the different mechanisms can have vastly different policy implications. Hence, understanding and disentangling the micro-foundations are important steps towards effective urban policy-making.

I follow the typology used by Duranton and Puga (2004), and categorize the micro-foundations by sharing, matching, and learning effects. They also provide more rigorous mathematical models showing how each mechanism results in productivity gains. Additionally, I discuss other mechanisms, including social capital, neighbourhood effects, and more, that are not traditionally considered with agglomeration economies. Their omission is primarily a result of their ambiguous returns to scale. Nonetheless, a discussion of their potential impacts is still worthwhile.

Sharing

The sharing of fixed costs is the primary channel through which increasing returns arise from sharing mechanisms. Perhaps the most apparent benefit of agglomeration is the ability to share the fixed costs of indivisible facilities that would otherwise be too costly or infeasible (Buchanan, 1965; Scotchmer, 2002). These include physical infrastructure, e.g. airports, but also institutions and markets. Although not framed as such, farmers in von Thünen’s model cluster around the market in the city due to its indivisible nature (von Thünen, 1826). If the market were divisible, e.g. selling software online, we would not expect agglomeration forming around cities solely from the effects in von Thünen's model.

As the name suggests, sharing mechanisms also underlie sharing economies (Davidson and Infranca, 2016). Although often enabled by information technology, sharing economies fundamentally work by sharing large fixed costs of goods or services with low utilization rates between its users. For example, on-demand mobility services share the cost of cars (and chauffeur services), that otherwise may have been prohibitively expensive, between their users.

Duranton and Puga (2004) make the case for variety and specialization also being results of sharing. For example, a small town might have a single recreational field, perhaps a baseball diamond that is also occasionally used for other activities but lacks dedicated "infrastructure" for other sports. On the other hand, in a large city you will typically find a variety of venues, specialized for each discipline, e.g. basketball court, ice hockey rink, football field, etc. In this example, a critical mass of demand is necessary to sustain each facility due to the indivisible fixed cost. Thus, the presence of variety arises from the ability to share the fixed costs of indivisible facilities. Of course, this is not unique to athletics facilities; agglomeration economies enable firms to produce a greater variety of goods and services. Greater aggregate demand for intermediate inputs, i.e. input sharing, also lowers the average production costs of suppliers. Similarly, worker training in specific skills is a fixed cost with respect to production. Thus, following the same principles, sharing in agglomeration economies enable workers to train in a broader variety of more specialized skills (Rosen, 1983).

A secondary, or at least less apparent, channel of increasing returns through sharing is by sharing risks. Specifically, a thicker labour market, i.e. one with both more firms and workers, is better able to absorb idiosyncratic shocks and fluctuations at the level of individual workers or firms (Krugman, 1991). Marshall (1980) described this effect as a "constant market for skill". For example, a firm experiencing a positive productivity shock will see the cost of labour increase less in a thick labour market compared to a thin one. Similarly, a worker who has been laid off due to a negative idiosyncratic shock at the previous employer is more likely to find employment the more other employers exist in the market.

Matching

Sharing mechanisms enable cities to sustain variety. While the benefits of variety may be intuitive to some, matching mechanisms formalize these. If preferences are heterogeneous, e.g. individuals might prefer certain goods or firms might prefer certain skillsets, having more different options increases the chances of finding matches that fit the specific taste preferences, thus increasing the expected match quality (Helsley and Strange, 1990). This mechanism is also captured by consumer surplus measures in choice models, so-called logsums (Ben-Akiva and Lerman, 1985). In the labour market context, this most obviously applies to the quality of matches between employers and worker skills; variety in the labour pool allows firms to hire workers that possess the exact skills required for the task. In turn, this increases firm production and worker wages. It similarly applies to matches between suppliers and buyers of intermediate inputs and matches final products and consumers.

Agglomeration not only improves the quality of matches but also increases the chances of matches occurring, reducing search times and costs. This mechanism was first explored in the frictional search literature (Mortensen, 1986; Petrongolo and Pissarides, 2001). The intuition here is straight-forward; having access to more firms and workers, agents can explore more options simultaneously, thus increasing the rate of matching (Coles and Smith, 1998). As a consequence of lower search friction, we also expect rational agents to raise their reservation match quality, i.e. the threshold for quality of a match that an agent would accept, thus further raising the expected match quality.

Finally, the availability of outside options can mitigate hold-up problems in case of incomplete contracts. For example, consider a location with a single employer and workers who are either skilled or unskilled. Skilled workers are more productive but becoming skilled requires training, which is an upfront cost incurred by the worker. However, if contracts are incomplete, workers can be held hostage by the firm’s monopsony power, and are thus discouraged to invest in training, despite it being socially optimal. Introducing other firms that can make use of the workers’ skills removes the firms’ monopsony power and the hold-up problem. The same line of reasoning can be applied to inter-firm relations between suppliers and buyers; suppliers are discouraged from investing in R&D if the buyers of their products hold monopsony power. This matching mechanism contributes to the innovative advantages of agglomeration economies that I elaborate on in the subsequent section along with learning mechanisms.

Learning

Learning and knowledge spill-overs are very compelling explanations for the increasing returns to agglomeration and have been central to both Marshall's (1890) and Jacobs' (1961) works. However, whereas the assumptions underlying sharing and matching mechanism, e.g. existence of fixed costs and heterogeneous preference, are intuitive and require little additional explanation, the assumptions underlying learning mechanisms are often somewhat more nebulous and the mechanisms themselves remain somewhat of a black box (Carlino and Kerr, 2015). Broadly, they appeal to the notion that learning, i.e. the transfer of information between people and/or firms, is an inter-personal activity best facilitated by proximity. For example, Helsley (1990) treats knowledge as a by-product of firms' production that serves as an intermediate input for other firms. It diffuses spatially through contacts between firms, which in turn decay with distance. However, he does not further specify the mechanism of knowledge transfer. Similarly, Glaeser (1999) assumes, in a theoretical model of learning in cities, that every encounter between two people of different skill levels carries with it a possibility of learning. Encounters are more frequent in denser cities, per the discussion above on matching rates. He shows that the logical consequence of these assumptions is that the average skill level is higher in cities and that there is a higher variance of skill levels in cities. Exploring innovation in firm life cycles, Duranton and Puga (2001) assume that new firms produce inefficiently until they discover their optimal production process. This learning is less costly in diverse clusters (urbanization economies) due to the availability of variety in both worker skills and intermediate inputs. Once firms have determined their optimal production process they move to specialized industrial clusters (localization economies). As illustrated by these examples, in the urban economics literature, increasing returns from knowledge spill-overs are often micro-founded through sharing and matching mechanisms.

Social mechanisms

A variety of concepts and mechanisms from the social sciences, e.g. social capital, neighbourhood effects, intuitively relate to agglomeration economies. These examine the social determinants of choices and outcomes, i.e. the interdependencies between the behaviours amongst individuals in a group. There are numerous reasons for such interdependences to exist. For example, individuals might have an intrinsic psychological desire to imitate the behaviours of other group members, or they might conform to group norms, consciously or not, to build social capital. Alternatively, the actions of one group member might alter the costs or constraints of others in the group, or later decision-makers might be able to observe and learn from the outcomes of earlier decisions, so-called social learning (Durlauf, 2004). These social mechanisms do not intrinsically exhibit increasing returns to scale but can in some cases amplify or be amplified by agglomeration. For example, if the average skill levels are higher in cities as in Glaeser's (1999) model, neighbourhood effects could amplify these differences. Conversely, if social learning takes place, proximity could strengthen the informational signals. Unfortunately, to my knowledge, these social mechanisms have not been formally linked to agglomeration economies in a rigorous way. Theoretical models of neighbourhood effects, including role model and peer group effects, focus on exploring equilibria and phase transitions, rather than returns to scale (Durlauf, 2004), and empirical work brings with it a host of econometric challenges, including reflection problems, sorting problems, and spatially correlated shocks (Topa and Zenou, 2015). Conversely, the literature on social learning presents convincing micro-founded models of learning between agents; see e.g. Sobel (2000) for a survey. However, these models do not exhibit increasing returns to scale (Duranton and Puga, 2004). Finally, there is a vast literature on social capital. However, its interaction with agglomeration economies is unclear. For example, we expect social capital to decline with residential mobility but increase with proximity to people (Glaeser et al., 2002) – both attributes associated with dense cities. These issues are exacerbated by the nebulous definition and multitude of operationalizations of the concept.

The empirics of agglomeration economies

The micro-foundations provide plausible explanations for agglomeration economies, but they do not inform us concretely about the magnitude or scale of the effects. For that, we must turn to the real world. Evidence of the externalities of agglomeration is not difficult to come by. Bettencourt et al. (2007) famously documented urban increasing returns to scale for innovation, productivity, and wages using a dataset spanning more than several hundred cities across the United States, China, and Europe. However, these findings did not, nor did they attempt to, shed light on the micro-foundations of agglomeration economies or how the mechanisms work at a more local scale. Rather, they establish the robustness of the increasing returns to scale at an aggregate level. Much of the recent empirical literature on agglomeration economies attempts to address the identification challenges and endogeneity issues, as described further below, and also try to disentangle the different underlying mechanism discussed in the previous section. These efforts, however, very much remain works in progress. In the following section, I discuss the potential issues with endogeneity in estimating agglomeration effects, measures of agglomeration, and finally existing empirical evidence of the magnitude and scale of agglomeration economies.

Endogeneity

Estimation of agglomeration effects requires careful consideration of potential endogeneity issues at both the firm and local levels (Combes & Gobillon, 2015). Endogeneity at the firm level, or individual level when examining individual outcomes, arises when unobserved agent effects are correlated with location attributes. This typically occurs when agents sort spatially according to characteristics not controlled for by the model, e.g. unobserved advantages in production technology for firms or unobserved abilities for individuals. Similarly, endogeneity at the local level arises when a local characteristic is correlated with unobserved local effects. This can typically happen if an aggregate variable that affects both the local characteristics and the outcomes is missing. For example, an airport might attract firms to the area and increase the productivity of the firms in the area, thus simultaneously increasing agglomeration and the measured outcome. In this case, the estimated agglomeration effect would be upwards biased. Alternatively, reverse causality also causes endogeneity issues at the local level. In other words, higher local productivity might cause workers and firms to move to the area, thus the causal interpretation of agglomeration economies would be wrong.

With panel and micro-data becoming increasingly available, endogeneity is most commonly addressed by introducing time-invariant agent or location-specific effects (Combes & Gobillon, 2015). These control for all the unobserved characteristics that remain constant over time, e.g. workers' unobserved intrinsic ability or local endowments that have not been controlled for. However, this strategy is hardly a panacea and may fail to address endogeneity issues for several reasons: (1) unobserved characteristics that change over time can still cause endogeneity; (2) it also does not address endogeneity resulting from reverse causality; and (3) controlling for time-invariant effects means that the estimation of the coefficients of interest relies on changes in agglomeration and outcomes over time. If these are small, the estimated effects are susceptible to attenuation bias from measurement errors (Angrist & Pischke, 2008). Alternative approaches to handling endogeneity, particularly at the local level, include instrumentation with geological variables, e.g. bedrock depth, landslide hazard, etc. (Rosenthal & Strange, 2008), or historical variables, e.g. historical presence of a deep water port (Ciccone & Hall, 1993), or by exploiting natural experiments, such as the division and subsequent reunification of Berlin around the Cold War (Ahlfeldt et al., 2015). Finally, structural approaches can and have been used to handle endogeneity issues. For example, Baum-Snow and Pavan (2012) formulate a dynamic model inspired by the matching models in the job search literature. It jointly models workers' wages and location decisions, thus explicitly accounting for spatial sorting. Structural approaches offer a potentially extremely useful tool for addressing endogeneity issues. However, the theory-based estimation relies heavily on both structural and parametric assumptions. Or, as expressed by Rust (1994, p.3130), *"since structural models can be falsified but never proven true, their predictions should always be treated as tentative and subject to continual verification."* For illustrative examples of structural estimation in urban economics, see Holmes and Sieg (2015).

Measurement

Agglomeration economies refer to the external economic benefits arising from the agglomeration of people and economic activity. The urban economics literature typically measures these using employment, population, and production. Unfortunately, they are often highly correlated and may not allow for identification of separate parameters (Combes & Gobillon, 2015). Between the three, employment is usually the preferred explanatory variable because it more directly represents economic activity compared to population and suffers less from endogeneity issues than production. Where possible, the use of density instead of magnitude is preferred as it is more robust vis-à-vis modifiable areal unit problems (Ciccone & Hall, 1993). Related to modifiable areal unit problems, these elementary measures share a fundamental limitation. Namely, they imply that agglomeration effects are in full force within a zone but do not cross zonal borders, which are sometimes defined with little consideration for the spatial extent of economic activity. To incorporate the attenuating effect of distance more explicitly, some authors use Harris's (1954) market potential variable.

$$A_i = \sum_j \frac{d_j}{c_{ij}} \tag{2.1}$$

Here, c_{ij} is the distance between the zone of interest i and other zones j and d_j is the size of the market in of the other zones. This so-called gravity formulation was also adopted by transportation and urban planners to represent *accessibility*, or as Hansen (1959, p.73) defined it, *"the potential of opportunities for interaction"*. Gravity-based accessibility has also been formulated with other spatial decay functions. In particular, the negative exponential formulation is especially convenient, with the cost function evaluating to 1 for $c_{ij} = 0$.

$$A_i = \sum_j d_j \exp\left(-c_{ij}\right) \tag{2.2}$$

An important difference between applications of gravity-based measures in urban economics compared to transportation planning is the definition of the variable measuring spatial friction c_{ij}. Whereas, applications in urban economics, with few exceptions, consider distances, the transportation planning literature typically uses travel times or generalized costs on the real road/transportation network. From the perspective of transportation planning, space is not homogenous. The same physical distances can have entirely different implications for access, and by extension agglomeration economies, depending on the transportation network, built environment, and socio-demographics. The overlap with the transportation literature is hardly surprising. After all, *"all of the benefits of cities come ultimately from reduced transport costs for goods, people, and ideas"* (Glaeser, 1998, p. 140).

Beyond proximity, density, and accessibility, a number of other measures have been useful for characterizing local economy activity – in particular for distinguishing between location and urbanization economies. For example, Henderson et al. (1995) measures specialization at location j as the fraction of workers employed in a given sector k.

$$spec_j^k = \frac{emp_j^k}{emp_j} \tag{2.3}$$

Relatedly, the most common measure of industrial diversity is the inverse of the Herfindahl index (Duranton and Puga, 2001).

$$div_j^k = \left[\sum_k \left(spec_j^k\right)^2\right]^{-1} \tag{2.4}$$

An undesirable feature of this measure is that it is heavily dependent on how the researcher chooses to classify sectors k with a finer breakdown yielding more diversity. Several alternative measures and proxy measures have been used to capture urbanization economies. For example, the Krugman Index (Krugman, 1991) measures the difference between local and region-wide distributions of employment specialization across sectors. Alternatively, simpler measures, such as average firm size (Glaeser et al., 1992) or total local employment size (Rosenthal and Strange, 2004), have also been used.

Evidence

While there is ample evidence of the existence of agglomeration economies and the theoretical micro-foundations are well-established, relating the empirical evidence to the theory remains a challenge. The outcomes, i.e. increased productivity, innovation, etc., are observationally very similar regardless of the underlying mechanism (Rosenthal and Strange, 2004). In addition to comprising a combination of positive agglomeration mechanism, the observed effect also captures negative effects, such as congestion and crowding. Thus, disentangling and identifying which (combination of) mechanisms are at play is often not possible. Contributing to the issue is the fact that the measures of agglomeration tend to be much simpler than theory demands and are rarely able to capture the many nuances of the micro-foundations.

Despite these challenges, many studies have attempted to quantify the magnitude of agglomeration economies. Combes and Gobillon (2015) provide a comprehensive survey of these efforts. Unavoidably, findings vary depending contexts of the studies and the specifications of the models, e.g. the sector of study, time periods, country contexts, control variables, and whether or not individual unobserved heterogeneity and reverse causality were accounted for. However, across the literature, typical elasticity values range between 0.04 and 0.07 (Melo et al., 2009; Combes and Gobillon, 2015). Generally, firms' total factor productivity (TFP) is more elastic with respect to agglomeration variables than worker wages. Studies with fewer control variables tend to find larger magnitudes for their estimates of elasticity. Perhaps surprisingly, correcting for aggregate local level endogeneity has little effect on the estimates and the effect is not consistent in direction. On the other hand, correcting for individual level endogeneity yields smaller magnitudes, down to around 0.02.

Investigations by sector remain rare. The primary reasons for this gap are inadequate granularity in the available datasets and the challenge of finding appropriate sector-specific instruments to correct for endogeneity issues. Furthermore, conclusions from studies that distinguish between sectors are mixed, with some finding stronger agglomeration effects for manufacturing (Melo et al., 2009) and others the opposite (Gould, 2007; Matano and Naticchioni 2012). This is likely a consequence of the differences in how agglomeration is measured and further emphasizes the importance of being able to disentangle the underlying mechanisms.

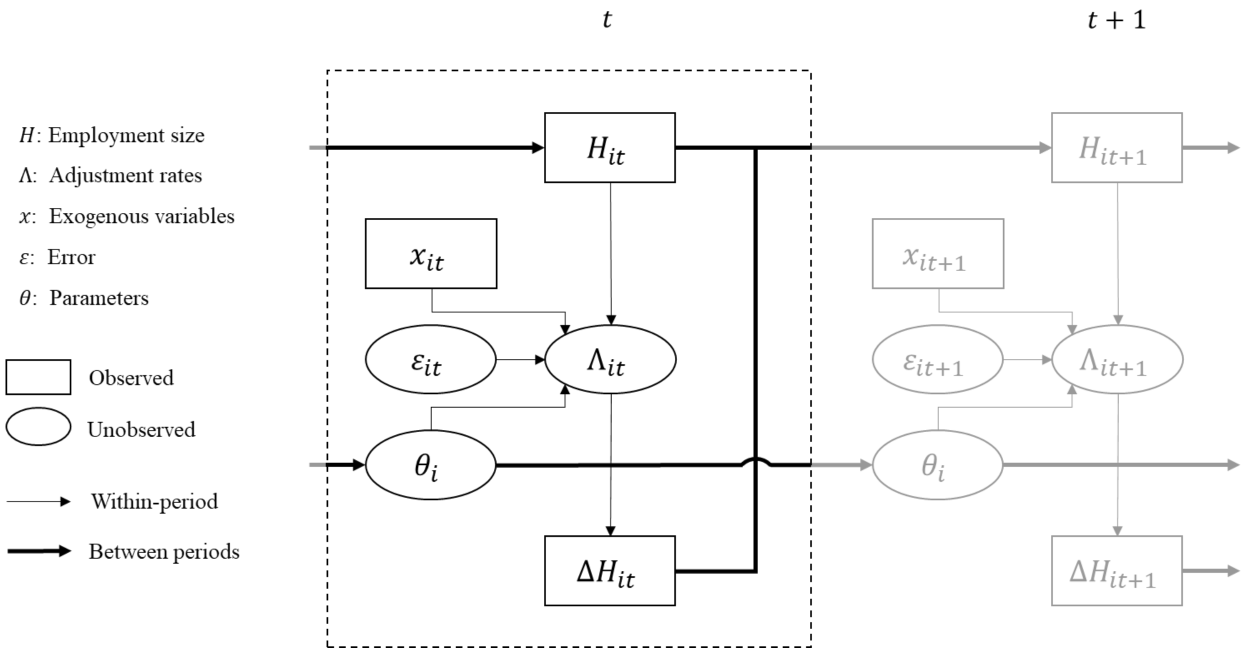
Despite fundamentally dealing with spatial proximity, the theory on agglomeration economies is not particularly informative with respect to spatial scale. Whether indivisible facilities can be shared across metres or miles is another question left for the empirics to answer. Notably, many of the studies discussed thus far have used zones defined by jurisdictional boundaries to measure space. This was for a long time the norm in the literature. As more granular data have become available, studying the spatial extent of agglomeration economies has become more common (Rosenthal and Strange, 2004). The empirical evidence suggests that the agglomeration effects decay rapidly with distance. Di Addario and Patacchini (2008) find that the population size of the local labour market positively correlates with worker wages with the effects being strongest within 4 kilometres and no longer significant beyond 12 kilometres. Similarly, Rosenthal and Strange (2008) found that the wage premium resulting from the number of workers within a 5-mile radius was 4-5 times that of the number of workers between 5 and 25 miles away. Holl (2012) uses a Harris-type measure and finds that population "market potential" increases regional wages in Spain. Such findings, of course, vary considerably by sector and inferred underlying mechanism. For example, Arzaghi and Henderson (2008) show that for the advertising industry in Manhattan, agglomeration benefits had completely vanished beyond merely 750 metres. Ellison et al. (2010) use industry colocation patterns to infer which mechanisms, input/output sharing, labour pooling, or knowledge spillovers, most strongly cause agglomeration. They find that industries relying on knowledge spillovers tend to be the most tightly clustered.

3 THE MODEL

I present a dynamic model of employment growth and decline of individual firms. Profit-maximizing businesses adjust their employment size in response to changes in their environment (physical, economic, or otherwise) that affect their profitability. However, hiring and firing employees is costly for a business, both directly, e.g. in the form of hiring bonuses or severance payments, and indirectly, e.g. through productive hours allocated to searching for an appropriate skills match, training, or re-organization of tasks and responsibilities. For example, surveys of HR professionals have found that it costs more than four thousand dollars, on average, to hire an employee (SHRM, 2016). It then takes on average eight months for the new employee to become fully productive (Allied, 2012). During this time, the firm still has to pay the employees full salary and potentially additional costs to train the employee. These figures naturally vary considerably depending on the nature of the job. Some estimates for the total costs of hiring and onboarding go up to 50%, 120%, and 400% of the employee's annual salary for entry-level, mid-level, and high-level or highly specialized employees, respectively (Borysenko, 2014). Given these non-negligible adjustment costs, and in the presence of discounted future values, adjustment decisions involve an intertemporal trade-off between adjustment costs and production profits. In turn, firms' employment size adjustments will be partial and/or gradual rather than instantaneous. Take for example a myopic firm, i.e. one that discounts future values heavily. All else equal, they would choose to make smaller adjustments. For them, the upfront costs of adjusting may well outweigh the benefits of increased profit in the short term. On the other hand, for a firm that does not discount the future heavily, the cost of rapid upfront adjustments can be offset by less foregone profit in subsequent periods. Adjustment costs that are non-linear with respect to the rate of adjustment is another potential cause of lagged responses. For example, attempting to fill numerous vacancies quickly when the labour supply is limited might require an expanded advertising effort, larger onboarding incentives, and potentially result in poorer skills matches. Such non-linear costs can further discourage rapid adjustments.

To model this behaviour, I formulate a dynamic programming Markov model in which firms' employment sizes are state variables and employment size adjustments serve as state transitions. Figure 3-1 below shows the model structure.

Figure 3-1: Overview of the dynamic programming Markov model structure



For each firm i and time period t, we have observations comprising $\{H_{it}, x_{it}, \Delta H_{it}\}$, where H_{it} is the firm's employment size at the beginning of the period, x_{it} are the relevant exogenous variables (including location-specific attributes such as agglomeration), and ΔH_{it} is the outcome variable, i.e. the total adjustment by the firm over the duration of the period. As modellers, we do not observe the time-invariant firm-specific parameters θ_i or the independent and identically distributed (IID) error term ε_{it}. Similarly, we do not observe the inner workings of the adjustment process, i.e. we do not know the exactly when during the period t each employee was hired or fired, much less the underlying adjustment rates Λ_{it}. To determine these adjustment rates, we assume that firms are rational, forward-looking decision-makers that choose Λ_{it} to maximize the net present value of all future profits. I model the hiring and firing as events in stochastic processes. Consequently, the firm is faced with a Markov decision process (MDP), whose solution, i.e. optimal policy, is Λ_{it}. As shown in Figure 3-1, the solution is a function of the employment size at the beginning of the period H_{it}, exogeneous variables x_{it}, a random term ε_{it}, and the firm-specific parameters θ_i.

$$\Lambda_{it}(H_{it}, x_{it}, \varepsilon_{it}, \theta_i) \tag{3.1}$$

Since adjustments are stochastic, the final outcome, i.e. the total adjustment ΔH_{it}, is also stochastic and its probability function is conditional on the adjustment rates. This serves as the model's likelihood function.

$$P(\Delta H|\Lambda_{it}) \tag{3.2}$$

Finally, the initial employment in the subsequent period $t + 1$ is the sum of the current employment size and the adjustment. Then the process repeats itself for each period.

$$H_{it+1} = H_{it} + \Delta H_{it} \tag{3.3}$$

The model structure is summarized by these three components: the firm's MDP and its solution (3.1), the likelihood function (3.2), and the transition to the subsequent period (3.3). In the following sections, I expand on these. Specifically, in Section 3.1 I present $\Lambda_{it}(H_{it}, x_{it}, \varepsilon_{it}, \theta_i)$ in full by showing the dynamic programming solution to the firm's MDP; in Section 3.2 I derive the conditional transition probability function $P(\Delta H_{it}|\Lambda_{it})$ and formalize the likelihood function. To estimate the model, I use a Bayesian Markov chain Monte Carlo (MCMC) algorithm. A Bayesian hierarchical model offers a convenient way to estimate firm-specific parameters, which capture serial correlation in firm's choices. In turn, this makes the IID assumption for ε_{it} more believable. I present this estimation procedure in Section 3.3.

The model presented here is closely related to the dynamic programming discrete choice structural models encountered in the economics literature with a few important differences that reduce the computational burden of the estimation procedure. I discuss how this model relates to the broader class of dynamic programming discrete choice structural models in 3.4, along with several miscellaneous topics, including how the model can fit into a larger IUM framework and its potential usefulness for other applications.

3.1 Firm Behaviour

Firm's Perspective

Our objective in this section is to derive the firm's decision rule for the adjustment rates Λ_t. For simplicity of notation, I drop the firm-specific subscript i in this section as the firm's decision rule does not involve other firms in the dataset. It is important to make a distinction between the researcher's perspective and the firm's perspective. Previously, Figure 3-1 presented the employment size adjustment process from the researcher's point-of-view, observing the time series of the exogeneous variables $(x_t, x_{t+1}, x_{t+2}, \dots)$ and inferring the distribution of the error terms based on the entire dataset. By contrast, the firm cannot access the whole dataset retrospectively but does observe the firm and period-specific terms that are stochastic to the researcher. At time t, the firm observes both the exogenous variables x_t and the idiosyncratic error ε_t as well as the time-invariant parameters θ specific to them. However, the firm does not know the subsequent values of x and ε, which to them are in the future. This is reflected by the firm's MDP, whose solution is only a function of the current values of the variables x and ε. That is to say, the periods $(t, t+1, t+2, \dots)$ refer to real time periods created as a product of the data collection process and observed by the researcher. For the period t, the firm observes the period-specific variables x_t and ε_t and solves an MDP to determine the optimal adjustment rates specific to that period Λ_t. In the subsequent period $t+1$, the firm will be faced with a new MDP with updated information x_{t+1} and ε_{t+1}, which it will solve to determine new optimal adjustment rates Λ_{t+1}. Put differently, each instance of the firm's MDP effectively exists within a single period. It is a means of determining the optimal adjustment rates Λ for that period. Importantly, the time dimension within the MDP is different from the *real* periods observed by the researcher $(t, t+1, t+2, \dots)$. Consequently, moving into the future within an MDP does not increment t.

The Markov Decision Process

Formally, let us consider an infinite time horizon where firms discount future cash flows at rate δ. The firm's problem is a continuous-time MDP. In other words, transitions between states can occur at any time and do not need to adhere to discrete time intervals of fixed duration as in traditional discrete-time MDPs. Carrying out the derivation with fixed time periods ultimately yields the same optimal decision rule but requires a few additional steps. Appendix A presents this alternative derivation.

An MDP can be described by the four components: the state space, the action space, the transition rate function, and the reward function. These are summarized in Table 3-1 below.

Table 3-1: Overview of the firm's Markov decision process

Component	**Description**	**Notation**
State space	Employment size is the only state variable; thus the state space is the set of possible employment sizes	$(H_t, H_t \pm 1, H_t \pm 2, \ldots, H_t^*)$
Action space	At each state H, the firm can choose to hire or fire employees at rate λ_H	Hire or fire employees at rate $\lambda_H \; \forall \; H$
Transition rate function	Given state H and adjustment rate λ_H, the time until the transition $H \rightarrow H \pm 1$ is an exponential-distributed random variable z_H	$z_H \sim Exponential(\lambda_H)$
Reward function	The reward is the firm's net profit during the time spent at state H	$R(H, \lambda_H)$

The states of the MDP are defined by the firm's employment size. If profit is globally concave with respect to employment size, the state space comprises only the interval between the initial employment size H_t and the optimal employment size H_t^* (inclusive) rather than all natural numbers. In other words, if the profit function is single-peaked, a profit-maximizing firm never has reason to move farther away from the optimal employment size.

The $\pm$ symbol is a plus (minus) if the optimal employment size H_t^* is greater (smaller) than the initial employment size H_t, in which case the firm will hire (fire) employees. It will choose the rate λ_H that maximizes the net present value of all expected future rewards. Adjustment rates λ_H are defined to be non-negative and represent the hiring and firing rates when firms are hiring and the firing, respectively. The set of these adjustment rates, each corresponding to a state in the state space, makes up the solution to the MDP at time t. Thus, rather than representing a single adjustment rate, Λ_t comprises a vector of rates.

$$\Lambda_t = \left(\lambda_{H_t}, \lambda_{H_t \pm 1}, \lambda_{H_t \pm 2}, \dots, \lambda_{H_t^*}\right) \tag{3.4}$$

The vector is denoted by the Λ (uppercase lambda), while individual adjustment rates are denoted by λ (lowercase lambdas). Although, I have referred to Λ_t as a single entity thus far, the fact that it comprises a vector of rates reflects that adjustment rates do not necessarily remain constant over the duration of a period t. From the firm's perspective, it is generally desirable to change the adjustment rate each time they hire or fire an employee. For example, as a firm approaches its optimal employment size, we expect adjustment rates to slow down and eventually reach zero, as the firm's incentive to make further adjustments diminishes. Altogether, Λ_t describes an employment size adjustment trajectory over time between the employment size H_t and the optimal employment size at time H_t^*.

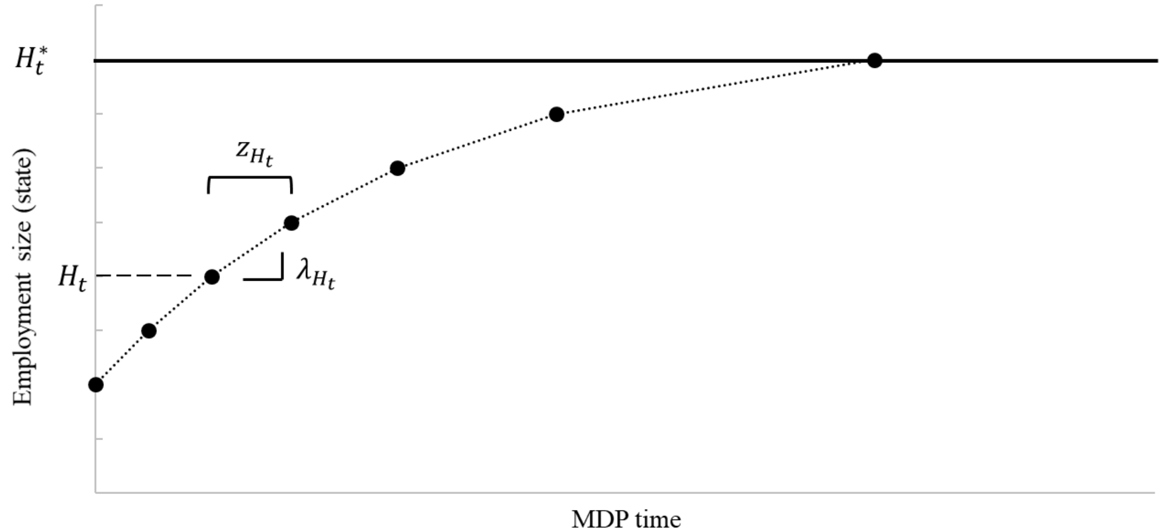
State transitions increment (if hiring) or decrement (if firing) the employment size, i.e. $H \rightarrow H \pm 1$. Let z_H be a random variable representing the time between subsequent adjustment events, i.e. the time during which the employment size is H. It is exponential-distributed with adjustment rate parameter λ_H, chosen by the firm. The parameter represents firms' effort, eagerness, and urgency in making adjustments.

$$\begin{gathered} z_H \sim Exponential(\lambda_H) \\ f_H(z_H) = \lambda_H e^{-\lambda_H z_H} \quad , \quad F_H(z_H) = 1 - e^{-\lambda_H z_H} \quad , \quad \tilde{F}_H(z_H) = e^{-\lambda_H z_H} \end{gathered} \tag{3.5}$$

$f_H(z_H)$, $F_H(z_H)$, and $\tilde{F}_H(z_H)$ are the PDF, CDF, and survival function of z_H, respectively. Note that the model only captures net adjustment, rather than hiring and firing separately.

Figure 3-2 presents an example trajectory graphically. The slope represents the adjustment rate and the horizontal distance between events represents the (expected) time between events. If profit is concave with respect to the employment size, i.e. the incentive to make further adjustments diminishes with each subsequent adjustment as the employment size approaches H_t^*, the rates in Λ_t are also diminishing, eventually reaching 0 for the rate corresponding to H_t^*.

Figure 3-2: Graphical representation of sample trajectory Λ_t



Finally, the reward function captures the firm's net profit during z_H, i.e. the time spent at state H, given the chosen adjustment rate (action). It is the net of production profits and adjustments costs. I describe the formulation of these in the following section.

Production profits π, adjustment costs α, and rewards R

Let $\pi(H; x_t, \varepsilon_t)$ denote a firm's annual profit associated with production during period t. Note that x_t and ε_t are subscripted with t, indicating that they remain constant in the firm's MDP for period t, whereas H varies.

$$\pi(H) = \pi(H; x_t, \varepsilon_t) = \bar{\pi}(H; x_t) + H\varepsilon_t \tag{3.6}$$

The profit function comprises a systematic component $\bar{\pi}$ and a random component. The systematic component is a function of the employment size H and exogenous variables x_t. I leave the complete specification of $\bar{\pi}$ for later but note for now that it is defined to be concave with respect to H, i.e. the profit exhibits diminishing marginal returns to employment size, and that a global maximum exists. This ensures that employment sizes do not tend to infinity. The random component is the product between the employment size H and a zero-mean error term ε_t, which represents heterogeneity in average worker productivity that is known to the firm but unobserved by the researcher. The firm's optimal employment size H_t^* given the values of x_t and ε_t at time t is the employment size which maximizes π. This profit function captures production revenues and costs but not the costs associated with hiring and firing. Absent these frictional effects, firms would choose an infinitely high adjustment rate in order to instantaneously adjust to the optimal employment size H_t^*.

Let $\alpha(\lambda_H)$ represent the annual costs associated with employment size adjustments. These costs are a function of the adjustment rate λ_H.

$$\alpha(\lambda_H) = \begin{cases} \mu_0^+ + \mu_1^+\lambda_H + \mu_2^+\lambda_H^2 & if\ hiring \\ \mu_0^- + \mu_1^-\lambda_H + \mu_2^-\lambda_H^2 & if\ firing \\ 0 & otherwise \end{cases}, \quad \mu_0^+, \mu_1^+, \mu_2^+, \mu_0^-, \mu_1^-, \mu_2^- > 0 \qquad (3.7)$$

The $+$ and $-$ superscripts are used to distinguish between parameters associated with hiring and firing, respectively, as these may well be different. Firms only incur adjustment costs for non-zero adjustment rates. The parameters μ_0, μ_1, μ_2, for either hiring or firing, characterize the adjustment costs. Specifically, μ_0 captures fixed costs, μ_1 captures linear costs, and μ_2 captures superlinear costs. The presence of fixed and linear costs is relatively intuitive. For example, placing a job advert is a fixed cost of hiring; once the advert has been posted its cost does not change regardless of how many employees the firm subsequently hires. Hiring bonuses are an example of linear hiring costs; each new employee will expect to be paid a similar hiring bonus. On the other hand, the intuition behind the superlinear adjustment costs, and the quadratic specification in particular, are perhaps less obvious. Numerous factors can contribute to costs rising superlinearly, e.g. decreasing quality of skills matches for each subsequent hire, increasing compensation demands, or increasingly complex reorganization of tasks following downsizing, etc. The quadratic specification is a relatively simple and a flexible way to capture these superlinear effects. There is also precedent for the use of a quadratic cost formulation in adjustment models. Specifically, the standard partial adjustment model, whereby adjustments in each period are proportional to the distance between the current and optimal values, arises as a result of quadratic adjustment costs (Kennan, 1979). These three cost specifications result in different modes of responses vis-à-vis firm's employment adjustment, as will become apparent in the following derivation.

The firm's net annual profit is the profit associated with production π less the adjustment costs α. The reward function $R(H, \lambda_H)$ is the present value of the net profit until the next adjustment, i.e. over time z_H.

$$\begin{aligned} R(H, \lambda_H) &= \int_0^{z_H} e^{-\delta s}\big(\pi(H) - \alpha(\lambda_H)\big)\, ds \\ &= \frac{\pi(H) - \alpha(\lambda_H)}{\delta}\left(1 - e^{-\delta z_H}\right) \end{aligned} \tag{3.8}$$

Having the discount rate δ in the denominator yields the present value of a perpetuity, thus the fraction in equation (3.8) represents the value of a perpetuity with the firm's net annual profit as the periodic cash flow. However, the reward function should only capture the value of this perpetuity up until time z_H, which is accounted for by the term in the brackets.

Intertemporal decision-making and the value function *V*

Given that adjustments are non-instantaneous and costly, and that costs increase with the rate of adjustment, a firm with a suboptimal employment size must consider the inter-temporal trade-off between production profit and adjustment costs. Specifically, if the firm chooses to adjust rapidly, it incurs large upfront adjustment costs but potentially makes up for it by increasing profits sooner. Conversely, slower adjustment lowers the upfront adjustment costs, but profits remain lower for longer. Optimal firm behaviour requires finding the trajectory between the initial employment size H_t and the optimal employment size H_t^* that maximizes the net present value of all future rewards – we call this the value function. We can now write the value function $V(H, \lambda_H)$, as a recursive Bellman equation (3.9).

$$\begin{aligned} V(H, \lambda_H) &= E\big[R(H, \lambda_H) + e^{-\delta z_H}\, V(H \pm 1, \lambda_{H\pm 1})\big] \\ &= \int_0^{\infty} f_H(z_H)\, \big[R(H, \lambda_H) + e^{-\delta z_H}\, V(H \pm 1, \lambda_{H\pm 1})\big]\, dz_H \end{aligned} \tag{3.9}$$

Since z_H is a random variable, we must consider the expectation over the distribution of z_H. The sign of the $\pm$ depends on the direction of adjustment with $+$ for hiring and $-$ for firing. Substituting in the PDF of z_H from equation (3.5), evaluating the integral, and rearranging yields equation (3.10). Due to the discontinuity in adjustment costs at 0, caused by the fixed costs μ_0 applying only to non-zero adjustment rates, we must consider the 0-solution separately.

$$V(H,\lambda_H) = \begin{cases} \dfrac{\pi(H)}{\delta+\lambda_H} - \dfrac{\mu_0+\mu_1\lambda_H+\mu_2\lambda_H^2}{\delta+\lambda_H} + \dfrac{\lambda_H\, V(H\pm 1,\lambda_{H\pm 1})}{\delta+\lambda_H} & if\ \lambda_H > 0 \\ \dfrac{\pi(H)}{\delta} & if\ \lambda_H = 0 \end{cases} \tag{3.10}$$

The firm chooses the λ_H that maximizes $V(H,\lambda_H)$. We determine the optimal adjustment rate for $\lambda_H > 0$ by taking first order condition. This yields a quadratic equation, whose positive root is the non-zero solution to the firm's MDP. The negative root yields a negative adjustment rate and is not permissible. Again, due to fixed costs, firms may decide against making any adjustments despite their current employment size being sub-optimal.

$$\lambda_H^* = \underset{\lambda_H}{\operatorname{argmax}}\, V(H,\lambda_H)$$

$$= \begin{cases} 0 & if\ V(H,0) \geq V(H,\lambda_H)\ \forall\lambda_H \\ -\delta + \sqrt{\delta^2 + \dfrac{\delta\, V(H\pm 1,\lambda_{H\pm 1}) - \pi(H) + \mu_0 - \delta\mu_1}{\mu_2}} & otherwise \end{cases} \tag{3.11}$$

Some quick intuition to make sense of the optimal adjustment rate: if hiring an additional employee increases the value function considerably, i.e. the difference $\delta V(H\pm 1,\lambda_{H\pm 1}) - \pi(H)$ is large, firms are encouraged to make adjustments more urgently, and vice versa. The adjustment costs affect firms' responses in two distinct ways: a hurdle effect and a retarding effect. The hurdle effect establishes a minimum improvement – a so-called hurdle that must be cleared – for adjustments to be worthwhile. This is reflected by the condition $V(H,0) \geq V(H,\lambda_H)\ \forall\lambda_H$ in equation (3.11). The retarding effect changes firms' adjustment rate, given that it is non-zero. This is reflected by parameters μ_0, μ_1, μ_2 for non-zero optimal adjustment rates in equation (3.11). The fixed costs, as captured by μ_0, encourage bundling of adjustments. In other words, they contribute to the hurdle effect, but if the hurdle is cleared, they encourage higher adjustment rates, contributing negatively to the retarding effect. In practice, an example of a fixed adjustment cost is the placing of a job advert. The cost is incurred only if the firm has an intention to hire. Once the advert has been posted, the firm does not incur any further costs associated with the advert regardless of how many new workers are actually hired. Linear adjustment costs, as captured by μ_1, include, for example, hiring bonuses, training costs, etc. and make adjustments more expensive overall. In turn, this contributes to both the hurdle and retarding effects provided firms discount future cash flows. Finally, the quadratic costs, captured by μ_2, also contribute to both the hurdle and the retarding effects but are chiefly responsible for the latter. Numerous factors could contribute to adjustment costs rising superlinearly, e.g. decreasing quality of skills matches for each subsequent hire, increasing compensation demands, or increasingly complex reorganization of tasks.

Dynamic programming solution

Armed with equations (3.10) and (3.11), we can now solve the MDP recursively using dynamic programming. Starting from the end – when the employment size is at its optimum H_t^* – we work our way backwards towards H_t. Table 3-2 illustrates the recursive algorithm. At optimal employment size, we know that $\lambda_{H_t^*}^*$ must necessarily be zero. Thus, the value function $V(H_t^*, 0)$ depends only on present profit and the discount rate – both of which are known. Stepping backwards, i.e. $H_t^* \mp 1$ with $-$ for hiring and $+$ for firing processes, the optimal rate $\lambda_{H_t^* \mp 1}^*$ depends on the downstream value function $V(H_t^*, 0)$, which we just determined. Evaluating the value function $V\left(H_t^* \mp 1, \lambda_{H_t^* \mp 1}^*\right)$ is also straightforward now that the optimal adjustment rate is known. At each step, we insert the optimal rate into the vector describing the firm's decision rule Λ_t. Repeating the recursive steps until H_t provides the full trajectory. Finally, note that in Table 3-2, I have omitted the hurdle effect, as the purpose is to illustrate the recursive solution. In practice, the hurdle potentially forces several (or all) adjustment rates to 0 starting from the end H_t^*. The overall shape of the trajectories will remain the same, with rates diminishing before eventually reaching 0, however they may do so prior to reaching the optimal employment size.

Table 3-2: Illustration of the recursive solution to the firm's MDP

H	λ_H^*	$V(H,\lambda_H)$	Λ_t
H_t^*	0	$\frac{\pi(H_t^*,x_t,\varepsilon_t)}{\delta}$	(0)
$H_t^* \mp 1$	$-\delta+\sqrt{\delta^2+\frac{\delta\, V(H_t^*,0)-\pi(H_t^*\mp 1)+\mu_0-\delta\mu_1}{\mu_2}}$	$\frac{\pi(H_t^*\mp 1)-a\left(\lambda_{H_t^*\mp 1}\right)+\lambda_{H_t^*\mp 1}V(H_t^*,0)}{\delta+\lambda_{H_t^*\mp 1}}$	$\left(\lambda_{H_t^*\mp 1},0\right)$
$H_t^* \mp 2$	$-\delta+\sqrt{\delta^2+\frac{\delta\, V\left(H_t^*\mp 1,\lambda_{H_t^*\mp 1}\right)-\pi(H_t^*\mp 2)+\mu_0-\delta\mu_1}{\mu_2}}$	$\frac{\pi(H_t^*\mp 2)-a\left(\lambda_{H_t^*\mp 2}\right)+\lambda_{H_t^*\mp 2}V\left(H_t^*\mp 1,\lambda_{H_t^*\mp 1}\right)}{\delta+\lambda_{H_t^*\mp 2}}$	$\left(\lambda_{H_t^*\mp 2},\lambda_{H_t^*\mp 1},0\right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
H_t	$-\delta+\sqrt{\delta^2+\frac{\delta\, V\left(H_t\pm 1,\lambda_{H_t\pm 1}\right)-\pi(H_t\ )+\mu_0-\delta\mu_1}{\mu_2}}$	$\frac{\pi(H_t)-a\left(\lambda_{H_t}\right)+\lambda_{H_t}V\left(H_t\pm 1,\lambda_{H_t\pm 1}\right)}{\delta+\lambda_{H_t}}$	$\left(\lambda_{H_t},\dots,\lambda_{H_t^*\mp 1},0\right)$

3.2 Likelihood Function

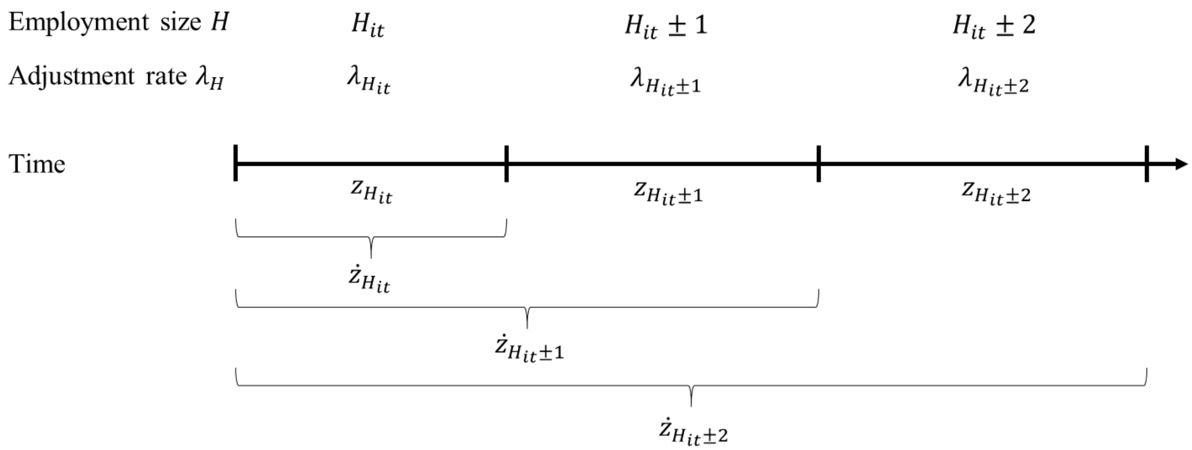
With the firm decision rule for Λ_{it} in hand, the next step is to write out the likelihood function. Let θ_i denote the vector of firm i's firm-specific parameters, including those of the adjustment cost function $\alpha(\lambda_H)$ and the yet unspecified profit function $\pi(H, x_{it}, \varepsilon_{it})$.

$$\boldsymbol{\mathcal{L}}(\theta_i \forall i, \varepsilon_{it} \forall it) = \prod_i \prod_t P(\Delta H_{it} | \Lambda_{it}) = \prod_i \prod_t P(\Delta H_{it} | \theta_i, \varepsilon_{it}) \tag{3.12}$$

The likelihood function is the joint probability of all the observations given the firms' chosen trajectories, which depend on the parameters and error terms as shown in Figure 3-1. I am treating the errors ε_{it} as known quantities rather than random variables; this circumvents the need for numerically integrating over the PDF of the error when evaluating the likelihood function. Instead, it is drawn alongside the parameters θ_i in the Bayesian estimation procedure, presented in the next section.

The conditional probability $P(\Delta H_{it} | \Lambda_{it})$ is the probability of observing ΔH_{it} given the firm's chosen trajectory Λ_{it}. If all the adjustment rates in Λ_{it} were equal, then the probability of ΔH follows the PMF of a Poisson distribution with that fixed rate parameter. However, as we saw in the previous section, the adjustment rates diminish as the employment size H approaches its optimum. Thus, we require a PMF of a stochastic process with exponential distributed inter-event times that allows varying rate parameters. To this end, some additional notation is helpful. Let $\dot{z}_H$ be an exponential-distributed random variable representing the time of the event associated with λ_H but measured from the beginning of period t. The timeline in Figure 3-3 illustrates the difference between $\dot{z}_H$ and z_H

Figure 3-3: Relationship between $\dot{z}_H$ and z_H



$\dot{z}_H$ is by definition greater than the upstream adjustment event time, i.e. $\dot{z}_{H_{it}} < \dot{z}_{H_{it}\pm 1} < \cdots < \dot{z}_{H_{it}\pm\Delta H_{it}}$ and follows the same exponential distribution as z_H with rate parameter λ_H described in (3.5). This can be seen by the memoryless property.

$$\frac{f_H(\dot{z}_H)}{\tilde{F}_H(\dot{z}_{H-1})} = \frac{\lambda_H e^{-\lambda_H \dot{z}_H}}{e^{-\lambda_H \dot{z}_{H-1}}} = \lambda_H e^{-\lambda_H(\dot{z}_H - \dot{z}_{H-1})} = \lambda_H e^{-\lambda_H z_H} = f_H(z_H) \tag{3.13}$$

Let T denote the length of period t. We can write the conditional probability $P(\Delta H_{it}|\theta_i, \varepsilon_{it})$ as the joint probability that all observed adjustments occur before time T and that the subsequent adjustment occurs after time T given that the adjustments happen sequentially.

$$\begin{aligned} P(\Delta H_{it}|\Lambda_{it}) = \Pr(&\dot{z}_{H_{it}} \leq T \cap \dot{z}_{H_{it}\pm 1} \leq T \cdots \cap\ \dot{z}_{H_{it}\pm\Delta H_{it}-1} \leq T \cap \\ &\dot{z}_{H_{it}\pm\Delta H_{it}} > T \mid \dot{z}_{H_{it}} < \dot{z}_{H_{it}+1} < \cdots < \dot{z}_{H_{it}+\Delta H_{it}}) \end{aligned} \tag{3.14}$$

To evaluate this joint probability, we must consider the possible values of all the event times $\dot{z}_H$, which yields a nested integral.

$$\begin{aligned} &P(\Delta H_{it}|\Lambda_{it}) \\ &= \int_0^T \cdots \int_{\dot{z}_{H_{it}+\Delta H_{it}-2}}^T f_{H_{it}}(\dot{z}_{H_{it}}) \frac{f_{H_{it}+1}(\dot{z}_{H_{it}+1})}{\tilde{F}_{H_{it}+1}(\dot{z}_{H_{it}})} \cdots \frac{\tilde{F}_{H_{it}+\Delta H_{it}}(T)}{\tilde{F}_{H_{it}+\Delta H_{it}}(\dot{z}_{\Delta H_{it}})} d\dot{z}_{H_{it}+\Delta H_{it}-1} \cdots d\dot{z}_{H_{it}} \end{aligned} \tag{3.15}$$

Substituting in the PDF and survival functions allows us to simplify the integrand.

$$
\begin{aligned}
&P(\Delta H_{it}|\Lambda_{it}) \\
&= \int_0^T \cdots \int_{\dot{z}_{H_{it}+\Delta H_{it}-2}}^T \lambda_{H_{it}} e^{-\lambda_{H_{it}}\dot{z}_{H_{it}}} \frac{\lambda_{H_{it}+1} e^{-\lambda_{H_{it}+1}\dot{z}_{H_{it}+1}}}{e^{-\lambda_{H_{it}+1}\dot{z}_{H_{it}}}} \cdots \\
&\qquad \frac{e^{-\lambda_{H_{it}+\Delta H_{it}}T}}{e^{-\lambda_{H_{it}+\Delta H_{it}}\dot{z}_{H_{it}+\Delta H_{it}}}} d\dot{z}_{H_{it}+\Delta H_{it}-1} \cdots d\dot{z}_{H_{it}} \qquad (3.16) \\
&= \left(\prod_{H=H_t}^{H_{it}+\Delta H-1} \lambda_H\right) e^{-\lambda_{H_{it}+\Delta H_{it}}T} \int_0^T \cdots \int_{\dot{z}_{H_{it}+\Delta H_{it}-2}}^T e^{-\left(\lambda_{H_{it}}-\lambda_{H_{it}+1}\right)\dot{z}_{H_{it}}} \ldots \\
&\qquad e^{-\left(\lambda_{H_{it}+\Delta H_{it}-1}-\lambda_{H_{it}+\Delta H_{it}}\right)\dot{z}_{H_{it}+\Delta H_{it}-1}} d\dot{z}_{H_{it}+\Delta H_{it}-1} \cdots d\dot{z}_{H_{it}}
\end{aligned}
$$

Evaluating these nested integrals and rearranging, we arrive at the probability of each observation:

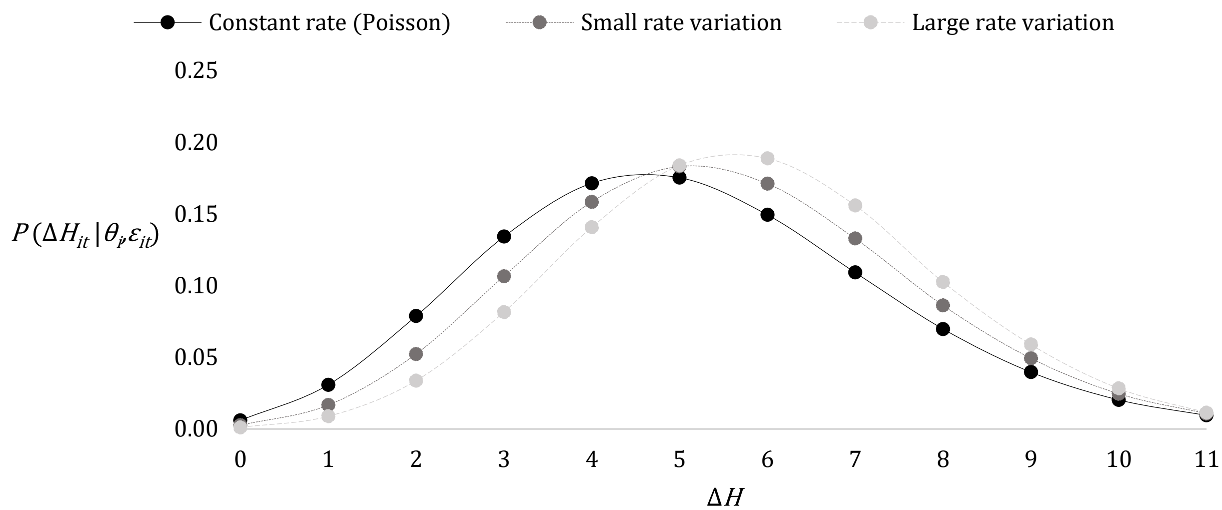
$$
P(\Delta H_{it}|\Lambda_{it}) = \left(\prod_{H=H_{it}}^{H_{it}+\Delta H_{it}-1} \lambda_H\right) \sum_{H=H_{it}}^{H_{it}+\Delta H_{it}} \frac{e^{-\lambda_H T}}{\prod_{\substack{H'=H_{it} \\ H'\neq H}}^{H_{it}+\Delta H_{it}} (\lambda_{H'} - \lambda_H)} \qquad (3.17)
$$

While convoluted, this can crucially still be evaluated analytically, which keeps the model estimation computationally tractable. As we might expect, this transition probability distribution has a structure that resembles that of the Poisson distribution with constant rate parameter λ.

$$
q(\Delta H|\lambda) = \lambda^{\Delta H} \frac{e^{-\lambda T}}{\Delta H!} \qquad (3.18)
$$

In fact, as the differences between the rates λ_H approach 0, the probability in (3.17) approaches a Poisson distribution with a fixed rate parameter, as shown in Figure 3-4. The more concave the profit function $\pi(H)$ is with respect to employment size, the more rapidly the adjustment rate diminishes with each change. In turn, this results in a more left-skewed (i.e. right-leaning) distribution.

Figure 3-4: Comparing probability functions with same mean rate but different levels of rate variation



Finally, the likelihood function (3.19) is the product of the transition probabilities over all the observations in the dataset.

$$\begin{aligned}
\boldsymbol{\mathcal{L}}(\theta,\varepsilon) &= \prod_i \prod_t P(\Delta H_{it}|\theta_i,\varepsilon_{it}) \\
&= \prod_i \prod_t P(\Delta H_{it}|\Lambda_{it}) \\
&= \prod_i \prod_t \left[\left(\prod_{H=H_{it}}^{H_{it}+\Delta H_{it}-1} \lambda_H \right) \sum_{H=H_{it}}^{H_{it}+\Delta H_{it}} \frac{e^{-\lambda_H T}}{\prod_{\substack{H'=H_{it} \\ H'\neq H}}^{H_{it}+\Delta H_{it}} (\lambda_{H'} - \lambda_H)} \right]
\end{aligned} \tag{3.19}$$

3.3 Bayesian Estimation Procedure

The likelihood function (3.19) is highly non-convex. Using a traditional maximum likelihood estimation technique, the estimation procedure would too often *get stuck* in local extrema. Instead, I formulate a hierarchical Bayesian model and use an MCMC sampling algorithm to estimate the parameters. I assume that the model's hyperparameters are normal-distributed. In other words, each firm i is associated with a vector of parameters θ_i; across all firms, these form a multi-variate normal distribution with unknown mean Θ and variance W. Additionally, I assume that the observation-specific error terms ε_{it} is Normal-distributed with zero-mean and variance σ^2. The objective of the estimation procedure is to repeatedly sample from these distributions until we can determine the hyperparameters Θ, W, and σ^2.

Let $K(\Theta, W, \sigma^2 | \Delta H_{it} \forall it)$ be the probability distribution of these parameters given all the observed outcomes. This is the so-called posterior distribution. By Bayes' theorem, we have

$$\begin{aligned} K(\Theta, W, \sigma^2 | \Delta H_{it} \forall it) &= \frac{P(\Delta H_{it} \forall it | \Theta, W, \sigma^2) k(\Theta, W, \sigma^2)}{P(\Delta H_{it} \forall it)} \\ &= \prod_i \prod_t \frac{P(\Delta H_{it} | \Theta, W, \sigma^2) k(\Theta, W, \sigma^2)}{P(\Delta H_{it})} \end{aligned} \tag{3.20}$$

where $k(\Theta, W, \sigma^2)$ is prior distribution, i.e. our best guess prior to observing the data. Reflecting the lack of a priori knowledge, I assume diffuse priors. Namely, I assume for Θ a Normal distribution with unboundedly large variance. For W, I assume an Inverse Wishart distribution with V degrees of freedom, where V is the length of the vector Θ, and scale matrix J, a V-dimensional identity matrix. Note however that for parameters with variances much smaller than unity, this diffuse prior is actually informative for practical sample sizes. In such cases, the use of alternative (Schuurman et al., 2016) or improper (Asparouhov, 2010) priors is preferred. For σ^2, I assume an Inverse Gamma distribution, which is the univariate version of the Inverse Wishart. Both shape and scale parameters for this diffuse prior are set to 1, such that we have $IG(1, 1)$.

Note that the denominator on the right-hand side $P(\Delta H_{it})$ of equation (3.20) is independent of the parameters Θ, W, and σ^2. Thus, we can replace the equality with a proportionality and get

$$K(\Theta, W, \sigma^2 | \Delta H_{it} \forall it) \propto \prod_i \prod_t P(\Delta H_{it} | \Theta, W, \sigma^2) k(\Theta, W, \sigma^2) \tag{3.21}$$

Sampling directly from (3.21) using a Metropolis-Hastings algorithm is theoretically possible. However, it requires repeated evaluations of numerical integrals, making it practically infeasible. Instead, we can consider the realizations of the model parameters for each firm θ_i and all the error terms ε_{it} part of the joint distribution that we sample from.

$$K(\Theta, W, \sigma^2, \theta_i \forall i, \varepsilon_{it} \forall it | \Delta H_{it} \forall it) \propto \left[\prod_i \phi(\theta_i | \Theta, W) \prod_t P(\Delta H_{it} | \theta_i, \varepsilon_{it}) \phi(\varepsilon_{it} | 0, \sigma^2)\right] k(\Theta, W, \sigma^2) \tag{3.22}$$

Now the right-hand side consists only of our likelihood function (3.19), the Normal probability density functions for the parameters and error terms, and the priors. We can sample from this joint posterior distribution by repeating the following five-step Gibbs sampling algorithm until convergence:

1. Draw θ_i for each firm i conditional on b, W, $\varepsilon_{it} \forall t$ and the observed data $\Delta H_{it} \forall t$.
   I use a Metropolis-Hastings algorithm to make these draws; evaluating trial values by the conditional probability (3.23)

$$K(\theta_i | b, W, \varepsilon_{it} \forall t, \Delta H_{it} \forall t) \propto \prod_t P(\Delta H_{it} | \theta_i, \varepsilon_{it}) \phi(\theta_i | b, W) \quad \forall i \tag{3.23}$$

2. Draw ε_{it} conditional on σ^2, θ_i, and the observed data ΔH_{it}.
   Again, I use a Metropolis-Hastings algorithm to make these draws. Here we evaluate trial values by the conditional probability (3.24).

$$K(\varepsilon_{it} | \sigma^2, \theta_i, \Delta H_{it}) \propto P(\Delta H_{it} | \theta_i, \varepsilon_{it}) \phi(\varepsilon_{it} | 0, \sigma^2) \quad \forall it \tag{3.24}$$

3. Draw b conditional on W and $\theta_i \forall i$.
   This is a draw from a Normal distribution (3.25)

$$K(b|W, \theta_i \forall i) = \mathcal{N}\left(\frac{\sum_i \theta_i}{N}, \frac{W}{N}\right) \tag{3.25}$$

where N is the total number of firms.

4. Draw W conditional on Θ and $\theta_i \forall i$.

   This is a draw from the Inverse Wishart distribution (3.26)

$$K(W|\Theta, \theta_i \forall i) = IW\left(V + N, \frac{VJ + N\bar{S}}{V + N}\right) \tag{3.26}$$

where $\bar{S} = \frac{\sum_i (\theta_i - \Theta)(\theta_i - \Theta)'}{N}$, V is the number of parameters to be estimated, i.e. the length of each θ_i, and J is a V-dimensional identity matrix

5. Draw σ^2 conditional on $\varepsilon_{it} \forall it$

   This is a draw from the Inverse Gamma distribution (3.27)

$$K(\sigma^2|\varepsilon_{it} \forall it) = IG\left(1 + N, \frac{1 + N\bar{s}}{1 + N}\right) \tag{3.27}$$

where $\bar{s} = \frac{\sum_{it} (\varepsilon_{it} - \sigma^2)(\varepsilon_{it} - \sigma^2)'}{R}$, and R is the total number of observations in the dataset.

I provide descriptions of the Gibbs and Metropolis-Hastings sampling algorithms in Appendix B. The likelihood function is evaluated in draws 1 and 2. Each of these evaluations require that we solve the firms' MDPs for each firm in each period.

3.4 Discussion

The proposed model structure offers a flexible analysis framework for modelling firms' employment size decisions over time. In this dissertation, I use it to examine the impacts of agglomeration economies. Specifically, I estimate the elasticity of productivity and labour market frictions with respect to various agglomeration measures in Chapter 4 and consider two scenario impact studies in Chapter 5 that explore the effects of a light-rail transit (LRT) line extension and quantify the costs of labour market frictions. However, the framework can be used to analyse any variety of subjects relevant for the evolution of the employment size of businesses. To examine specific subjects of interest, the researcher simply needs to specify the profit and adjustment cost functions with the relevant variables. For example, to examine the impact of economic policy, such as enterprise tax credits, the researcher only needs to make the appropriate changes to the profit function. Technological change can be modelled by appropriately changing the production function. For example, automation might change the total factor productivity and the returns to capital and labour inputs, and similarly, the ability to work remotely might change the returns to land and capital. The specification of the adjustment cost function is also a potentially fruitful avenue for research. By including the relevant variables, we can explore the impacts of spatial mismatches between the skills of workers and those demanded by businesses. If such mismatches increase the costs of hiring, we can model steady state unemployment at the micro-level. Furthermore, with measures of social network density and quality, we can also explore their impact on the job search costs. I present these examples to illustrate the flexibility of the framework. It merely provides a structure for modelling the decisions of businesses that are profit-maximizing and forward-looking.

Challenges and limitations

The modelling efforts presented in this chapter take several steps towards more economically sound representations of firm evolution at the microscopic level. However, these advances are certainly not without their limitations and areas for improvement. First and foremost, the complexity introduced by the structural approaches such as this make the models considerably more esoteric; they are less transparent, and their results are more difficult to interpret. In turn, this demands more precise communication from the researcher and more attention and effort from the audience. Furthermore, the added sophistication of the model relies on detailed panel datasets and may not worthwhile if data of adequate quality are not available.

In practice, uncertainty and imperfect information about current conditions are likely significant contributors to inaction and inertial responses on firms' parts. The proposed model, in its present form, does not account for these effects explicitly. Instead, they may inadvertently be captured, partially or fully, by the μ parameters. Relatedly, the model assumes that firms are agnostic with respect to the future values of exogeneous variables x and instead merely assume that current conditions persist. However, in reality expectations about future conditions likely play an important role in firm decision-making. Executives will form expectations and make decisions based on subject matter expertise and historical trend extrapolation. Although these rarely align perfectly with actual development, they are more informative than the constancy assumption made in the model. Introducing these expectations to the firms' MDP would not add considerably complexity to the MDP or its solution. However, gathering data and formulating a model for the expectations requires considerable effort and is beyond the scope of this dissertation.

The exponential-distributed inter-event times imply that adjustment events occur one at a time. Although perhaps technically accurate at the most microscopic level, this is likely not an accurate representation of all hiring and firing decisions. In particular, one can imagine employment size decisions being more individual for higher skill jobs in small firms, whereas they are likely made less on an individual basis for large, low-skill firms. Furthermore, hiring and firing are generally interspersed between each other in a firm's employment size trajectory. However, the model structure does not allow that. Specifically, the likelihood function assumes that each observation only comprises adjustments in one direction. At first glance, this may appear quite restrictive. However, its impact should be small in practice. The model effectively disregards direct replacements, e.g. if an employee quits and the firm replaces them, as these changes are not resulting from changing profits. In other words, the issue only persists when a firm hires and fires employees due to profit fluctuations within observations. However, in many cases, the employment data is of the same or higher granularity than the explanatory variables.

Relation to Dynamic Programming Discrete Choice Models

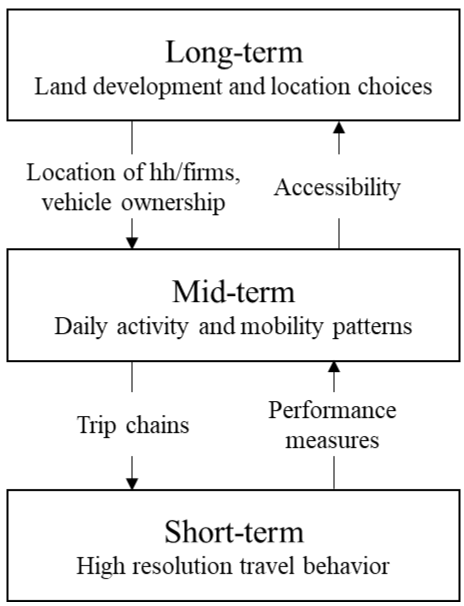
The model presented in this chapter is closely related to the class of models known as dynamic programming discrete choice (DPDC) models. As their name suggests, these models combine discrete choice models with dynamic programming. Specifically, traditional discrete choice models do not explicitly have a temporal dimension. Thus, decision-makers implicitly maximize instantaneous utility. On the other hand, DPDC models explicitly consider the temporal dimension by assuming that future utility depends on current choices and that decision-makers maximize (the present value of) lifetime utility instead of instantaneous utility. Such models of intertemporal decision-making processes have found application in many different contexts. For example, whether or not to have a child (Wolpin, 1984), job search (Miller, 1984), patent renewals (Pakes, 1986), bus engine replacement (Rust, 1987), occupation and schooling decisions (Keane and Wolpin, 1997), retirement behaviour from the labour force (Rust and Phelan, 1997; Karlstrom et al., 2004). My originally proposed dissertation work was another such model aimed at examining the impact of accessibility and agglomeration economies on workers' wage progression. The central sections from that proposal are available in Appendix C. Aguirregabiria and Mira (2009) provide an overview of methods and issues related to DPDC models.

Although similar, the model presented here differs from the typical DPDC model in a few ways that simplify the solution of the MDP and the estimation. Most obviously, firms do not choose from a discrete set of choice alternatives but instead decide on their adjustment rates, which can take any non-negative value. However, more importantly, the assumptions made here for the firm's MDP significantly simplify its solution. Specifically, the state space is defined only by the firm's employment size because the environment, i.e. exogenous variables x and crucially the error terms ε, are assumed constant. Furthermore, the relatively simple MDP enable the use of Bayesian estimation methods. Such methods would otherwise be infeasible, since re-solving a more complex MDP at each iteration of the Gibbs sampler quickly becomes too computationally burdensome (Aguirregabiria and Mira, 2009). In turn, the Bayesian estimation procedure allows me to simulate the errors. This considerably simplifies the evaluation of the likelihood function, which otherwise would require integration over the distribution of the error.

Integration in IUM framework

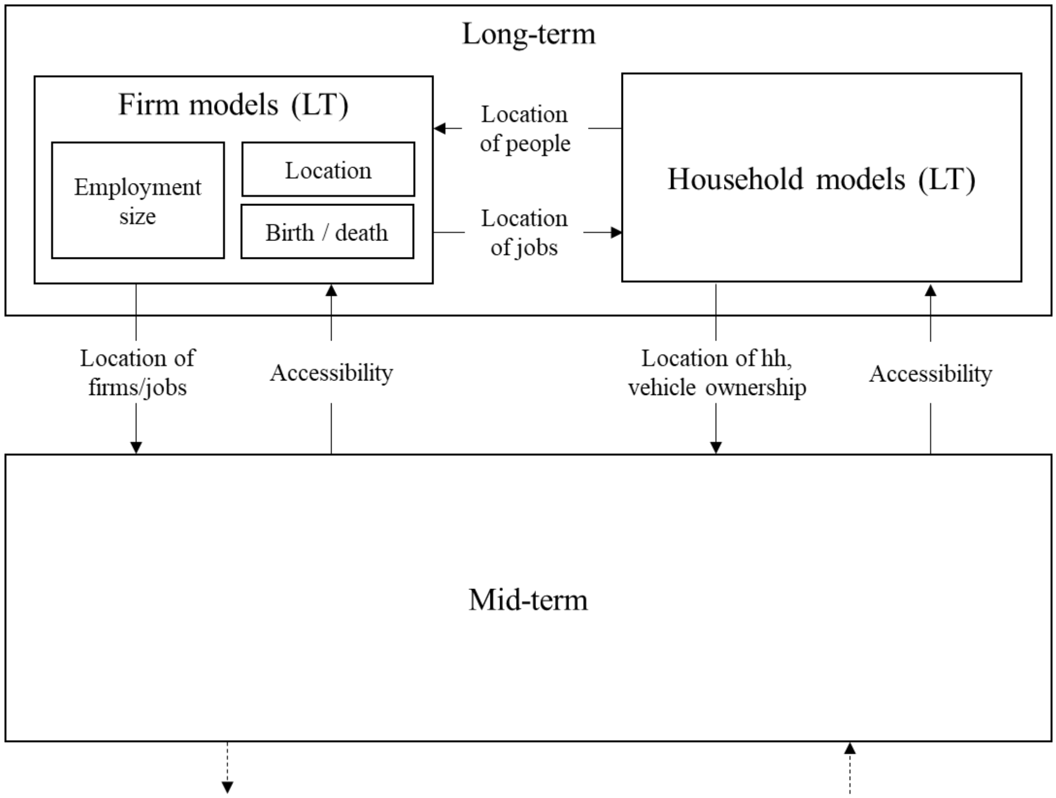
The modelling work presented in this dissertation focuses on the hiring and firing of employees. Of course, this only represents one component of urban development, and the intended use of the model is as a part of a larger IUM framework. Figure 3-5 shows the structure of the SimMobility framework, one of the state-of-the-art activity-based IUMs (Adnan et al., 2015). It models urban activity at three different timescales, ranging from second-to-second decisions in traffic in the short-term module, through day-to-day activity decisions in the mid-term module, to life and lifestyle decisions in the long-term module.

Figure 3-5: SimMobility framework, adapted from Adnan et al. (2015)



Within this framework, our firm employment size model fits in the long-term module. Figure 3-6 presents an example schematic of how the employment size model could be integrated within a SimMobility-like framework.

Figure 3-6: Example integration of employment size model in SimMobility-like framework



Although the proposed employment size model is able to operate in continuous time, there is generally no benefit to doing so in an integrated model. Instead it will adhere to the time steps of the larger framework. For SimMobility LT, this is typically year-by-year. The flexibility of the proposed model does mean that it can be adapted to operate with time steps of any size in a theoretically consistent manner.

Firm behaviour, as it is relevant to integrated urban models, is generally captured by three (types of) models: birth and death models, location models, and employment size growth and decline models. Although all three models describe firm behaviour, they do not interact directly. For each time step, a birth model creates new firms for each sector. Initially, these firms will comprise a single employee and will not yet be assigned a location. Then, an awakening model determines the subset of firms that were looking to relocate. This model can be formulated as a survival model or a simple binary choice. SimMobility determines residential relocations by modelling the real estate market with a bidding model that simulates households' willingness-to-pay for available locations. A similar bidding model could be developed for commercial real estate. Once all firm locations, including those of new firms, have been modelled, we can use the employment size model to determine the growth and decline of each firm. Finally, firm closures can also be handled by the employment size model, i.e. a firm closes if its optimal employment size is non-positive. However, this requires that the data used to estimate the employment size model includes firm deaths. If this is not the case, deaths should be modelled as a separate process.

Naturally, firms interact with the larger urban system in numerous ways. This is reflected by firm model outputs serving as inputs for other models and vice versa. Within the long-term module, firm models interact with household models with flow of information in both directions. A firm's labour pool depends on the location and composition of households. Additionally, the location of people is a key determinant in market access for consumer-facing firms. Conversely, the firm models collectively provide the location and quantity of jobs, which influence household residential location decisions.

The assignment of workers to jobs should ideally be a matching model that finds the best match between the workers' skill sets and the tasks associated with available jobs – similar to the residential bidding model with wages representing prices.

The firm models also interact with the models in the mid-term module with data being passed in both directions. The transportation system plays a central role in facilitating agglomeration economies and is naturally a key determinant of the costs of transporting goods between firms. These inputs come from models in the mid-term module. On the other hand, the location of jobs and firms serve as destinations for work trips as well as many leisure and shopping trips.

The focus on employment size adjustments in this dissertation should primarily be attributed to limited data availability. In particular the primary firm dataset, the historical micro-level business records from InfoGroup, does not describe firm birth and death events nor does it consistently identify relocations. Furthermore, a satisfactory location model requires additional real estate market data, including transaction prices, asking prices, availabilities, etc. I provide detailed descriptions of the datasets in this dissertation in Chapter 4.1.

Applications beyond employment

The presented model was developed specifically for employment size adjustment processes. However, it is potentially useful for other applications beyond employment. In general, the model describes random incidence processes where the event rate changes between subsequent events. A timely example of an alternative application is for modelling the timing of repeated Covid-19 infections. Since the development of antibodies following an infection lowers the risk of subsequent reinfection, the event rate effectively changes.

In the urban transportation context, the archetypical applications of Poisson processes are passenger arrivals at a bus stop and queuing systems. The presented varying-rate model can also be applicable for these purposes, especially in low-demand cases. For example, if on average only a few passengers arrive at a bus stop each hour, and several just arrived, it would seem likely that subsequent arrival rates are lower because demand has already been depleted. However, in most cases, particularly higher demand ones, the memoryless fixed-rate properties of the Poisson are adequate approximations and results in much simpler models. Activity scheduling models are perhaps a more appropriate and worthwhile application of the varying-rate model. For most people, activities such as shopping, leisure, and exercise happen with some level of regularity but are generally not fixed. For example, if someone usually makes two grocery shopping trips each week, it is much more likely that they will go on any given day if they have yet to go than if they had already gone twice this week. The events are not independent because they depend on the person’s demand for groceries which is finite within any given time period.

4 APPLICATION TO BOSTON

4.1 Data

In the following sections, I describe the datasets used for the empirical application of the model to the Greater Boston Area. No new data were collected specifically for this project. Instead, I make use of a combination of existing datasets, including commercial business intelligence data, transportation and land use data provided by local planning agencies, and publicly available economic datasets. The use of already existing datasets both demonstrates what can be achieved without a dedicated data collection effort and pinpoints where better data would be beneficial. Table 4-1 below provides an overview of the datasets I used, including the spatial and temporal resolutions of each dataset.

Table 4-1: Overview of datasets for empirical application in the Greater Boston Area

Dataset	**Source**	**Year(s) available**	**Temporal resolution**	**Spatial resolution**
Historical Micro-level Business Records	InfoGroup (2014)	2003-13	Months	Latitude, longitude; addresses
Population and Employment Forecasts	Metropolitan Area Planning Council (2009)	2000, 2030	Months (interpolated)	TAZ
Travel Time and Cost Matrices	Central Transportation Planning Staff (2007)	2005	N/A	TAZ
Massachusetts Travel Survey	Central Transportation Planning Staff (2012)	2010-11	N/A	Block groups; PUMA

Dataset	Source	Year(s) available	Temporal resolution	Spatial resolution
Land Parcels	Massachusetts Bureau of Geographic Information (2019)	2009	N/A	Parcels
GDP by Industry	U.S. Bureau of Economic Analysis. (2021b)	2003-13	Years	N/A (Boston-Cambridge-Newton MSA average)
Compensation by Industry	U.S. Bureau of Economic Analysis. (2021c)	2003-13	Years	N/A (Boston-Cambridge-Newton MSA average)
Value Added by Industry	U.S. Bureau of Economic Analysis. (2021e)	2003-13	Years	N/A (U.S. average)
Employment by Industry	U.S. Bureau of Economic Analysis. (2021a; 2021d)	2003-13	Years	N/A (U.S. average; Boston-Cambridge-Newton MSA average)
Weighted Average Cost of Capital by Industry	Damodaran (2021)	2021	N/A	N/A

One overarching drawback of using such disparate data sources is the additional cleaning and processing necessary to ensure alignment in spatial and temporal resolution. Inevitably, the aggregation to shared units introduces some error. Table 4-2 summarizes and illustrates the differences in spatial resolution between public use micro areas (PUMAs), traffic analysis zones (TAZs), block groups, and parcels.

Table 4-2: Summary of spatial resolutions

Resolution	Count	Mean area	Median Area
Public Use Micro Areas (PUMA)	38	256 km^2	170 km^2
Traffic Analysis Zones (TAZ)	2728	2.69 km^2	1.08 km^2
Block Groups	3337	2.11 km^2	0.595 km^2
Parcels	1039437	0.00485 km^2	0.00277 km^2

Historical Micro-level Business Records

The historical micro-level business records are a panel dataset, providing information about company names, locations (addresses and latitude-longitude), employment sizes, which sectors they operate in, and more (InfoGroup 2014). The dataset is compiled by InfoGroup, a data and marketing services company based in the U.S. The business records used here are one of their commercial products that has been made available for academic purposes at MIT and Harvard through the Harvard Dataverse. InfoGroup compiled and continuously updated the data by combining telephone interviews with publicly available data sources, including yellow pages, government sources, public company filings, points-of-interest compilations. The academic-use dataset is published in annual snapshots, saved each December between 2003 and 2013.

The variables of interest for this study remained present in the dataset in all snapshots. However, their values were unfortunately not always consistent. In other words, while business name, address, and coordinates were available in all years, their values for the same business in the same location sometimes change. For example, the same business might change their name, be recorded with/without corporate suffixes (e.g. ltd. Or inc.), or simply be misspelled in certain years. Similarly, locations might be recorded with inconsistent address numbering, e.g. when located in large malls spanning multiple numbers, with/without unit information, or with inconsistent latitude-longitude information jumping between parcel centroid coordinates and curbside coordinates. To alleviate challenges linking records of the same business across different years, I use InfoGroup's own business ID flag (known as ABI, short for American Business Information). On inspection, it appears to handle and overcome inconsistencies in other identifying fields. However, due to the proprietary nature of the data and ABI field definition, it is unknown how records of the same business are linked over time. In particular, this can be important for identifying relocations versus business birth/death.

The employment size is one of the key variables that InfoGroup elicited through telephone interviews. However, these calls are made continuously and irregularly, as opposed to on a fixed schedule or with a fixed frequency. Consequently, businesses generally do not have updated records in every annual snapshot. This poses several challenges for the analysis. First, the dataset is an unbalanced panel with many observations for some firms and few for others. Fortunately, the model formulation is built to handle irregular inter-observation durations. Second, the lack of a fixed frequency makes it impossible to identify firm births and deaths. Periods without observations before the first and after the last data point could represent either gaps in the call schedule or the business no longer existing. Crucially, if the latter is prevalent, it can lead to attrition bias in the estimation. I elaborate on this in the limitations section.

For this Boston study, I only consider records within the Greater Boston Area as defined by the Boston Metropolitan Area Planning Council (MAPC). I discard records where businesses' employment sizes were not verified by phone interviews. Additionally, I discard records for which the provided address and coordinates could not reliably be geocoded to a parcel in the MassGIS parcel data (see below for a description of the parcel data). Geocoding was attempted by three methods: (1) matching the address strings between the records and parcels, (2) matching the location of the record coordinates and the parcels, and (3) finding the nearest parcel to the record coordinates. The last method was necessary because a large number of record coordinates were placed curbside, rather than within the parcel boundaries. Records whose coordinates did not have a parcel within 50 metres were discarded. Finally, I discard firms with only a single record since the analysis requires information about employment size over time. In total, the final dataset comprises 253,826 records from 79,360 unique businesses. Table 4-3 presents record and firm counts for the three sectors of interest.

Table 4-3: Records and firms in manufacturing, professional services, and food and accommodation services

Sector	Records	Firms
Manufacturing	13745	3939
Professional	29225	8518
Food and accommodation	15846	4989

(InfoGroup, 2014)

Population and Employment Forecasts

The spatial distributions of population and employment are required to measure agglomeration. Using forecasts from the MAPC's 2030 regional plan, known as MetroFuture (Metropolitan Area Planning Council, 2009), I interpolate population and employment by TAZ for each month between 2003-13. The plan takes population and employment numbers by TAZ based on the 2000 US Census and information from the Central Transportation Planning Staff (CTPS) and projects their development to 2030. Note that while the historical business records from InfoGroup provides information about individual firms' employment, they do not cover every business or provide expansion factors. Thus, the MAPC forecasts are likely reliable for aggregate counts. That said, the use of forecasts rather than observed data introduces its own problems. Specifically, their appropriateness as a proxy for actual development depends on the accuracy of the forecasting model and interpolation. This inevitably introduces additional error. On the other hand, since forecasts are by definition made independently of the actual development, they are useful for addressing potential simultaneity concerns.

Travel Time and Cost Matrices

The transportation system is represented by travel time and cost matrices connecting all TAZs to each other. These matrices are provided by the CTPS (2007). The data describe travel times and costs by auto (single and high occupancy vehicles) and transit modes. These are based on CTPS's 2005 model. Additionally, the dataset also includes a distance matrix, which is useful for modelling travel impedance for active modes. Unfortunately, these travel time and cost data were only available for a single point in time, i.e. they are cross-sectional. Thus, this study does not capture the impacts of changes to the transportation system over the analysis period.

Massachusetts Travel Survey 2010-11

The Massachusetts Travel Survey elicits information about individuals' and households' travel preferences, e.g. mode choices and vehicle ownership decisions, and relate these to various socio-demographic characteristics (CTPS, 2012). Specifically, the survey comprised a 24-hour weekday trip diary and a follow-up questionnaire administered by phone or mail. The data collection period spanned from May 2010 to October 2011. During this period, a total of 15,033 households in Massachusetts completed the survey. Of these, 7,661 resided in the Greater Boston Area (i.e. within MAPC jurisdiction). The sample was then expanded by iterative proportional fitting (IPF) to align with the 2006-2010 American Community Survey (ACS). The Massachusetts Department of Transportation (2012) report provides a complete description of the survey methods and results.

I use the data from the Massachusetts Travel Survey for two purposes: First, I estimate a commute mode choice model, whose coefficients I use for calculating accessibility measures. Second, I extract and expand survey responses vis-à-vis education to calculate average local education levels, which I control for in the model estimation. Similar to the travel time and cost matrices, these data are unfortunately cross-sectional only.

Massachusetts Land Parcel Database

The Massachusetts Bureau of Geographic Information (MassGIS) maintains a publicly accessible database of all the land parcels in the commonwealth (MassGIS, 2019). It has information about parcel sizes, number of occupants, publicly assessed values, land use designation, etc. I join this dataset with the historic micro-level business records to determine how much land each business occupies. The dataset being cross-sectional likely does not introduce considerable error, since parcels would mostly be limited to renovation rather than redevelopment while occupied. However, the dataset is updated to 2019, meaning that parcels and building stock may not match conditions in 2003-13, especially in rapidly developing areas.

Aggregate Economic Data

To control for macroeconomic trends, I use several aggregate economic statistics, including GDP, average compensation per employee, and value-added per employee. All statistics are stratified by year and sector. The data are published by the U.S. Bureau of Economic Analysis and are publicly accessible. They are available at different levels of aggregation. Specifically, GDP and compensation data can be found for the Boston-Cambridge-Newton MSA, while value-added data are only available at the national level. Although the Boston-Cambridge-Newton MSA does not perfectly overlap with the Greater Boston Area as defined by the MAPC jurisdiction, it covers the majority of the same economic centres. Thus, the statistics here are likely more representative of the Greater Boston Area than state or national averages.

Weighted Average Cost of Capital

Weighted average cost of capital (WACC) is a measure of a firm's cost to raise money and is a key parameter in determining its profitability. Damodaran (2021) collected data for 7582 businesses in the U.S and calculated the WACC for different sectors. For this study, I use these WACC as the discount rate, as specified in Chapter 3.1, of the firms in the respective sectors.

4.2 Descriptive Statistics

In this section, I present descriptive statistics for the three sectors of interest: manufacturing, professional services, and food and accommodation services. In particular, I aim to examine the differences aggregate patterns vis-à-vis labour and land inputs, and unevenness of growth across space and over time.

Table 4-4 shows the mean and median employment size and land area for businesses in each of the three sectors. To avoid overrepresenting more frequently observed firms, these statistics are calculated using only the first observation (chronologically) of each business.

Table 4-4: Employment size and land area by sector

Sector	**Employment size**		**Land area** (sq. metres)	
	Mean	*Median*	*Mean*	*Median*
Manufacturing	24.0	7	1437.5	559
Professional	12.8	4	716.7	287
Food and Acc.	15.9	8	950.5	316

(Infogroup, 2014; Massachusetts Bureau of Geographic Information, 2019)

Across the board, the mean sizes are larger than medians, suggesting that firm sizes are right-skewed in all three sectors, i.e. there are many smaller businesses and a few very large ones. Comparisons across the sectors largely yield expected outcomes. Manufacturing is the most land intensive sector, both in absolute terms and on a per-employee basis. On the other hand, professional services firms are the smallest in terms of both employment size and land area. Notably, in the food and accommodation services sector, the mean-to-median ratio is considerably larger for land (3.0) than employment size (2.0), suggesting that land is more right-skewed. This can likely be explained by the composition of the sector; food services and accommodation services, as their names imply, make use of space in very different ways.

Next, I examine employment size changes in each of the three sectors. The total observed employment size changes in the InfoGroup dataset were -1493 in manufacturing, +1911 in professional services, and -2228 in food and accommodation services. For more than three-quarters of the observations in each of the three sectors, firms neither created new jobs nor eliminated existing ones. While many of these may indeed reflect stable employment sizes, some can likely be attributed rounding and erroneous reporting during phone interviews. To address potential underreporting of employment size changes, I estimate a zero-inflated model. I elaborate on this in Chapter 4.5.

Table 4-5: Summary of job creation and destruction in manufacturing, professional services and food and accommodation services in the InfoGroup dataset

Sector	**Total change**	**% No change**	**Job creation**		**Job destruction**	
			Mean	*Median*	*Mean*	*Median*
Manufacturing	-1493	78.7%	8.8	3	-9.1	-3
Professional	+1911	76.8%	5.6	2	-4.6	-2
Food and accommodation	-2228	77.5%	5.0	3	-5.6	-3

(Infogroup, 2014)

The job creation columns in Table 4-5 show the mean and median jobs created per observation for observations with positive change. Conversely, the job destruction columns show the mean and median jobs destroyed per observation for observations with negative change.

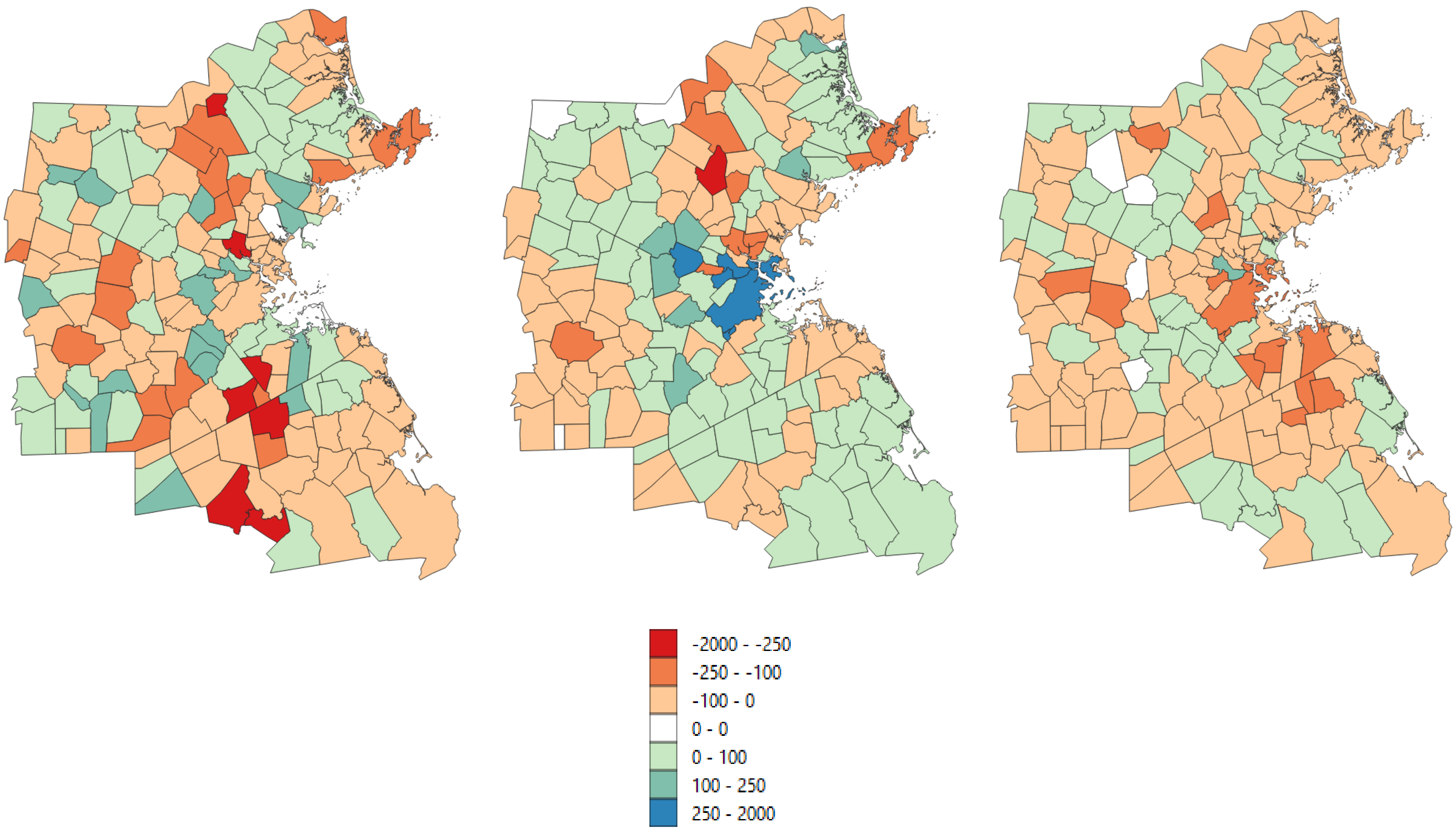
Figure 4-1 shows the unevenness of these changes over space. We can observe manufacturing decline in satellite towns and the exurbs. Discrepancies in direction of development between this map and the specialization map are a result of growth and decline in other sectors in the same areas. For professional services, growth is centered around the urban core and the immediate suburbs. Changes in food and accommodation services are more uniform with less apparent peaks and troughs. This could be evidence that agglomeration economies play a smaller role for this sector and that it instead follows overall population and employment growth patterns.

Figure 4-1: Employment size growth in towns by sector 2003-13

Manufacturing

Professional services

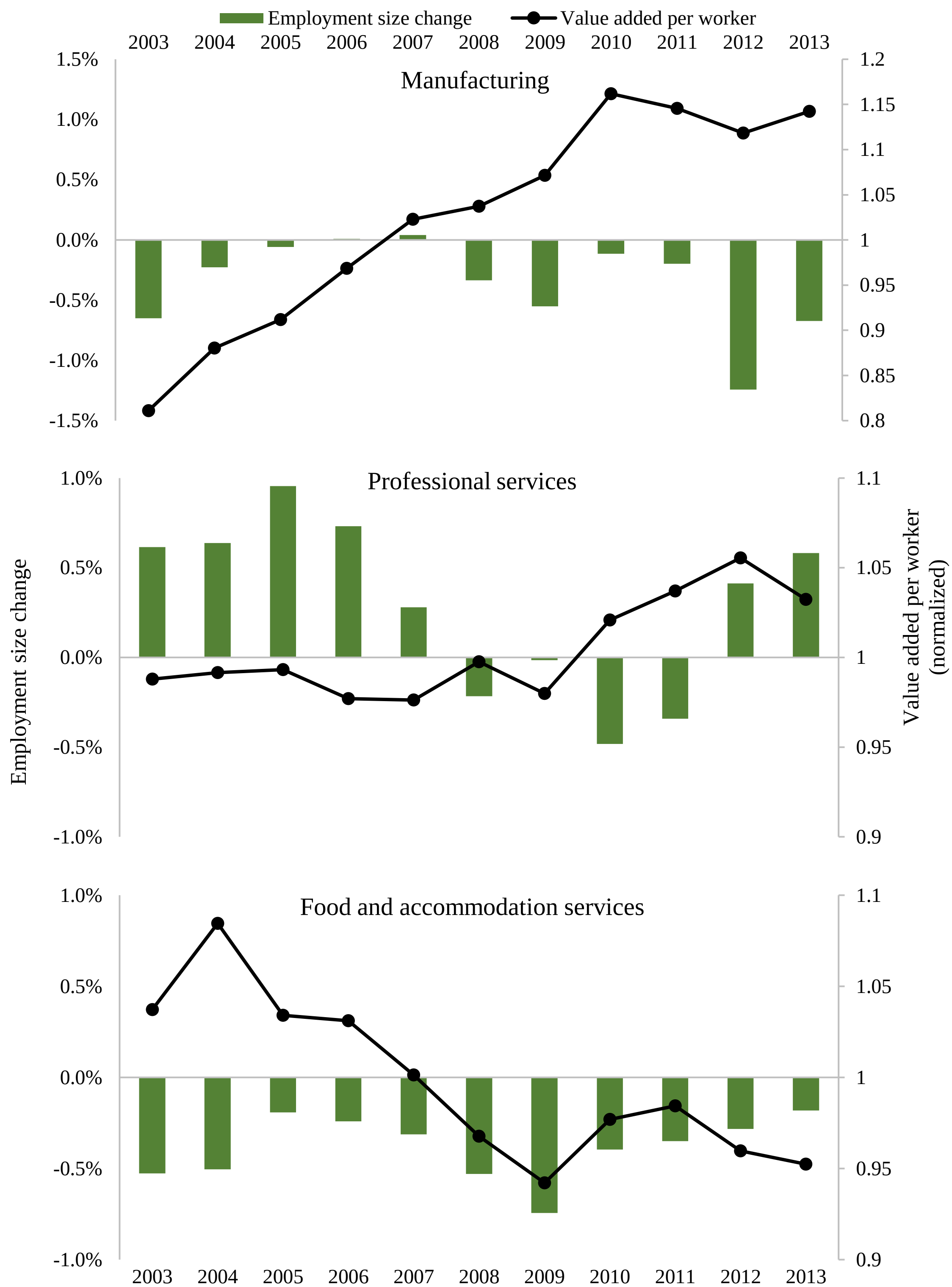
Food and accommodation services



(Metropolitan Area Planning Council, 2009)

Figure 4-2 compares annual average employment size changes to value added per worker in each sector. The latter are data for the entire U.S. and have been normalized, such that the mean for the period 2003-13 is equal to 1. Although the relationship between these two variables hardly tells complete or conclusive stories, we can glean some interesting information from the graphs. For manufacturing, employment has declined despite increased worker productivity. This trend is likely a consequence of the increased automation of the sector. On the other hand, the employment in food and accommodation services for the most part mimic the productivity trends. The 2008-09 financial crisis is also apparent in the productivity time series, albeit to varying degrees. Interestingly, the response in terms of employment changes differ in time between the sectors. For manufacturing and especially food and accommodation services the response in employment adjustments is immediate, whereas for professional services there appears to be a delay, both in eliminating jobs and in the subsequent recovery.

Figure 4-2: Annual employment size change and value added per worker by sector



(InfoGroup, 2014; U.S. Bureau of Economic Analysis, 2021a, 2021d, 2021e)

4.3 Measures of Agglomeration

My study explores two channels through which agglomeration economies affects the evolution of firms' employment sizes: (1) by increasing productivity; and (2) by reducing labour market frictions associated with searching for and hiring new employees. To quantify agglomeration, I use the gravity formulation. Popularized in transportation planning as accessibility, gravity-based accessibility fundamentally captures scale of activity and proximity. Although agglomeration economies encompass a multitude of different mechanisms, they are ultimately all enabled by a combination of scale and proximity.

Gravity-based accessibility consists of two components: the count of opportunities d_j (e.g. jobs, people, firms) in each zone j, representing the scale of activity; and a generalized cost function c_{ij}, representing the spatial friction between the zone of interest i and each zone j.

$$A_i = \sum_j d_j \exp(-\eta c_{ij}) \tag{4.1}$$

Crucially, the gravity formulation assumes a gradual attenuation of the strength of effects, rather than a binary step function defined by (often arbitrary) zone boundaries. The parameter η reflects how rapidly agglomeration effects decay over space. Thus, by varying this parameter we can examine the spatial extent of the agglomeration economies. Ideally, η should be estimated as a parameter in the Bayesian estimation procedure. However, that would require recalculating equation (1) and its more than 7.4 million ij pairs at every iteration of the Gibbs sampler. Unfortunately, this is not feasible with existing computational resources. Instead, I estimate the model repeatedly with various fixed values of η (0.25, 0.5, 1.0, 2.0, 5.0) to determine the best-fit in what is effectively a manual nested fixed-point algorithm.

For the generalized costs c_{ij}, I make use of the logsums from a commuting mode choice model. This captures the expected maximum utility (or minimum disutility) of making a trip accounting for all viable modes. Appendix D provides a description of the mode choice model, its utility functions, and estimated parameters. Finally, turning to the activity component d_j, I consider three different formulations. These are:

- **Population**: pop_j is defined as the total population in each TAZ j. This uses the population data from MAPC (Metropolitan Area Planning Council, 2009) and measures access to people at their place of residence.
- **Employment**: emp_j is defined as the total employment in each TAZ j. This uses the employment data from MAPC (Metropolitan Area Planning Council, 2009) and measures access to economic activity or workers at their place of work.
- **Specialization**: $spec_j^k$ is defined as the fraction of employment in TAZ j belonging to sector k. Thus, this is a captures concentration of activity in a single sector, rather than the scale of total like the previous two measures did.

$$spec_j^k = \frac{\widetilde{emp_j^k}}{\widetilde{emp_j}} \tag{4.2}$$

Here, I use the tilde superscript to denote that the data source for this agglomeration measure is the InfoGroup historical business records (InfoGroup, 2014). Using this dataset for calculating summary statistics has its drawbacks. Namely, it is unlikely to be a perfectly representative sample and InfoGroup does not provide any expansion factors. However, the MAPC dataset does not differentiate employment by sectors and data from the Bureau of Economic Analysis is too aggregated spatially to be of use.

In addition to varying across space, all three measures also vary over time by virtue of both the InfoGroup and MAPC data being longitudinal. Ideally, the travel time and cost data underlying the generalized cost matrix should also vary over time to reflect changes in the transportation system. However, these data were only made available as a single snapshot. While these three measures suitably describe agglomeration of activity in space, we should be under no illusion that any of them perfectly align with any single mechanism of agglomeration economies. Rather, they are simplified measures that capture partially overlapping bundles of different mechanisms. Unfortunately, fully disentangling all these underlying mechanisms remains infeasible with the available data.

Table 4-6 presents the mean normalized agglomeration measures ($\eta = 1$) of firm locations by sector. This shows the expected pattern with professional services being located with best access to population and employment, followed by food and accommodation services. The comparisons between specialization measures might at first appear counter-intuitive with manufacturing having the lowest average. However, since each measure is normalized within its own sector, the comparison shows which sector is more *centralized*. For example, if all firms in one sector are located in the same area, whereas the firms in another sector are spread out, the former will yield a much higher average normalized specialization score. In this case, professional services firms are largely centralized in the downtown core, while manufacturing firms are spread out over the region. This somewhat counter-intuitive outcome is not an issue for estimation purposes, where scores are only compared within-sector.

Table 4-6: Average agglomeration ($\eta = 1$) for firm locations by sector

	Population	Employment	Specialization
Manufacturing	0.76	0.67	0.97
Professional	1.27	1.48	1.40
Food and accommodation	1.14	1.22	1.21

The following section presents maps of the agglomeration measures. Two figures are presented for each measure, one showing the 2003 levels and showing change between 2003 and 2013. Additionally, each figure shows maps for η values of 0.25, 1.0, and 5.0. All the values have been normalized such that the mean of each measure is 1. Figure 4-3 and

Figure 4-4 show maps of population agglomeration, while Figure 4-5 and Figure 4-6 show maps of employment agglomeration. The maps show continued urbanization with growth centered around the urban core and along major highways, particularly near interchanges. In fact, drawing lines between high-growth areas would yield a decent approximation to the highway network in the Greater Boston Area. The η parameter acts as a slider between regional (small η) and local (large η) lenses. A small η reflects slow spatial decay and captures far-reaching agglomeration effects. Conversely, a large η reflects rapid spatial decay and captures agglomeration effects acting in the immediate vicinity

.

Figure 4-3: Population agglomeration by TAZ in 2003

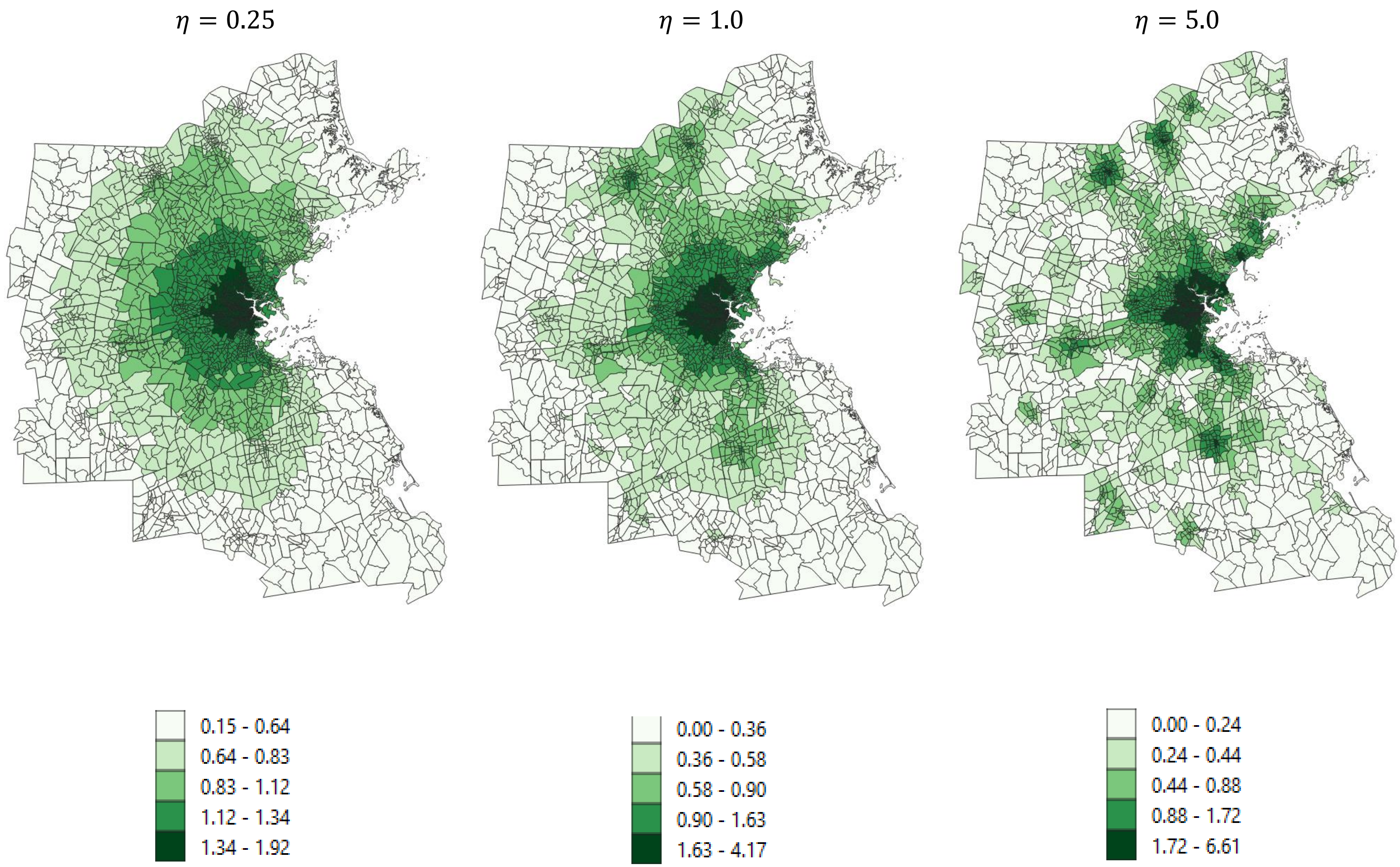


Figure 4-4: Population agglomeration change by TAZ in 2003-13

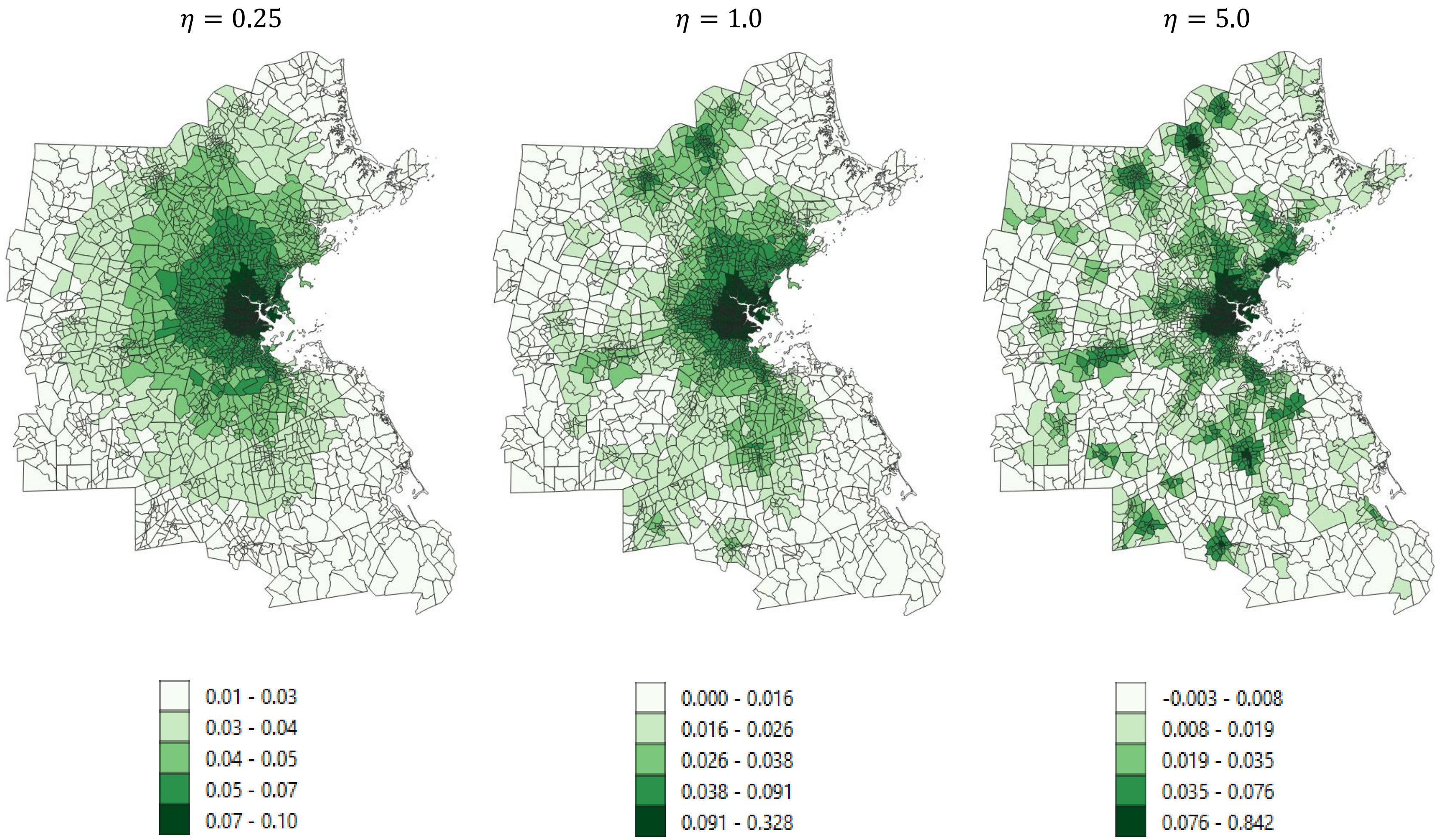


Figure 4-5: Employment agglomeration by TAZ in 2003

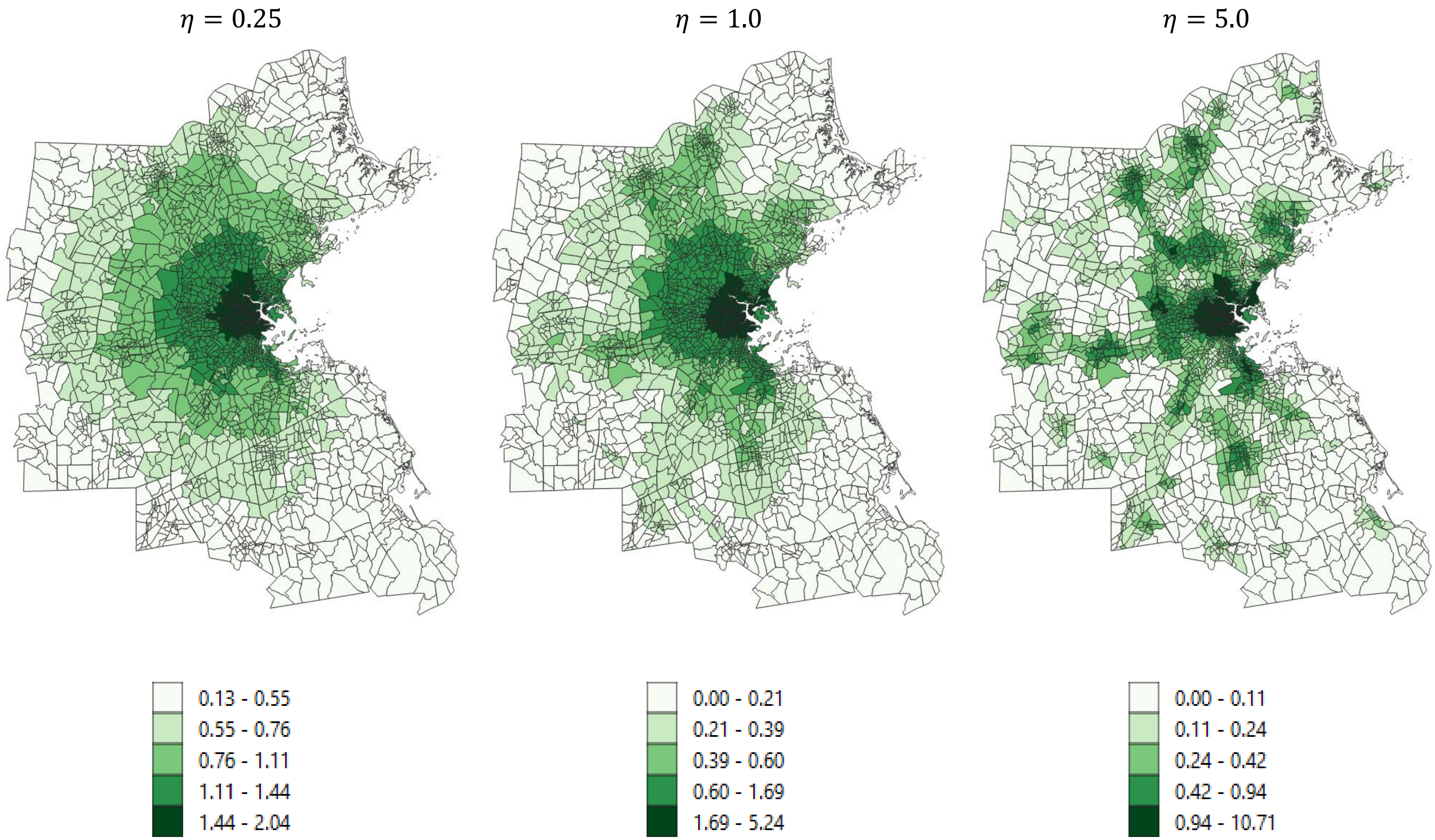


Figure 4-6: Employment agglomeration change by TAZ in 2003-13

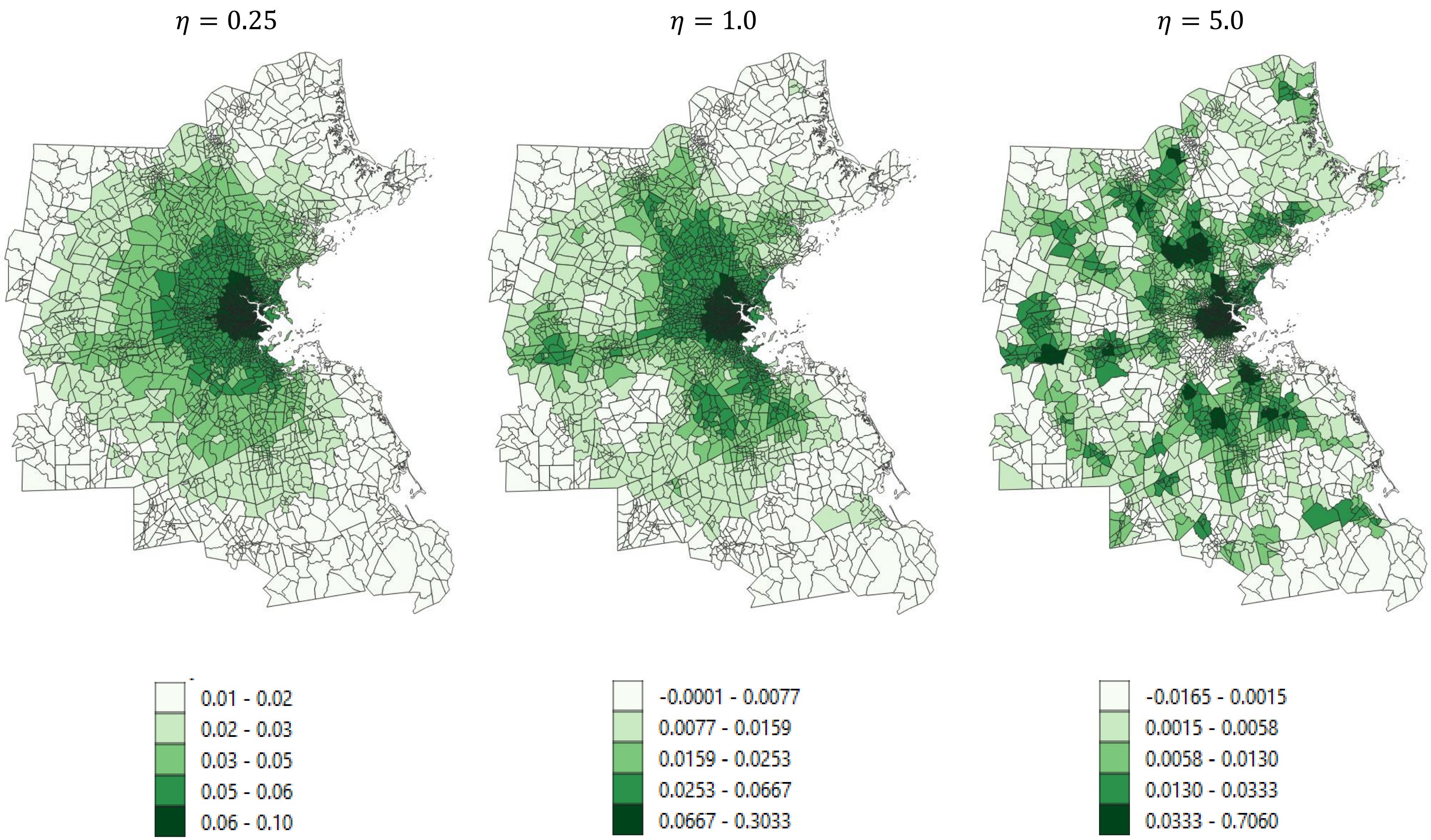


Figure 4-7 through Figure 4-12 present maps of specialization in each of the three sectors of interest. For manufacturing, there are several clusters ($\eta = 5.0$) in satellite towns outside the urban core with the regional centre ($\eta = 0.25$) located north of the city proper. Specialization in manufacturing increased in the aforementioned satellite towns but decreased in the core. Professional services are, as expected, centered around the urban core, with the growth in the sector continuing that trend. However, there also appears to be a professional services cluster forming in the Taunton area south of Boston. The pattern of development for food and accommodation appears like a combination of the two other sectors. Specialization is highest in the core but its distribution is more homogenous across the region than professional services.

Figure 4-7: Manufacturing specialization by TAZ in 2003

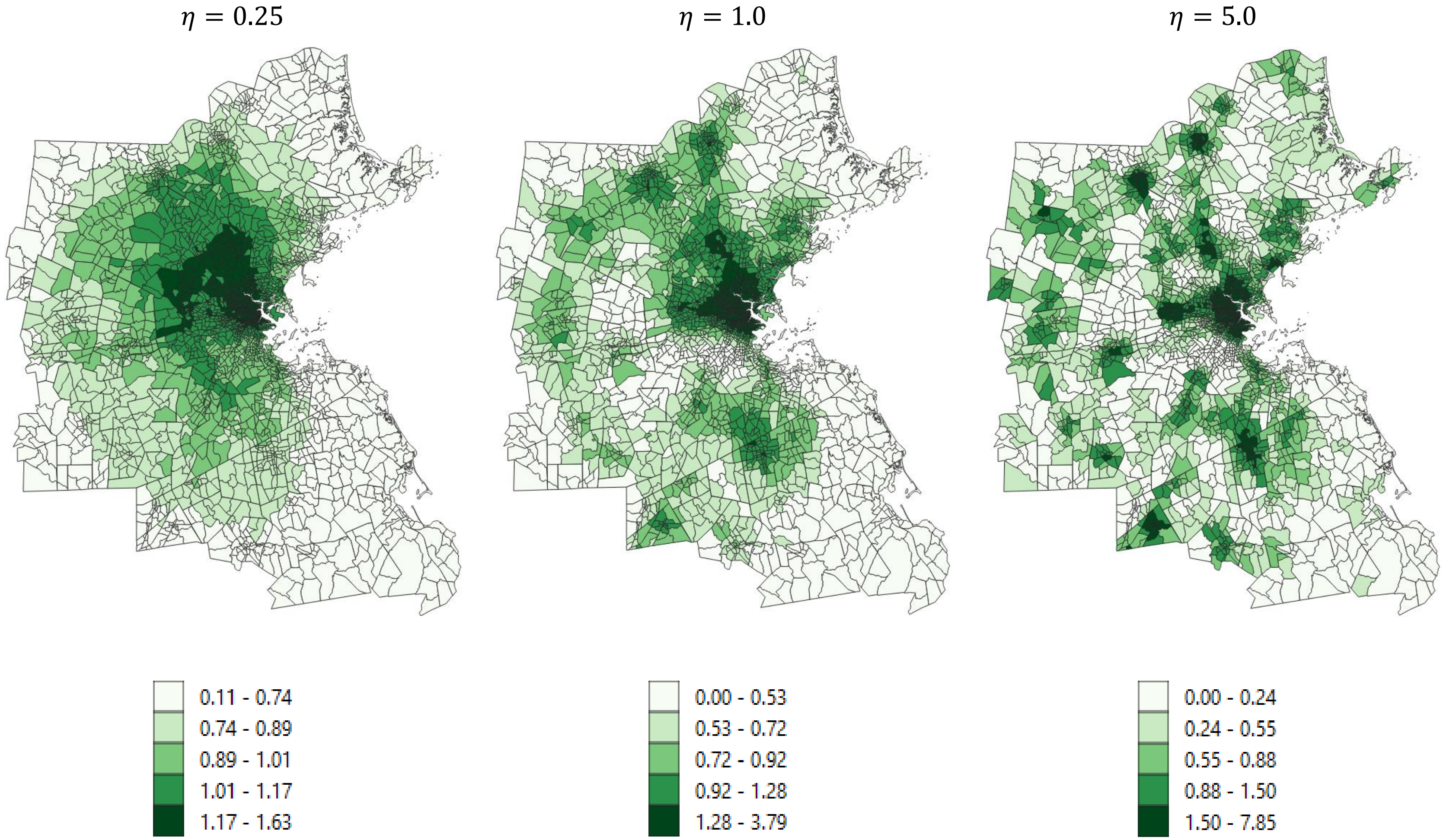


Figure 4-8: Manufacturing specialization change by TAZ in 2003-13

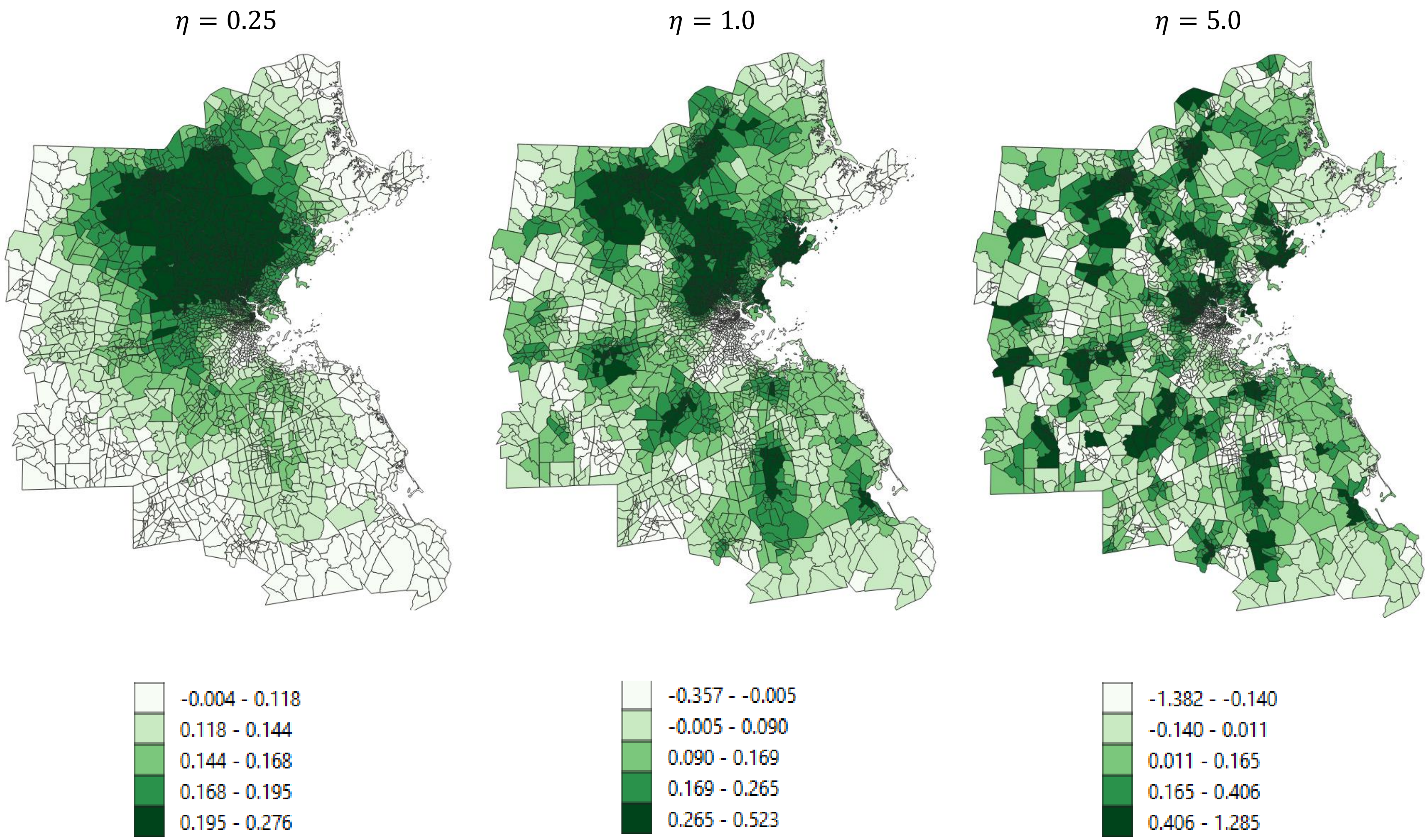


Figure 4-9: Professional services specialization by TAZ in 2003

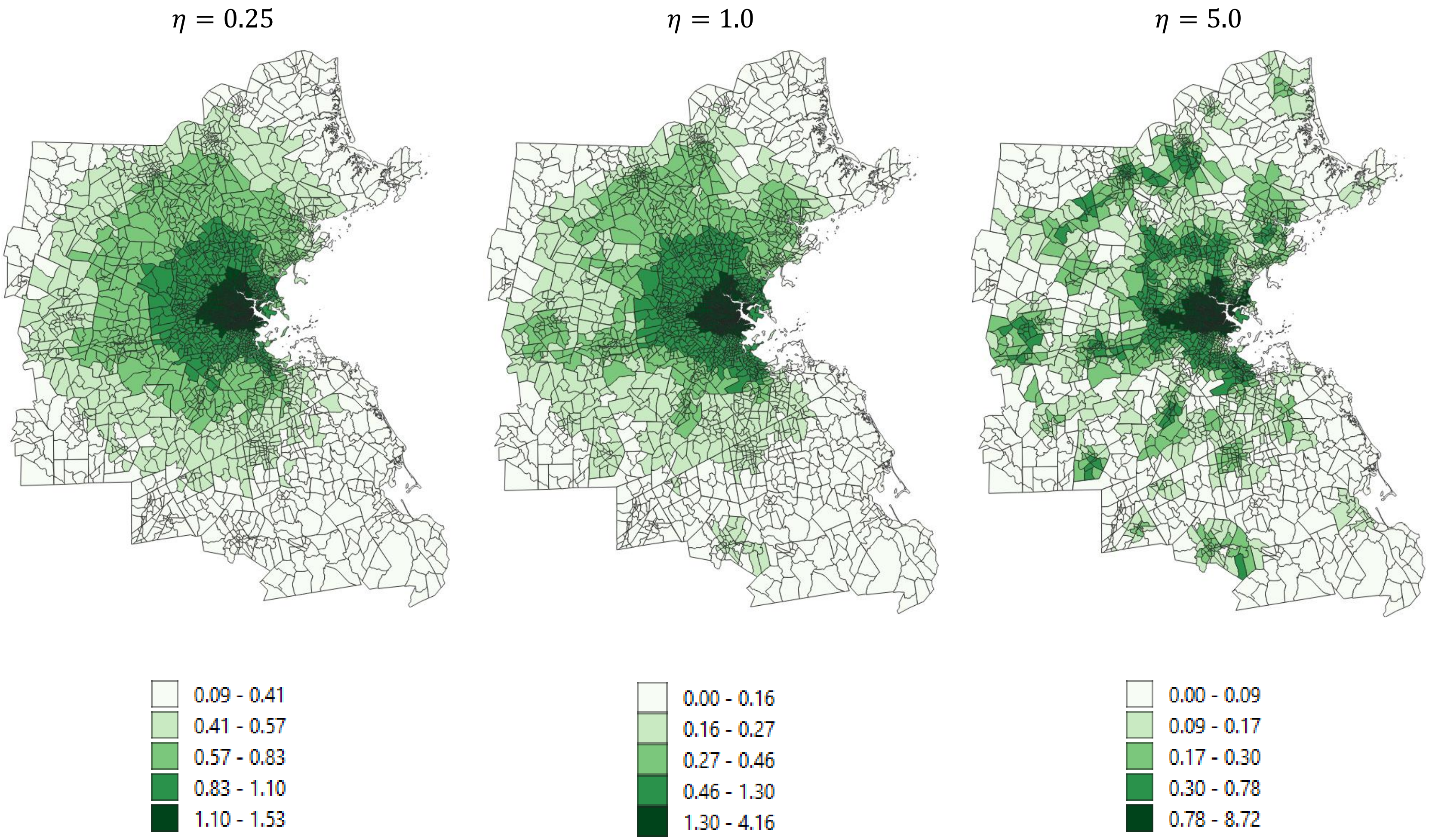


Figure 4-10: Professional services specialization change by TAZ in 2003-13

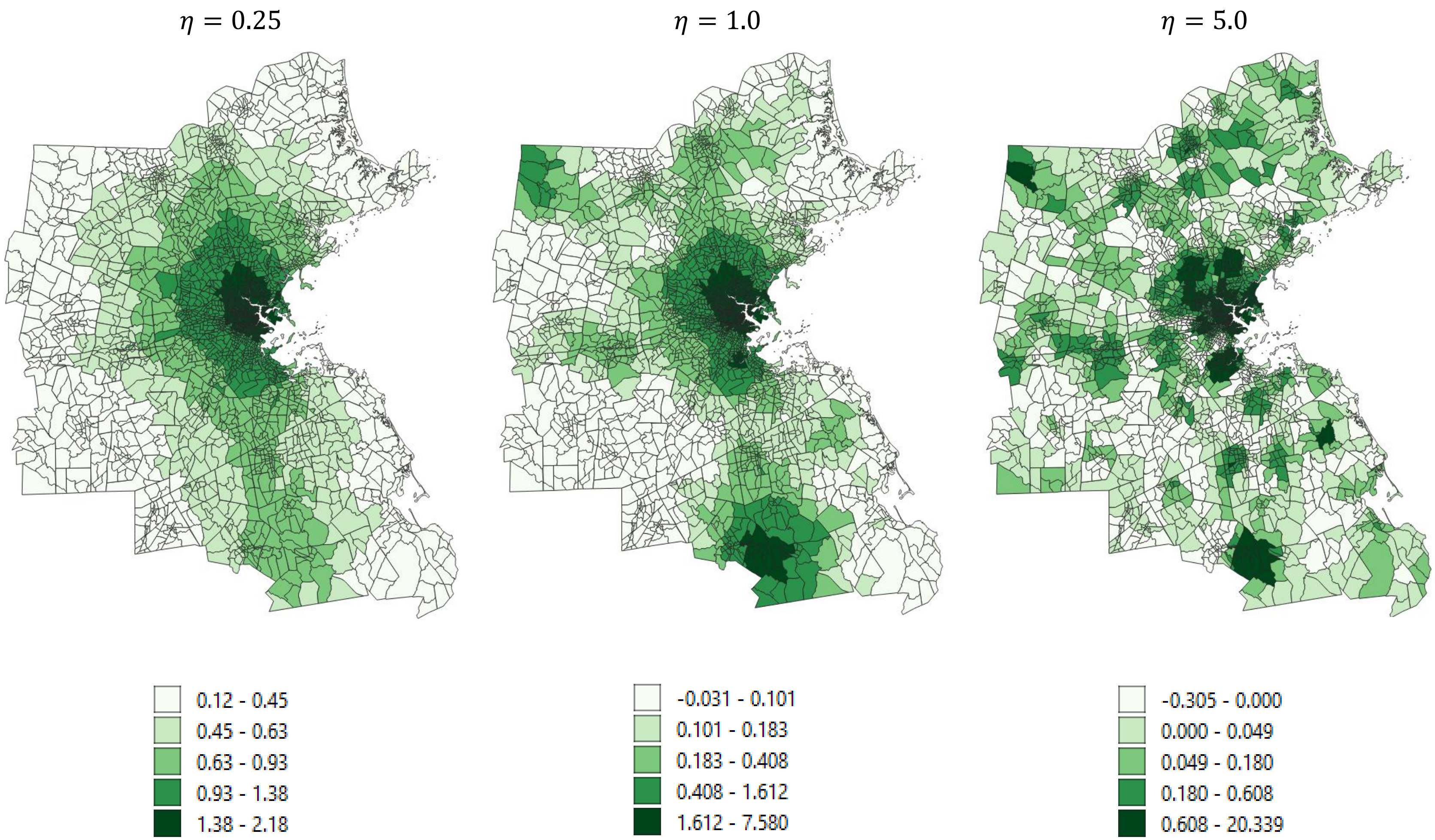


Figure 4-11: Food and accommodation services specialization by TAZ in 2003

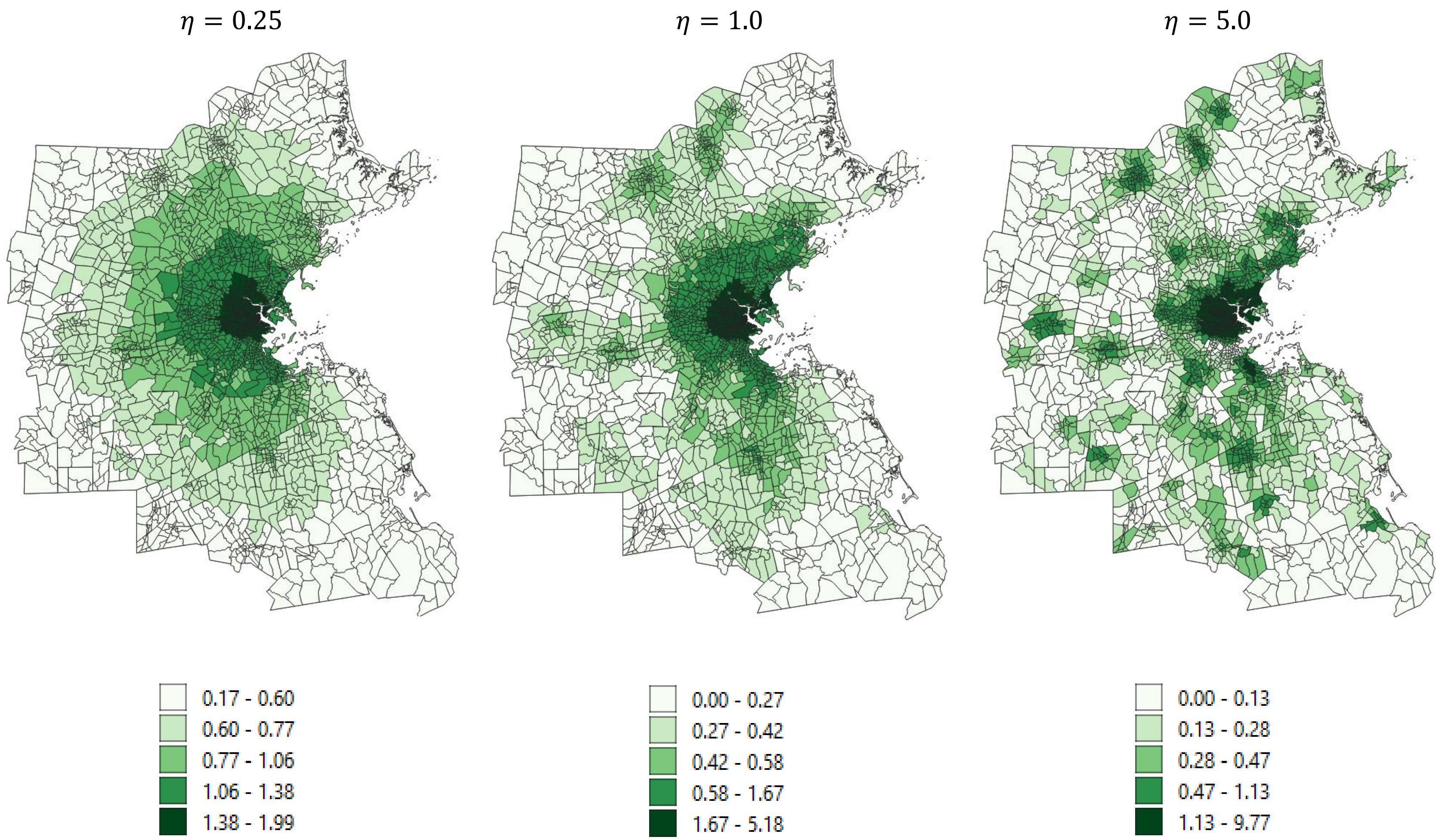
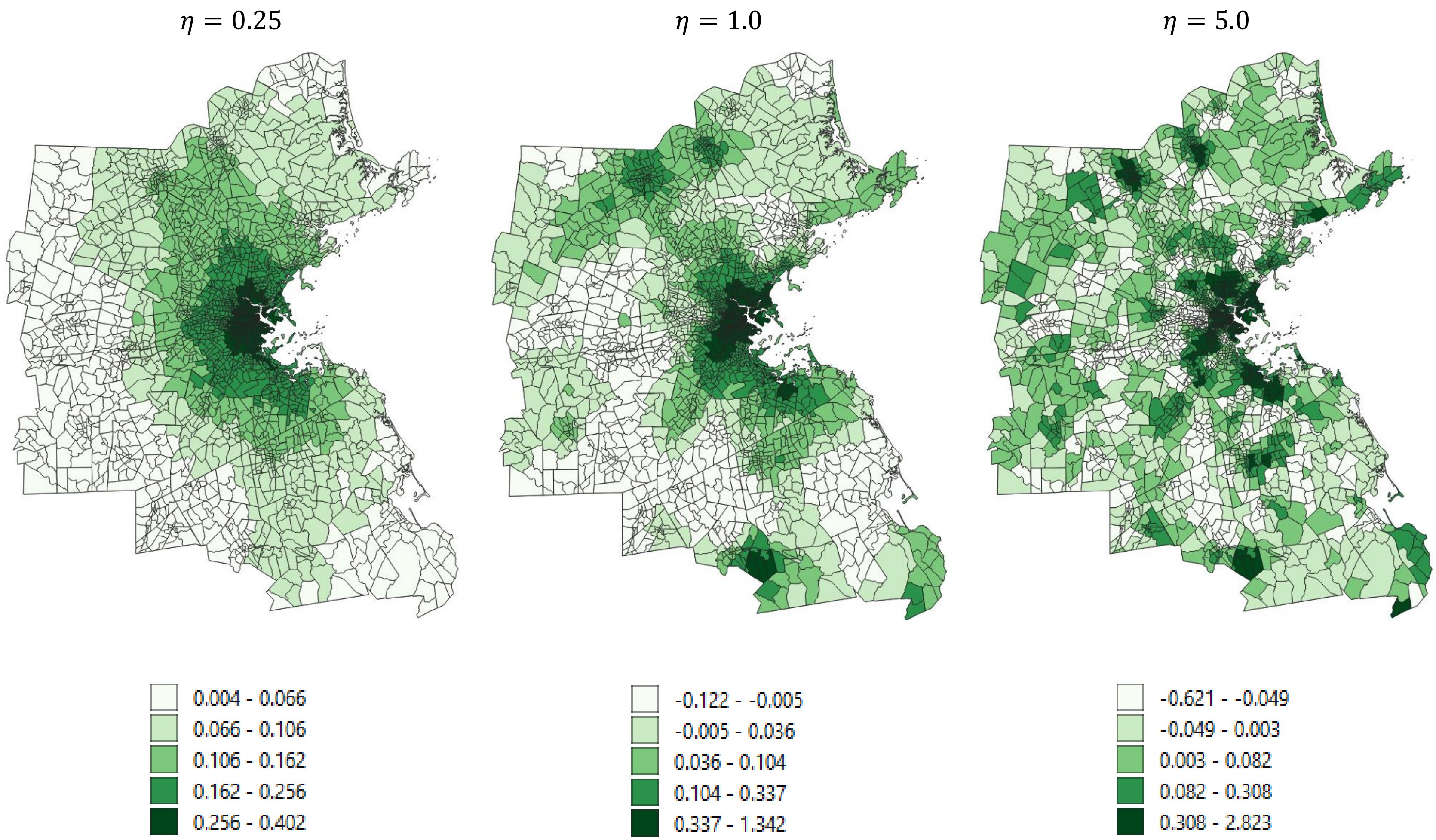


Figure 4-12: Food and accommodation services specialization change by TAZ in 2003-13



4.4 From Mechanisms to Measures

In this section, I link the mechanisms of agglomeration economies to the measures formulated in Chapter 0. Specifically, I formulate hypotheses for which mechanisms I believe are relevant for each sector based on the literature review in Chapter 2.2 and then determine which measure best captures those mechanisms. This is summarized by Table 4-7 below. The following provides brief descriptions of the main mechanisms at play. The review of mechanisms and micro-foundations of agglomeration economies in Chapter 2.2 elaborates on these.

- **Input/output sharing:** The sharing of intermediate suppliers and buyers enable businesses to operate at larger scale. In turn, this allows for efficient use of indivisible equipment and facilities. Additionally, clustering also reduces the transport costs of physical goods. In this study, these mechanisms are primarily relevant for the manufacturing sector.
- **Learning**: Learning mechanisms encompass innovation, spill-overs of existing production technologies, and transfer of tacit knowledge through face-to-face interactions, and can occur both within sectors (localization economies) and across sectors (urbanization economies). These mechanisms become more pertinent the more knowledge-intensive a sector is. Thus, they apply to professional services and, to a lesser extent, manufacturing.
- **Match quality:** A thick labour market improves the average quality of matching between firms and workers, making better use of workers' skills. Matching is most applicable to skills-intensive sectors, again, professional services and, to a lesser extent, manufacturing.
- **Labour pool:** Access to a larger labour pool allows businesses to find workers with a suitable skillset more quickly. Population and specialization capture different aspects of the labour pool, namely size and appropriateness of the skillsets available. The latter is likely more important to skill-intensive sectors, i.e. professional services and manufacturing.
- **Competition and market potential:** Competition and market potential have been included in the table even though they do not affect productivity directly. Instead they

affect prices and in turn revenue through supply and demand. However, I do not have adequately detailed price data to control for local price variations. Thus, the estimated coefficients of the agglomeration measures will likely also capture these market effects. Incorporating a spatial input-output model that determines local price variations from make/use tables is a potential solution to address this. However, the spatial granularity of these models is typically at the regional scale, whereas I examine the effects agglomeration within a metropolitan area.

The expectations for the spatial extent of each mechanism depend primarily on who the interacting agents are. For match quality and frequency, which are interactions between employers and employees at the job search stage, it seems sensible to measure agglomeration at a spatial scale similar to people's willingness to commute. I denote this as "metropolitan" in Table 4-7. Learning mechanisms are usually associated with interactions between employees at or near their workplace, e.g. in an office complex. Hence, the expected spatial extent is relatively small. I denote this as "local". Finally, input/output sharing, presence of competition, and access to markets are all interactions between firms and their suppliers and buyers. Thus, the spatial extent of these mechanisms depends on the nature of the trade linkages. For example, a manufacturing firm likely sources its intermediate inputs regionally and beyond, whereas the market area of a typical restaurant is far more local. For I/O sharing, which is primarily relevant for manufacturing firms, my hypothesis is that the spatial extent is "regional". For market potential and competition, the spatial extent varies by sector but are likely decreasing in the following order: manufacturing, professional services, then food and accommodation services.

Table 4-8 translates Table 4-7, such that rows represent agglomeration measures. It summarizes, by adding together all the +'s and –'s, the total expected effect of all mechanisms captured by each measure, allowing me to compare the a priori hypotheses with empirical findings.

Table 4-7: A priori expectations of how the mechanisms interacts with each sector, which channel they act through, their spatial extent, and which measure they are captured by

Mechanism	**Manufacturing**	**Professional**	**Food and accommodation**	**Spatial extent**	**Agglomeration measure**
I/O sharing	+++			Regional	Specialization
Learning	+	+++		Local	Employment Specialization
Match quality	+	++		Metropolitan	Population Specialization
[Match frequency]	[++]	[+++]	[+]	Metropolitan	Population Specialization
Competition	-	-	- - -	Varies (see text)	Specialization
Market potential	+	+	+++	Varies (see text)	Population Employment Specialization

[] indicate mechanisms acting through the labour market friction channel as opposed to productivity

Table 4-8: Hypothesized relative magnitudes of agglomeration measures

Agglomeration measure	**Spatial extent**	**Mechanisms**	**Manufacturing**	**Professional**	**Food and accommodation**
Population	Local	Market potential[3]			+++
	Metropolitan	Match quality[123] [Match frequency][123]	+ [++]	++ [+++]	[+]
Employment	Local	Learning[2] Market potential[3]		+++	+++
Specialization	Local	Learning[12] Competition[3]	+	+++	- - -
	Metropolitan	Match quality[12] [Match frequency][123] Competition[2]	+ [++]	+ [+++]	[+]
	Regional	I/O sharing[3] Competition[3] Market potential[3]	+++		

[] indicate mechanisms acting through the labour market friction channel as opposed to productivity

[1] applies to manufacturing

[2] applies to professional services

[3] applies to food and accommodation services

4.5 Model Specification

Table 4-9 presents the exogeneous variables used in the model specification that were not already discussed in the agglomeration measures section. These include employment size, land, wages, and various control variables.

Table 4-9: Variable descriptions

Variable	**Symbol**	**Description**
Employment size	H	Number of workers employed by firm
Land	L	Floorspace used by firm
Wages	ω	Average compensation by sector and year in the Boston-Cambridge-Newton MSA
GDP	γ_1	GDP by sector and year in the Boston-Cambridge-Newton MSA
Productivity	γ_2	Value added per employee by sector and year in the US
Education	γ_3	Percent of workers with college degree or above (for professional services and manufacturing) Percent of workers and residents with college degree or above (for food and accommodation services)

Profit $\bar{\pi}$ is the difference between revenue and costs. I model firm production with a Cobb-Douglas production function with employment size H and land L inputs and total factor productivity τ. Revenue is given by the product between output prices p and the firm's production.

$$\begin{aligned} revenue &= p\ \tau H^{\beta_0} L^{\beta_1} \\ \tau &= \beta_2\ \gamma_1^{\beta_3} \gamma_2^{\beta_4} \gamma_3^{\beta_5}\ A_1^{\beta_6} A_2^{\beta_7} \ldots \end{aligned} \tag{4.3}$$

β_0 and β_1 are the elasticity parameters associated with employment and land, respectively. We normally expect that $0 < \beta_0, \beta_1 < 1$, such that production is concave with respect to each input. The A's denote relevant agglomeration measures based on Table 4-8. Unfortunately, accurate information about the prices of the goods and services each firm produces is not available. Instead, I assume that firms are price-takers and that they can sell their entire production at a fixed market price. In estimating the model, this value will be captured by β_3, the constant term in the total factor productivity.

For production costs, only those that are a function of employment size are relevant for our purpose.

$$cost = \omega H + k \tag{4.4}$$

Constant cost terms k get cancelled out since we only consider differences in profit at various employment sizes. It is for this reason that the error term has also been specified as a function of the employment size H. The firm's profit π is the difference between its revenue and costs.

$$\begin{gathered} \bar{\pi} = p\tau H^{\beta_0} L^{\beta_1} - \omega H + k \\ \pi = p\tau H^{\beta_0} L^{\beta_1} - \omega H + k + H\varepsilon \end{gathered} \tag{4.5}$$

Section 3.2 discussed the formulation of adjustment costs and their different impacts on firm behaviour. However, that discussion did not consider agglomeration economies. The literature on thick market effects suggest that having access to a larger labour pool enables quicker matching between employers and employees, all else equal. Thus, I include agglomeration measures capturing labour pool effects in the adjustment costs specification – see Table 4-8. Specifically, I include it in the superlinear term in hiring costs, since this most directly governs adjustment rates.

$$\begin{gathered} \alpha(\lambda_H) = \begin{cases} \mu_0^+ + \mu_1^+ \lambda_H + \mu_2^+ A_3^{\mu_3} A_4^{\mu_4} \cdots \lambda_H^2 & if\ hiring \\ \mu_0^- + \mu_1^- \lambda_H + \mu_2^- \lambda_H^2 & if\ firing \\ 0 & otherwise \end{cases} \\ where\ \mu_0^+, \mu_1^+, \mu_2^+, \mu_0^-, \mu_1^-, \mu_2^- > 0 \end{gathered} \tag{4.6}$$

To address potential “excess zeroes” in the dataset, i.e. observations where firms erroneously reported that no employment size adjustment were made, I estimate a zero-inflated model. This approach introduces and mixes-in an additional zero-generating process. For this purpose, I use a binary logit.

$$\Pr(excess\ zero) = Z = \frac{1}{1 + e^{-\zeta}} \tag{4.7}$$

where ζ is a parameter to estimate. Now, we might observe zeroes as a result of the original data-generating process or this new zero-generating process.

$$\begin{aligned} \Pr(\Delta H_{it} = 0) &= Z + (1 - Z) \cdot P(0|\theta_i, \varepsilon_{it}) \\ \Pr(\Delta H_{it} > 0) &= (1 - Z) \cdot P(\Delta H_{it}|\theta_i, \varepsilon_{it}) \end{aligned} \tag{4.8}$$

4.6 Results and Discussion

Using Bayesian estimation procedure discussed in Chapter 3.3, I estimate separate models for each of the three sectors, manufacturing, professional services, and food and accommodation services. Table 4-10 presents estimation results for non-agglomeration variables. These variables are used in the models for all three sectors. The only exception is worker productivity for manufacturing, which is highly correlated with the GDP variable for that sector. Table 4-11, Table 4-12, and Table 4-13 below present the coefficients associated with agglomeration measures for each sector. To determine which measures best capture the relevant mechanisms of agglomeration economies for each sector, I use Table 4-8 as starting point. Then, I test alternatives and compare based on how well the estimated coefficient means and standard deviations align with the a priori hypotheses. I also consider overall model goodness-of-fit as measured by the mean of the posterior predictive distribution (PPD). However, this metric is dominated by the zero-inflation component of the model and thus not especially informative. Unlike traditional least squares or maximum likelihood-based approaches, results are presented as the mean and standard deviations of the parameter distributions. These make up the hyperparameters Θ and W.

For ease of interpretation, the models have been specified in units of dollars and years. That is, for unit inputs and absent the effects of other factors, the average employee in manufacturing produces 127 thousand dollars worth per year. Of course, the TFP constants should not be assigned too much meaning, since they consist of an amalgamation of unobserved factors and unit inputs are not necessarily meaningful. Nonetheless, it is unsurprisingly the largest for professional services, followed by manufacturing then food and accommodation services. Examining the elasticities of the production inputs, we find that they are all less than 1, as expected, indicating diminishing marginal returns. Their relative magnitudes also seem largely intuitive: manufacturing firms scale better with land, whereas professional services scale better with human capital inputs. For food and accommodation services, the elasticity with respect to land is unexpectedly low. This is likely an artefact of the sector definitions. Specifically, restaurants and hotels make use of land very differently, thus attenuating the coefficient. If the data permit, separate models could be considered for the food services and accommodation services.

Table 4-10: Estimation results, means and (standard deviations)

Variable	Manufacturing	Professional	Food & Accommodation
TFP constant β_2	126.77 (29.47)	159.74 (34.58)	84.85 (7.71)
Employment size β_0	0.25 (0.07)	0.34 (0.07)	0.29 (0.02)
Land β_1	0.28 (0.08)	0.23 (0.07)	0.18 (0.07)
GDP β_3	-0.24 (2.45)	-0.09 (1.82)	0.89 (0.36)
Productivity β_4		0.22 (0.29)	0.35 (1.19)
Education β_5	0.64 (1.68)	0.28 (0.58)	0.25 (0.22)
Fixed adj. cost, hiring μ_0^+	4.04 (1.00)	9.71 (1.18)	5.85 (0.28)
Fixed adj. cost, firing μ_0^-	5.24 (1.39)	7.26 (0.68)	7.01 (0.36)
Linear adj. cost, hiring μ_1^+	9.47 (1.80)	11.87 (1.39)	5.82 (1.37)
Linear adj. cost, firing μ_1^-	27.60 (8.09)	17.26 (1.78)	10.88 (1.09)
Quadratic adj. cost, hiring μ_2^+	3.50 (1.25)	5.86 (1.83)	2.38 (0.43)
Quadratic adj. cost, firing μ_2^-	5.47 (2.15)	9.44 (3.24)	4.88 (1.20)
Zero-inflation ζ	1.19 (0.77)	1.01 (0.84)	1.16 (0.87)
Error st. dev. σ	14.94	20.13	7.22
Mean PPD	0.67	0.66	0.67

Surprisingly, GDP and productivity have little bearing on the employment size of firms outside the food and accommodation services sector. The aggregate nature of the measures mean they may not be representative of the behaviour of individual firms. The controls for education have positive coefficients as expected. However, their magnitudes are surprisingly larger for the sectors typically less associated knowledge intensive work and the estimates are associated with considerable variance. For the manufacturing sector, this could reflect the types of manufacturing conducted in different areas, e.g. microchip manufacturing is very different from steel manufacturing in terms of technology and factor of production. In any case, the education variables only vary cross-sectionally, not longitudinally, and are thus easily biased by unobserved local fixed effects.

As discussed in Chapter 3, the adjustment cost function encapsulates a myriad of different costs associated with hiring and firing employees, e.g. hiring bonuses, severance pay, training, reorganization, but also union rules and local legislation. However, in total, the coefficients suggest that adjustment costs increase with level of knowledge and skill-intensity of the sector. In particular, the professional services are associated with the largest coefficients for all but the linear firing costs. Adjustment costs are generally higher for manufacturing than for food and accommodation services. However, interestingly, the fixed cost of hiring is higher for the latter. Overall, the patterns align with the a priori expectation that adjustments, and in particular hiring, is more costly for higher skilled sectors as they have narrow skills requirements and are more likely to invest in worker training. Interpreting the coefficients for firing is less straightforward. Beyond simply adjustment costs, the coefficient could also be measures of how agile a firm is or the stringency of local labour market legislation.

The zero-inflation parameters suggest that a large portion of the dataset comprise excess zeroes. This could be indicative of the poor data or poor adjustment cost specification – or both. The error terms represent average unexplained productivity per worker. Thus, the standard deviations of 15, 20, and 7 thousand dollars per year indicate that considerable variation remains unexplained.

Next, we examine the coefficients associated with agglomeration measures and compare them to the a priori hypotheses from Table 4-8. Note that for coefficients associated with match frequency, more [+]'s represent a larger hypothesized magnitude. However, this translates into a more negative coefficient since we expect the effect to be a friction reduction. For manufacturing, we find that local specialization considerably increases productivity (0.10). This potentially suggests that the improved quality of matches between workers and firms arise more locally and rely more on being in a specialized labour pool than anticipated. On other hand, regional specialization was only found to have a minute impact on productivity (0.01). While agglomeration of people does not appear to have a major impact on manufacturing firms' hiring costs (-0.01), local specialization considerably reduces these (-0.34). It is worth pointing out that while the coefficient is several times larger than typical estimates for agglomeration economies found in the literature, the magnitude of the coefficient by itself is not a cause for concern. The coefficient estimated here captures the extent to which specialization allow firms to adjust their employment size faster, whereas the typical ranges discussed in the literature review, Chapter 2.2, are for wage or productivity increases. As such these cannot be compared directly. Moreover, the spatial resolution of the measures used here is finer than what is commonly found in existing studies of agglomeration economies.

Table 4-11: Estimation results, manufacturing agglomeration effects, means and (standard deviations)

Agglomeration measure (η)	**Parameter estimates**	**Mechanisms**	**Hypothesized magnitude**
Specialization (walk only, 2.0)	0.10 (0.05)	Learning Match quality*	++
Specialization (0.5)	0.01 (0.03)	I/O sharing Competition Market potential	+++
Population (2.0)	-0.01 (0.04)	[Match frequency]	[++]
Specialization (walk only, 2.0)	-0.34 (0.16)	[Match frequency]*	[++]

[] indicate mechanisms acting through the labour market friction channel as opposed to productivity
* indicate deviations from the hypothesis in terms of spatial extent

For professional services, agglomeration of people was not found to have a beneficial effect of firm productivity (-0.05) regardless of the spatial extent of the measure. However, this estimate likely also captures some of the effects associated with competition as the measure was highly correlated with that of local specialization, thus only allowing for the inclusion of one. As expected, the agglomeration of jobs has a positive effect (0.10) on the productivity of professional services firms. For reducing labour market frictions, having access to a large regional employment pool is an advantage (-0.08), so is being located in a specialized local cluster (-0.06), albeit to a lesser extent.

Table 4-12: Estimation results, professional services agglomeration effects, means and (standard deviations)

Agglomeration measure (η)	**Parameter estimates**	**Mechanisms**	**Hypothesized magnitude**
Population (0.25)	-0.05 (0.06)	Match quality* Competition*	++
Employment (5.0)	0.10 (0.05)	Learning Market potential*	+++
Population (0.25)	-0.08 (0.06)	[Match frequency]*	[+++]
Specialization (5.0)	-0.06 (0.07)	[Match frequency]*	[+++]

[] indicate mechanisms acting through the labour market friction channel as opposed to productivity
* indicate deviations from the hypothesis in terms of spatial extent

For food and accommodation services, agglomeration of people (0.05) and jobs (0.05) are both beneficial to their market potential, although the spatial extent of the residential market area is larger than expected. Local specialization, and by extension, competition appears to have an overall negative effect (-0.05) but is associated with considerable variance. Finally, while hypothesized that a larger labour pool also allows restaurants and hotels to find employees faster, this effect appears minimal (-0.01).

Table 4-13: Estimation results, food and accommodation services agglomeration effects, means and (standard deviations)

Agglomeration measure (η)	Parameter estimates	Mechanisms	Hypothesized magnitude
Population (1.0)	0.05 (0.02)	Market potential*	+++
Employment (walk only, 5.0)	0.05 (0.03)	Market potential	+++
Specialization (walk only, 5.0)	-0.05 (0.18)	Competition	- - -
Population (1.0)	-0.01 (0.05)	[Match frequency]	[+]

[] indicate mechanisms acting through the labour market friction channel as opposed to productivity
* indicate deviations from the hypothesis in terms of spatial extent

Challenges and limitations

The computational burden of estimating these models is quite significant. Each model estimation run requires between several days and a few weeks to complete on a typical personal computer[2], depending on the starting values. From a computational point-of-view, three obstacles combine to prevent the estimation from running faster: (1) the draws of the Bayesian estimation procedure are fundamentally sequential, severely limiting the effectiveness of speed-ups through parallel processing; (2) although the dynamic programming problem is relatively simple, it must be solved for each firm and each time period in each iteration of the Bayesian estimation procedure; and (3) similarly, the likelihood function is evaluated firm-wise rather than simultaneously for the whole dataset because the nested products and sums of varying lengths in the likelihood

[2] Intel(R) Core(TM) i7-8650U CPU @ 1.90GHz; 8.00 GB RAM.

function do not lend themselves to efficient matrix algebra. Although model estimation, unlike simulation, is typically considered an upfront one-time cost, the sluggishness has made testing alternative specifications extremely cumbersome.

The complex model structure results in some parameters converging relatively slowly, further exacerbating the long estimation times. In these cases, choosing the starting value close to the true value is particularly important. Appendix E presents convergence diagnostics, including Heidelberger-Welch p-values and Gelman-Rubin $\hat{R}$ statistics. Several factors potentially contribute to the slow convergence. There is some evidence of poor mixing, which typically arises if the likelihood surface is very uneven or poorly connected. Potential strategies to address this include using a more advanced sampler, for example ones that use simulated or parallel tempering. Making different distributional assumptions could potentially also help. Another potential cause of the convergence issues is poorly identified parameters. The highly non-linear model structure means that parameters can interact in potentially unpredictable ways that usual linear-in-parameters models do not. Of particular concern is the simultaneous identification of the fixed adjustment cost and the zero-inflation parameter. Both serve the function of allowing more "zero" observation, albeit one as a hurdle and the other as a separate data-generating process. However, in testing, removing the zero-inflation model component yielded clearly unreasonable values for the fixed adjustment cost. This suggests that including the additional zero-generating process is appropriate. On the other hand, omitting the fixed cost largely eliminates the hurdle effect, and with it the "*long periods of constancy broken by infrequent large jumps"* that is characteristic of firms' employment size adjustment at the micro-level (King and Thomas, 2006, p.782).

When estimating agglomeration effects, we must consider potential endogeneity issues at both the firm and local levels. Endogeneity at the firm level arises when unobserved firm-specific effects are correlated with agglomeration variables. This typically occurs when firms sort spatially according to characteristics not controlled for by the model, e.g. unobserved advantages in firms' production technology. Although no panacea, the use of panel data is crucial for dealing with potential endogeneity at the firm level. In particular, introducing time-invariant firm-specific effects, e.g. the TFP constant, controls for unobserved firm characteristics – at least those that remain constant over time. Endogeneity at the local level arises when the agglomeration variables are correlated with unobserved local effects. This typically happens if a variable that affects both local agglomeration and firm profits is missing from the model. For example, an airport might attract firms to an area and simultaneously increase worker productivity and by extension employment size, thus increasing agglomeration and the measured outcome at the same time. In this case, the estimated agglomeration effect would likely be upwards biased. Unfortunately, the model is less well-equipped to handle potential local level endogeneity. Estimating location-specific fixed effects for each of the 2728 TAZs is not feasible given data and computational constraints. Furthermore, insofar as prices and wages vary locally, the lack of spatially disaggregated price and wage data might also contribute to local level endogeneity. While including agglomeration variables in the cost term could address the issue, in practice it leads to further identification challenges.

The InfoGroup historical micro-level business records do not have observations at regular intervals and thus make for an unbalanced panel. This is potentially problematic because of attrition / survivorship bias. In other words, we do not observe firms when they go out of business, thus the dataset is not representative of all firms but rather the subset that *survived*. The interaction of this bias with the agglomeration variables specifically is not obvious. However, the TFP of surviving firms is likely higher than that of those that closed.

At its best, structural estimation is a powerful tool for quantifying structural – or so-called policy invariant – parameters. However, structural approaches are certainly not without their limitations and also bring with them additional drawbacks compared to traditional reduced-form approaches. In particular, the flexibility and believability of any such model are tied to its structural and parametric assumptions, such as those discussed in Chapter 3, and communicating and interpreting the model results is considerably more involved. Additionally, the data-hungriness of agent-based models is only exacerbated by the introduction of dynamics in the structural approach. Nonetheless, the ever-increasing computational power and continued gathering of microscopic datasets make addressing most these limitations feasible – if not now – in the future.

5 IMPACT ANALYSES

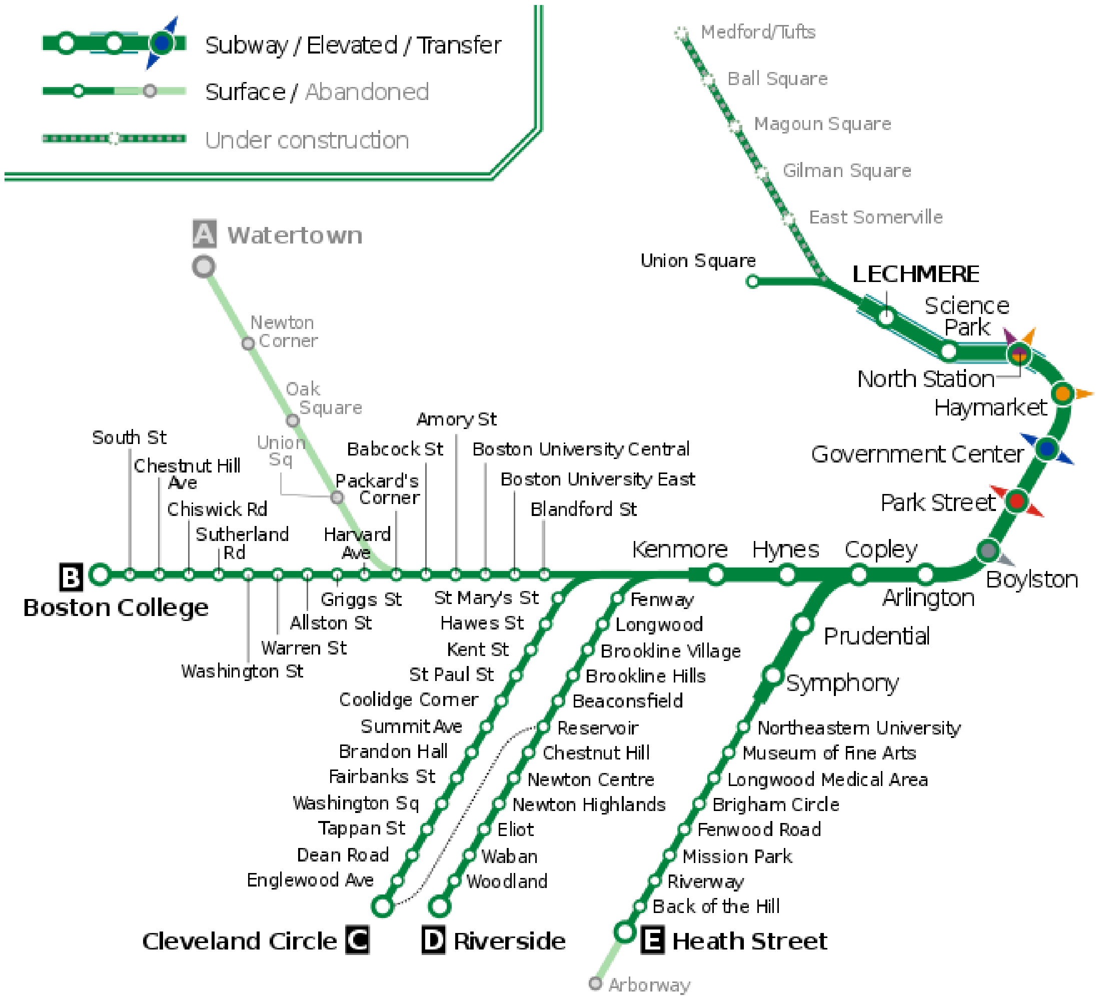
To demonstrate the usefulness of the proposed approach, I conduct two impact studies. These examine aspects of urban development that traditional static modelling approaches cannot shed light on. The first study examines the potential impacts of the extension of the MBTA Green Line in Boston – and importantly, how long it takes for these impacts to take effect. The second study quantifies the profits that firms forego due to labour market frictions and examines to what extent agglomeration can reduce these deadweight losses.

5.1 Green Line Extension

The Green Line is a light rail transit route part of the MBTA network in Boston. Figure 5-1 shows the Green Line route map. The line has four western branches, which merge before going through downtown Boston. It exits the downtown core, going north to Cambridge, terminating at Lechmere station. Various proposals for northwards extensions have been considered since 1922. Planning for the currently proposed alignment took place in the 2000's with construction finally beginning in 2012. The branch to Union Square opened in March 2022, while the extension to Medford/Tufts is scheduled to open later the same year.

In this scenario study, I examine the impacts of the Green Line extension on the employment size decisions of firm along the extension corridor. Specifically, I quantify job creation, production increase, and the time required to reach a new steady state. Using the estimates from Chapter 4, I evaluate separately the impacts for professional services and food and accommodation services firms. Impacts on manufacturing firms were not considered in this analysis, since the two agglomeration measures that were found to affect the productivity of manufacturing firms, local (walk only, 2.0) and regional (5.0) specialization, do not change considerably as a result of the Green Line extension.

Figure 5-1: Map of the Green Line and its extension



(Wikipedia, 2022)

Modelling Assumptions

Unfortunately, the analysis is inevitably anachronistic because the firm dataset only covers up until 2013. Thus, the results reflect the impact of the Green Line extension, had it completed its construction in 2013. Improvements to the transportation system are captured directly by the agglomeration measures. Specifically, the generalized cost term c_{ij} in equation (4.1) is the logsum of a travel mode choice, which includes the (dis)utility of travel by transit. Thus, reducing the travel times used to calculate the agglomeration measures effectively reduces the rate of spatial decay of the agglomeration effects.

To model the changes to agglomeration variables resulting from the Green Line extension, I consider two scenarios: transportation network improvements only (GLX); and transportation network improvements and densification (GLX-D). I consider the transport network improvements separately for within-corridor travel and travel to/from the corridor. Specifically, for:

- Within-corridor travel, I reduce the transit in-system travel times by 45% percent when both trip origin and destination are within a 500-metre buffer of the corridor. This is based on a comparison between the travel speed on other dedicated right-of-way transit services in Boston and the current travel speed by bus along the corridor.
- Travel to/from corridor: I reduce transit in-system travel times by 5 minutes if either trip origin or destination is within a 500-metre buffer of the corridor. This is half the of the travel time reduction for traversing the entire corridor (10 minutes). While a flat 5-minute reduction will not accurately represent the improvements for all travellers and locations, it is a reasonable first-order that can be obtained without a detailed analysis of local travel patterns.

These changes apply to both the GLX and GLX-D scenarios. It should be noted that these changes likely do not account for all the improvements to the transportation system resulting from the new transit service. In particular, they do not capture the potential reduction in congestion along parallel roadways from travellers choosing to take transit rather than to drive and from removing busses (and their frequent stops) from occupying lane space. Furthermore, the two cases do not cover through-travel. However, the speed of through-travel should have very limited impact on agglomeration within the corridor.

We typically associate mass transit with higher density. However, since I do not model feedback effects (from the transport system to land development), modelling the densification resulting from the Green Line extension has to be accomplished by brute force. That is, for the GLX-D scenario, I increase both population and employment density by 20% within a 500-metre buffer of the corridor. Naturally, this is unlikely to be an accurate representation of future development – especially as densification would also be gradual rather than instantaneous with the opening of the Green Line extension. However, it will provide an impression of the order of magnitude of changes that can be expected.

Simulation

The model estimated in Chapter 4 contained multiple sources of randomness. Specifically, the error terms ε representing idiosyncratic shocks to average worker productivity and the stochastic adjustment process with exponential-distributed inter-event times z. Omitting these sources of randomness would not accurately represent the likelihood of future development. Hence, I use simulation to capture the variance of this stochasticity. For each scenario-sector combination, the simulation repeats the following four steps:

1. Initialize at baseline conditions, i.e. without the Green line extension
2. Allow burn-in to ensure system is in steady state
3. Introduce the GLX / GLX-D scenarios as shocks to the system
4. Simulate the response of each firm until steady state (and beyond) using the estimated firm-specific parameters and drawing error terms ε and inter-event times z

I run 100 simulations of each firm's response and summarize the mean and 95% confidence interval of additional jobs created and increase in gross production. Note that for this study, I determine the point in time by which steady state has been reached by inspection of the aggregate employment size trajectories. A better method would use a standardized definition of steady state, e.g. once employment sizes remain within a certain number of standard deviations within a moving average. This would yield more rigorous *time to steady state* and allow for calculation of confidence intervals for this result as well. However, I leave the testing and implementation of rules for future work.

Furthermore, it should be noted that comparisons are made between scenarios under static conditions, i.e. the exogeneous variables such as GDP and agglomeration measures do not change over the course of the analysis. Furthermore, as mentioned, feedback effects are not captured. Modelling these feedbacks require integration into a larger IUM framework, as discussed in Section 3.4. Thus, the comparisons should be considered all else equal – and remaining equal.

Results and Discussion

Table 5-1 and Table 5-2 present results for professional services and food and accommodation services, respectively. The results represent the impacts of agglomeration economies once the new steady state has been reached, i.e. number of jobs created at steady state and the increase in gross annual production at steady state. Note however, that the raw values reflect impact on firms in the InfoGroup dataset and not the entire firm population.

Table 5-1: Impact of Green line extension and densification of professional services firms

Professional	**Scenario**			
	GLX		*GLX-D*	
Job creation	3.9% [1.8% – 5.9%]	56.3 [26.7 – 85.9] jobs	4.6% [2.3% – 6.9%]	66.1 [34.9 – 97.3] jobs
Production increase	3.2% [2.5% – 3.9%]	9.4 [7.3 – 11.6] million dollars	3.7% [2.9% – 4.5%]	10.9 [8.5 – 13.3] million dollars
Time to reach steady state	12.4 years		16.4 years	

Table 5-2: Impact of Green line extension and densification of food and accommodation services firms

Food & Accommodation	**Scenario**			
	GLX		*GLX-D*	
Job creation	2.2% [-1.7% – 6.1%]	18.6 [-14.7 – 51.6] jobs	3.0% [-0.3% – 6.3%]	25.4 [-2.7 – 53.6] jobs
Production increase	0.84% [0.8% – 1.6%]	0.54 [-0.19 – 1.27] million dollars	1.9% [0.9% – 2.9%]	1.2 [0.6 – 1.8] million dollars
Time to reach steady state	9.8 years		11.3 years	

The analysis shows that impacts on the professional services sector is greater than it is on food and accommodation services. This is not surprising, both intuitively, and considering the estimation results from the previous chapter, where professional services were found to benefit considerably from agglomeration of employment.

At first glance, the impacts may seem small overall – we usually associate new mass transit lines with large-scale development. But these impacts are for firms already existing in the area, i.e. on the intensive margin only. The extension will likely also attract businesses from elsewhere, impacting the extensive margin. The analysis also highlights the importance of modelling feedback effects, rather than merely comparative statics. While building the Green Line extension clearly improves travel speed and access for many, focusing only on the individual traveller misses a crucial benefit of mass transit – namely increasing capacity. Although speed and capacity are obviously intertwined, the latter is more often neglected since the supply-side is less commonly modelled. In turn, this emphasizes the potential value of integrated modelling frameworks. Despite their issues, limitations, and lacking of widespread traction, there is clearly a need for IUMs to fulfill. For example, in a larger integrated urban modelling framework, we could model the evolution of land use, and in the long term, urban form. As professional services firms would be willing to pay more for real estate in dense areas due to the productivity benefits they derive from agglomeration, they will slowly displace other sectors, like manufacturing. The proposed model provides a theoretically consistent way of quantifying the valuation of agglomeration economies for different sectors.

The last rows in Table 5-1 and Table 5-2 show the time until a new steady state is reached. Because I model the trajectories of firm's employment sizes, I can predict how long it takes before the impacts take effect. If the temporal dimension is of interest – which it should be more often – the model can produce full development trajectories as shown in Figure 5-2.

Figure 5-2: Job creation over time for food and accommodation services firms under the GLX-D scenario

The trajectory also illustrates the potential dangers of more static before/after-type approaches. If we measured the impact after 2 years, we would only capture a third of the effect. On the other hand, if we assumed that the full impact would take effect immediately, we would have been very wrong for the first five to ten years. Again, urban development is not just about what and where, but importantly also, when.

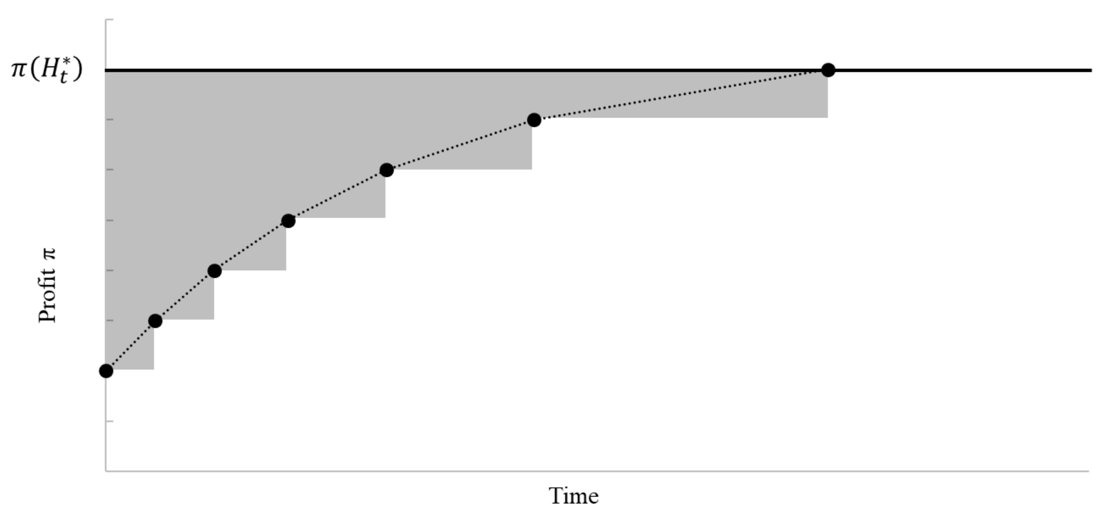
5.2 Cost of Labour Market Frictions

Most empirical work on agglomeration economies examines its innovation-promoting and productivity-enhancing effects. Although matching benefits are well-understood in both the urban and labour economics literature, few have attempted to quantify them in a micro-analytic framework. Uniquely, the modelling of individual businesses' employment size trajectories allows me to determine the cost of labour market frictions and to what extent better matching between employers and employees can alleviate these. Specifically, the gap between profits at the actual and optimal employment sizes are effectively foregone profits resulting from poor matching, incomplete or imperfect knowledge, or other frictions in the labour market.

$$Loss(H) = \pi(H_{it}^{*}) - \pi(H) \tag{5.1}$$

This is depicted visually as the shaded area in Figure 5-3. The analysis in this section focuses specifically on businesses looking to expand their employment size. It is in hiring processes that matching benefits have effect, whereas there is no apparent or direct theoretically compelling link between agglomeration economies and firing processes. For negative parameters, the agglomeration measures associated with the quadratic adjustment costs, i.e. the μ_2 term in the model specification (4.6) increase the rate of hiring. This allows firms to approach their optimal employment size faster, in turn reducing the deadweight loss. In Figure 5-3 this corresponds to a steeper slope on the trajectory, which shrinks the shaded area representing lost profits.

Figure 5-3: Foregone profits due to labour market frictions



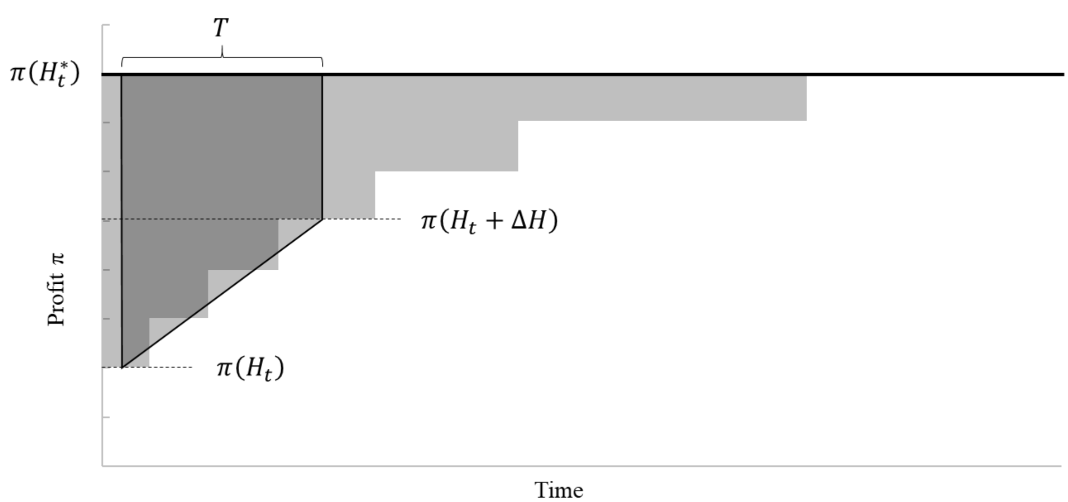
Quantifying the Costs of Labour Market Frictions

In practice, quantifying the firms' actual foregone profits exactly is not straightforward since we do not observe firm's full employment size trajectories. Rather they are inferred stochastically based on a subset of observed data points. Calculating the expected foregone profits given these observations requires integration of the loss function (5.1) over the joint probabilities of the exact time of each adjustment event, which does not lead to a convenient integral. Instead, I approximate the expected losses using the already derived model likelihood function as described by equation (5.2) and depicted in Figure 5-4.

$$UB = \sum_{\Delta H=0}^{H_{it}^{*}-H_{it}} P(\Delta H|\theta_i, \varepsilon_{it}) \cdot T \cdot \left(loss(H_{it} + \Delta H) + \frac{loss(H_{it}) - loss(H_{it} + \Delta H)}{2} \right) \quad (5.2)$$

The idiosyncratic errors ε_{it} and firm-specific parameters θ_i come from the model in Chapter 4. The expected loss is represented by the probability-weighted average over the cases where the firm's trajectory is constant, i.e. a straight line between H_{it} and $H_{it} + \Delta H$. Due to the concavity of actual employment size trajectories, the straight-line approximation will on average over-estimate losses.

Figure 5-4: Approximation of expected foregone profits



I calculate the losses as a percentage of the optimal profit, i.e. LB / $\pi(H_{it}^{*})$ and UB / $\pi(H_{it}^{*})$, since the interpretation of the losses by themselves is not particularly meaningful.

Similarly, I calculate the approximate loss in total production. While firms act to maximize their profits, the impact on production is of more interest for an economic and policy-making perspective. Finally, I explore to what extent agglomeration economies can alleviate the production losses by conducting a sensitivity analysis. I increase by 10% each of the agglomeration measures that were found to improve hiring rates and repeat the calculation above. Specifically, the relevant measures were, for:

- Manufacturing: population (2.0), specialization (walk-only, 2.0);
- Professional services: population (0.25), specialization (5.0); and
- Food and accommodation services: population (1.0).

Results and Discussion

Table 5-3 presents the results of this analysis for each of the three sectors, manufacturing, professional services, and food and accommodation services. Overall, firms forego between 2.71% and 3.90% of their potential profits due to labour market frictions. For scale, the 3.90% of potential profits for professional services firms corresponds, on average, to around $55,000 per year for each firm.

Table 5-3: The costs of labour market frictions

	Manufacturing	**Professional**	**Food & Accommodation**
Approximate Expected Profit Loss	2.71%	3.90%	3.30%
Approximate Expected Production Loss	9.17%	13.17%	11.66%
Elasticity (10%)	-0.23%	-0.11%	-0.01%

The losses are generally smaller, percentagewise, in the manufacturing sector. While the expected loss percentages are not too far apart for professional services and food and accommodation services, the absolute dollar value of the losses is much greater for the former.

For production, the approximate expected losses are 9.17%, 13.17%, and 11.66% percent for each of the sectors respectively. Put differently, a professional services firm that could produce a million dollars worth of output, instead only produces around 870,000 dollars. To test how much agglomeration can help reduce these losses I increase the relevant agglomeration measures by 10% and examine the impact. I find that the effects are very small. A 10% increase in the relevant agglomeration measures reduces expected losses by between 0.01% and 0.23%. Manufacturing is the most sensitive to these changes, which is unsurprising given the large benefits this sector derives from a specialized local labour pool as we found in the previous chapter. On the other hand, food and accommodation services derive little benefit from the increased agglomeration measures. However, the adjustment costs they face, specifically as represented by the μ_2^+ parameter, were already the lowest. While the small elasticities are somewhat disappointing, this is ultimately not too surprising since spatial frictions only make up a small portion of the total search costs – especially as ICT has become more prevalent. This does perhaps suggest that policies aimed at addressing spatial mismatch should be considered and targeted very carefully.

6 CONCLUSION

6.1 Summary of Contributions and Findings

This dissertation makes several contributions to the modelling of firms, jobs, and agglomeration economies in the context of integrated urban models (IUM). In Chapter 3, I presented a novel approach to modelling the dynamics of firms' employment expansion and contraction decisions. In particular, I designed a dynamic programming Markov model that predicts the evolution of employment size trajectories over time as opposed to merely static outcomes. This enables urban modellers to answer questions about *when* development happens in addition to the usual *what* and *where*. Crucially, as urban development is inertial and path-dependent, disregarding the former (when) can potentially result in erroneous answers about the latter (what and where). Firms in the model are forward-looking and maximize the net present value of all future profits. I formulated the firm's Markov decision process (MDP), whose solution, the optimal trajectory found via dynamic programming, is the firm's decision variable. I derived a bespoke likelihood function (3.19) that is effectively a rate-varying Poisson, reflecting the varying slope of the firm's desired employment size trajectory. To estimate the model, I formulated a hierarchical Bayesian model in which parameters are firm-specific and distributed according to a set of hyperparameters. Repeatedly sampling from the joint posterior distribution using a Gibbs sampling algorithm recovers the model parameters.

In Chapter 4, I applied the model to the Greater Boston Area. Using a panel dataset of historical business records between 2003-13, I explored the heterogenous effects of agglomeration economies on manufacturing, professional services, and food and accommodation services firms. Specifically, I found that manufacturing firms benefit primarily from specialization; it both increases productivity and reduces adjustment costs. However, the spatial extents of the effects are more local than expected. Professional services benefit from agglomeration of local employment, regardless of sector. They also benefit from having access to a larger regional labour pooling. However, these benefits only materialize through matching frequency, lowering the cost hiring, and not match quality, which would increase productivity. For food and accommodation services, agglomeration of people and jobs increase their market potential, however local specialization also increases competition. Additionally, I found that adjustment costs, i.e. the costs associated with hiring and firing employees, appears to increase with the skills and knowledge intensity of the sector. As expected, when tasks become more specialized, finding the right skills match becomes increasingly difficult and firing employees whom you have invested training in becomes less desirable.

In Chapter 5, I demonstrated some of the unique capabilities of the proposed model. Specifically, I use the estimation results from the previous chapter to examine the impacts of the Green Line extension in Boston and to quantify the foregone profits resulting from adjustment costs. The Green Line extension was found to increase professional services jobs by 3.9% and gross production by 3.2%, reaching a new steady state after 12.4 years. Similarly, for food and accommodation services it increases jobs by 2.2% and gross production by 0.84%, reaching steady state after 9.8 years. For manufacturing, the changes were not modelled because manufacturing firms benefit from specialization which change minimally as a result of the Green Line extension. Next, examining the costs of labour market frictions, I found that manufacturing, professional services, and food and accommodation services firms forego 2.71%, 3.90%, and 3.30% of their potential profits, respectively, due to the costs of hiring. This corresponds to an loss in production of 9.17%, 13.17%, 11.66%, respectively.

6.2 Future Research Avenues

The effort to model the dynamics of firms' employment size decisions presented in this dissertation is not conclusive. It builds upon a vast existing IUM literature and calls on future research to make further improvements. These potential improvements pertain to both the model itself and to how it is used.

The main challenge for the current model formulation is that it is not well-suited for modelling very large firms. The fundamental assumptions about the adjustment process, i.e. that the occur one-at-a-time and sequentially, are likely not very appropriate for large firms – and especially for downsizing. Modelling large firms is also associated with practical challenges. Large, and very rapid, adjustments slow down the evaluation of the likelihood function, whose computation time scales with $O(\Delta H_{it}^2)$. Furthermore, large adjustments yield very small likelihoods, to the point where numerical errors become a potential concern. Thus, formulating a natural way to accommodate larger firms could not only contribute to the credibility of the approach but also make its application less computationally demanding. Beyond the handling of large firms, the approach would benefit from additional testing of structural and distributional assumptions. For example, some parameters could possibly be more accurately modelled as log-normal distributed rather than the current normal assumption, or alternative specifications of adjustment cost functions could be considered.

Applications of the current model also present numerous potential avenues for research. Most obviously, estimations for sectors other than those covered here will paint a more complete picture of the impacts of agglomeration economies. Applications to other locations would inform us about the degree to which agglomeration economies are generalizable or context-dependent. As noted in the impact analyses in Chapter 5, the omission of feedback effects is one of the key limitations of using the model on its own. Modelling these feedbacks and interactions with the larger urban system is main argument for the IUM approach. Chapter 3.4 provides an overview and a starting point for how to integrate the model into a larger IUM framework. However, the specifics, which comprise by far the bulk of the work, have been left for future research. Finally, the dynamic approach adopted here could be beneficial for the modelling other urban processes that exhibit inertial responses. In particular, literature already exists for dynamic job search and wage progression models from the workers' perspective. However, these models are very limited in their representation of space and agglomeration economies. Incorporating job search into the suite of urban models is potentially a fruitful avenue of research. Furthermore, the temporal dimension of existing relocation and vehicle ownership models is for the most part rudimentary or non-existent. Applying a duration modelling approach to these processes could improve the credibility of the long-term predictions made by IUMs.

APPENDIX A: ALTERNATIVE DERIVATION OF OPTIMAL FIRM POLICY

Let us consider the same infinite time horizon where firms discount future cash flows at rate δ, as well as the same profit and adjustment cost functions as described in Chapter 3.1. However, with discrete periods of length T, the reward function becomes

$$R(H_t, \lambda_t) = \big(\pi(H_t) - \alpha(\lambda_t)\big)\, T \tag{A.1}$$

We can write the *value* function as a recursive Bellman equation:

$$V(H_t, \lambda_t) = R(H_t, \lambda_t) + d(T)E[V(H_{t+1}, \lambda_{t+1})] \tag{A.1}$$

where $dE[V(H_{t+1}, \lambda_{t+1})]$ is the expected value in the subsequent period. The optimal policy, i.e. decision rule, for λ is that which maximizes the value function V.

Before we can proceed, we need to be able to evaluate the expected value in the subsequent period $E[V(H_{t+1}, \lambda_{t+1})]$. If the adjustment rate λ is constant, then the probability of making an adjustment of size ΔH follows a Poisson PMF $q(\Delta H, \lambda T)$, and we can evaluate the expectation of the subsequent period accordingly.

$$E[V(H_{t+1}, \lambda_{t+1})] = \sum_{\Delta H=0}^{\infty} q(\Delta H, \lambda_t T)\, V(H_t + \Delta H, \lambda_{t+1})$$

$$V(H_t, \lambda_t) = \big(\pi(H_t) - a(\lambda_t)\big)\, T + d(T) \sum_{\Delta H=0}^{\infty} q(\Delta H, \lambda_t T)\, V(H_t + \Delta H, \lambda_{t+1}) \tag{A.2}$$

However, the adjustment rate does not generally remain constant as the employment size changes. In particular, we would expect that adjustment rates diminish as the employment size approaches its optimum, as seen in standard partial adjustment models. To circumvent this issue we consider infinitesimal period lengths T effectively changing to continuous time. However, before moving to continuous time, let us re-write the value function for convenience. First, we separate out the case with no adjustments made, i.e. $\Delta H = 0$:

$$\begin{aligned}
V(H_t, \lambda_t) &= \big(\pi(H_t) - a(\lambda_t)\big)\, T + d(T) \sum_{\Delta H=1}^{\infty} q(\Delta H, \lambda_t T)\, V(H_t + \Delta H, \lambda_{t+1}) \\
&\quad + d(T)\left(1 - \sum_{\Delta H=1}^{\infty} q(\Delta H, \lambda_t T)\right) V(H_t, \lambda_t) \\
&= \big(\pi(H_t) - a(\lambda_t)\big)\, T + d(T) \sum_{\Delta H=1}^{\infty} q(\Delta H, \lambda_t T)\, V(H_t + \Delta H, \lambda_{t+1}) \\
&\quad + d(T)\, V(H_t, \lambda_t) - d(T) \sum_{\Delta H=1}^{\infty} q(\Delta H, \lambda_t T)\, V(H_t, \lambda_t)
\end{aligned} \tag{A.3}$$

Then rearrange and divide both sides by T.

$$\begin{aligned}
&\big(1 - d(T)\big) V(H_t, \lambda_t) \\
&\quad = \big(\pi(H_t) - a(\lambda_t)\big)\, T + d(T) \sum_{\Delta H=1}^{\infty} q(\Delta H, \lambda_t T)\, V(H_t + \Delta H, \lambda_{t+1}) \\
&\quad - d(T) \sum_{\Delta H=1}^{\infty} q(\Delta H, \lambda_t T)\, V(H_t, \lambda_t) \\
&\frac{\big(1 - d(T)\big)}{T} V(H_t, \lambda_t) \\
&\quad = \pi(H_t) - a(\lambda_t) + d(T) \sum_{\Delta H=1}^{\infty} \frac{q(\Delta H, \lambda_t T)}{T}\, V(H_t + \Delta H, \lambda_{t+1}) \\
&\quad - d(T) \sum_{\Delta H=1}^{\infty} \frac{q(\Delta H, \lambda_t T)}{T} V(H_t, \lambda_t)
\end{aligned} \tag{A.4}$$

Now, we move to continuous time by letting $T \to 0$. For the Poisson distribution we have the following limits:

$$\lim_{T \to 0} \frac{q(1, \lambda T)}{T} = \lambda$$

$$\lim_{T \to 0} \frac{q(\Delta H, \lambda T)}{T} = 0 \quad for\ \Delta H > 1 \tag{A.5}$$

And we can write the per-period discount rate $d(T)$ in continuous time as $e^{-\delta T}$.

$$\lim_{T \to 0} \frac{1 - \mathrm{e}^{-\delta T}}{T} = \delta \tag{A.6}$$

Thus, as $T \to 0$, equation (A.4) becomes:

$$\delta V(H_t, \lambda_t) = \pi(H_t) - a(\lambda_t) + \lambda_t\, V(H_t \pm 1, \lambda_{t+1}) - \lambda_t\, V(H_t, \lambda_t) \tag{A.7}$$

The sign of the $\pm$ depends on the direction of adjustment with $+$ for hiring and – for firing. Next, rearranging and writing out the adjustment costs

$$V(H_t, \lambda_t) = \frac{\pi(H_t)}{\delta + \lambda_t} - \frac{a(\lambda_t)}{\delta + \lambda_t} + \frac{\lambda_t V(H_t \pm 1, \lambda_{t+1})}{\delta + \lambda_t}$$

$$V(H_t, \lambda_t) = \frac{\pi(H_t)}{\delta + \lambda_t} - \frac{\mu_0 + \mu_1 \lambda_t + \mu_2 \lambda_t^2}{\delta + \lambda_t} + \frac{\lambda_t V(H_t \pm 1, \lambda_{t+1})}{\delta + \lambda_t} \tag{A.8}$$

We can now determine the optimal adjustment rate using the first order condition.

$$\frac{\partial V(H_t, \lambda_t)}{\partial \lambda_t} = 0$$

$$= -\frac{\pi(H_t)}{(\delta + \lambda_t)^2} + \frac{\mu_0 - \delta\mu_1 - 2\delta\mu_2\lambda_t - \mu_2\lambda_t^2}{(\delta + \lambda_t)^2} + \frac{\delta\, V(H_t \pm 1, \lambda_{t+1})}{(\delta + \lambda_t)^2} \tag{A.9}$$

This yields a quadratic equation. Finally, we can find the roots of the quadratic to arrive at the optimal adjustment rate $\lambda^{*\prime}(H_t)$.

$$\mu\lambda_t^2 + 2\delta\mu\lambda_t - (\delta\, V(H_t \pm 1, \lambda_{t+1}) - \pi(H_t) + \mu_0 - \delta\mu_1) = 0 \tag{A.10}$$

$$\lambda^{*\prime}(H_t) = -\delta + \sqrt{\delta^2 + \frac{\delta\, V(H_t \pm 1, \lambda_{t+1}) - \pi(H_t) + \mu_0 - \delta\mu_1}{\mu_2}}$$

The negative root yields a negative adjustment rate and is not permissible.

APPENDIX B: BAYESIAN ESTIMATION PROCEDURE

We have the posterior distribution of the model parameters from equation (3.22):

$$K(\Theta, W, \sigma^2, \theta_n \forall n, \varepsilon_{nt} \forall nt | \Delta H) \propto \left[\prod_n \phi(\theta_n | \Theta, W) \prod_t P(y_{nt} | \theta_n, \varepsilon_{nt}) \phi(\varepsilon_{nt} | 0, \sigma^2)\right] k(\Theta, W, \sigma^2) \tag{3.22}$$

We can sample from this joint posterior distribution by combining two MCMC methods, namely Gibbs sampling and the Metropolis-Hastings algorithm. I describe these two sampling methods in this appendix.

Gibbs sampling

Gibbs sampling is a useful technique for sampling from multivariate distributions, when drawing from the joint distribution directly is difficult but drawing from the conditional distributions is relatively easier. For illustrative purposes, let us consider the random variables ϵ_1 and ϵ_2. We want to sample from the joint distribution $f_{1,2}(\epsilon_1, \epsilon_2)$ but doing so directly is not feasible. However, we know how to sample from the conditional distributions $f_{1|2}(\epsilon_1|\epsilon_2)$ and $f_{2|1}(\epsilon_2|\epsilon_1)$. Then we can use the following steps to achieve draws from the joint distribution.

1. Initialize – set $i = 0$
2. Choose any initial value for ϵ_1 with non-zero density and label it ϵ_1^0
3. Draw an initial value for ϵ_2 from $f_{2|1}(\epsilon_2|\epsilon_1^0)$ and label it ϵ_2^0
4. Draw the subsequent value for ϵ_1 from $f_{1|2}\left(\epsilon_1\middle|\epsilon_2^i\right)$ and label it ϵ_1^{i+1}
5. Draw the subsequent value for ϵ_2 from $f_{2|1}\left(\epsilon_2\middle|\epsilon_1^i\right)$ and label it ϵ_2^{i+1}
6. Increment – set $i = i + 1$
7. Repeat steps 4 through 6

For large enough i, the draws of ϵ_1 and ϵ_2 approximate the joint distribution $f_{1,2}(\epsilon_1, \epsilon_2)$.

Metropolis-Hastings

The Metropolis-Hastings algorithm is another technique for circumventing direct sampling. It is useful when we know and can evaluate a function that is proportional to the density. Let $f(\epsilon)$ be the distribution of interest. The Metropolis-Hastings algorithm then proceeds as follows:

1. Initialize – set $i = 0$
2. Choose an initial value for ϵ and label it ϵ^0
3. Get a trial value for the subsequent draw of ϵ

$$\widehat{\epsilon^{i+1}} = \epsilon^i + \eta \tag{B.1}$$

   where η is a random variable drawn from a symmetric distribution $g(\eta)$, e.g. Normal, with 0-mean.

4. Choose whether to accept the trial value or keep the previous value as the subsequent draw of ϵ.

$$\epsilon^{i+1} = \begin{cases} \widehat{\epsilon^{i+1}} \;\; if \;\; \dfrac{f\left(\widehat{\epsilon^{i+1}}\right)}{f(\epsilon^i)} \geq u \\ \epsilon^i \quad otherwise \end{cases} \tag{B.2}$$

   Here, u is a uniformly distributed random variable between 0 and 1.

5. Increment – set $i = i + 1$
6. Repeat steps 3 through 5

For sufficiently many iterations, the draws of ϵ converge to $f(\epsilon)$. Note that we can calculate the ratio $\frac{f\left(\widehat{\epsilon^{i+1}}\right)}{f(\epsilon^i)}$ if we know a function that is proportional to $f(\epsilon)$ since any constant cancels out.

Putting it all together

To sample from the joint distribution (C.1), we iterate over the five conditional distributions in a Gibbs sampling algorithm presented in section 3 (and repeated below for convenience). For 1. and 2. we nest Metropolis-Hastings samplers within the Gibbs algorithm. For each iteration of the outer Gibbs algorithm we also increment the inner Metropolis-Hastings algorithms a single time. Sampling from the conditional distributions 3., 4., and 5. is relatively straight-forward. These are draws from Normal, Inverse Wishart, and Inverse Gamma distributions. Most scientific computing software and libraries have built-in functions to carry out these draws.

1. Draw θ_n conditional on b, W, ε_{nt} and the observed data ΔH_n.

   I use a Metropolis-Hastings algorithm to make these draws; evaluating trial values by the conditional probability (3.23).

$$K(\theta_n|b, W, \varepsilon_{nt}, \Delta H_n) \propto \prod_t P(\Delta H_{nt}|\theta_n, \varepsilon_{nt})\phi(\theta_n|b, W) \quad \forall n \tag{3.23}$$

2. Draw ε_{nt} conditional on σ^2, θ_n, and the observed data ΔH_{nt}.

   Again, we use a Metropolis-Hastings algorithm to make these draws. Here we evaluate trial values by the conditional probability (3.24).

$$K(\varepsilon_{nt}|\sigma^2, \theta_n, \Delta H_{nt}) \propto P(\Delta H_{nt}|\theta_n, \varepsilon_{nt})\phi(\varepsilon_{nt}|0, \sigma^2) \quad \forall nt \tag{3.24}$$

3. Draw b conditional on W and θ_n.

   This is a draw from a Normal distribution (3.25).

$$K(b|W, \theta_n \forall n) = \mathcal{N}\left(\frac{\sum_n \theta_n}{N}, \frac{W}{N}\right) \tag{3.25}$$

   where N is the total number of firms.

4. Draw W conditional on Θ and θ_n.

   This is a draw from the Inverse Wishart distribution (3.26)

$$K(W|\Theta, \theta_n \forall n) = IW\left(V + N, \frac{VJ + N\bar{S}}{V + N}\right) \tag{3.26}$$

   where $\bar{S} = \frac{\sum_n (\theta_n - \Theta)(\theta_n - \Theta)'}{N}$, V is the number of parameters to be estimated, i.e. the length of each θ_n, and J is a V-dimensional identity matrix

5. Draw μ conditional on ε_{nt}

   This is a draw from the Inverse Gamma distribution (3.27)

$$K(\sigma^2|\varepsilon_{nt} \forall nt) = IG\left(1 + N, \frac{1 + N\bar{s}}{1 + N}\right) \tag{3.27}$$

   where $\bar{s} = \frac{\sum_{nt} (\varepsilon_{nt} - \sigma^2)(\varepsilon_{nt} - \sigma^2)'}{R}$, and R is the total number of observations in the dataset.

APPENDIX C: DYNAMIC MODELS OF JOB SEARCH AND ACCESSIBILITY

Background and Purpose

The central motivation for this study is to understand how cities facilitate economic activity and development. This question has been treated in several branches of the economics literature, including urban economics, economic geography, and to some extent, labour economics. Thus, the objective here is not to derive new theory, but rather to draw upon these ideas and examine empirically how they apply at the intra-metropolitan scale.

Although cities are much more than just places of economic opportunity, their role as labour markets is central to their existence (Bertaud, 2014). Of particular interest for job search modelling purposes are agglomeration benefits, i.e. urban external scale economies, such as thick market effects and knowledge spillovers. The thick market effect, first formalized by Helsley and Strange (1990), describes how having access to more jobs, i.e. a thick labour market, improves worker-job matches given a system of heterogeneous workers and jobs. This makes intuitive sense since access to more jobs affords the job-seeker more opportunities to find one that suits their skills and interests, in turn making them more productive. Additionally, thick markets also reduce the risk of extended unemployment following idiosyncratic shocks (Moretti, 2011). From the workers' perspective, these effects yield higher wages and shorter unemployment durations. The empirical evidence largely supports these hypotheses. For example, Petrongolo and Pissarides (2003) find, using a semi-structural model, that workers in larger cities in the UK have higher reservation wages, endogenizing the effects of better matching; Immergluck (1998) finds, in a reduced-form estimation, that unemployment rates at neighbourhood-level are lower where labour markets are thicker. Knowledge spill-overs, popularly associated with Marshall's (1890) quote "The mysteries of the trade become no mystery; but are as it were in the air", have also been the subject of much academic attention. From the worker's point-of-view, this is particularly relevant vis-à-vis learning benefits. Notably, numerous studies have documented the learning advantages of workers in large cities compared to those in small ones (Glaeser and Maré, 1994; de la Roca and Puga, 2017). Theoretical models of the mechanisms underlying knowledge spill-over generally rely on more frequent interactions between people in dense areas, resulting in more rapid learning (Glaeser, 1999). For a more detailed discussion on the micro-foundations of urban agglomeration economies, see e.g. Duranton and Puga (2004) and Puga (2010).

Another important distinction between the work of economists and that of urban and transportation modellers lies in the scope of analysis; the work of urban economists and economic geographers often examines regional dynamics, i.e. the interactions within a system of cities, whereas urban and transportation modellers generally focus their attention on intra-metropolitan dynamics. Consequently, the treatment of space and spatial friction, in particular transportation costs, in economics is often simplistic or abstract compared to that in urban and transportation modelling. For example, most of the aforementioned studies on agglomeration effects compare between small and large cities, treating whole cities as single uniform markets. In comparison, urban and transportation modellers often use exceptionally detailed representations of both land use and transportation networks. For example, the SimMobility simulation framework accounts for parcel-level land use and travel times by different travel modes at various times of the day as inputs for the land market bidding models (Adnan et al., 2015). The rigour and detail with which urban and transportation modellers consider space and spatial friction can potentially contribute to the understanding of the mechanisms underlying agglomeration economies. These effects fundamentally derive from the reduced spatial friction, i.e. transportation costs, in dense well-connected areas. However, agglomeration effects, beyond congestion, have largely been peripheral to the efforts of urban and transportation modellers.

Early efforts to model the individual job seekers problem were essentially statistical optimal sample size or optimal stopping problems (Stigler, 1962; Chow et al., 1971), i.e. the optimal number of job offers to sample before taking a job. Building upon these, labour economists formalized equilibrium models. e.g. equilibrium aggregate unemployment rates as function of search friction, unemployment benefits, and wages expected from working (Diamond and Maskin, 1965; Mortensen, 1978; Pissarides, 1979). Empirical estimation of such models is the focus of the literature on dynamic discrete choice (DDC) models, also sometimes known as dynamic programming discrete choice models. The methods in this study draw primarily from this literature. In essence, DDCs model decision processes where future utility depends on current choices and can be applied widely, e.g. having a child (Wolpin, 1984), bus engine replacement (Rust, 1987), and of course job search (Miller, 1984). Rust (1994) and Aguirregabiria and Mira (2009) provide comprehensive reviews of DDC modelling methods. There have been few applications of DDC that directly explore agglomeration effects in the labour market. However, two examples of are Petrongolo and Pissarides (2003) and Baum-Snow and Pavan (2012) who examine differences in wages and offer arrival rates between workers in small, mid-sized, and large cities.

From a theoretical point-of-view, the structural estimation of such models with a detailed representation of spatial frictions is useful for testing the various hypothesized sources of agglomerations effects at an intra-metropolitan scale. Distinguishing different agglomeration benefits, e.g. learning from matching, is challenging using reduced-form approaches and must rely on convenient features in the data. From a practical point-of-view, the findings are useful for testing counterfactuals, e.g. the labour market impacts of transportation infrastructure investments, land use policy, etc. It is also a timely contribution to the understanding of spatial barriers to labour mobility considering the structural transformation that the economy is undergoing as consequence of the accelerated rate of technological innovation. Furthermore, a formalized model of job search should be an important component to large-scale urban modelling efforts given the fundamental role that the labour market plays in urban development. However, it is often omitted or ad hoc. For example, job assignment in SimMobility remains a static location choice that does not account for wages, wage progression, or agglomeration economies. Yet, income is an important input to several models, including daily activity choice, vehicle ownership choice, and the residential bidding models.

The overall purpose of the study is to examine the impact of local job access, measured with high spatial resolution, on employment outcomes, such as unemployment rates and the rate of wage progression. More specifically, I will attempt to answer the following three questions:

1. Does local labour market thickness affect employment outcomes?
2. What inferences can we make vis-à-vis the job search processes in different sectors?

3. To what extents do matching and learning each contribute to the agglomeration benefits in job search?

Methods and Data

Modelling framework

Let us now formalize a model of the choices that individuals face in job search. We consider a sequence of discrete time periods, e.g. months or years, and a finite time horizon T counting periods since graduation until retirement age $t = 1, \dots, T$. In each period, individuals can either be employed, receiving wage w, or unemployed and searching for work, receiving unemployment benefit u. Let $\bar{R}_t$ in period t be the systematic utility a person derives in that period.

$$\bar{R}_t(\omega, c) = \omega + \beta_c c \tag{C.3}$$

where,

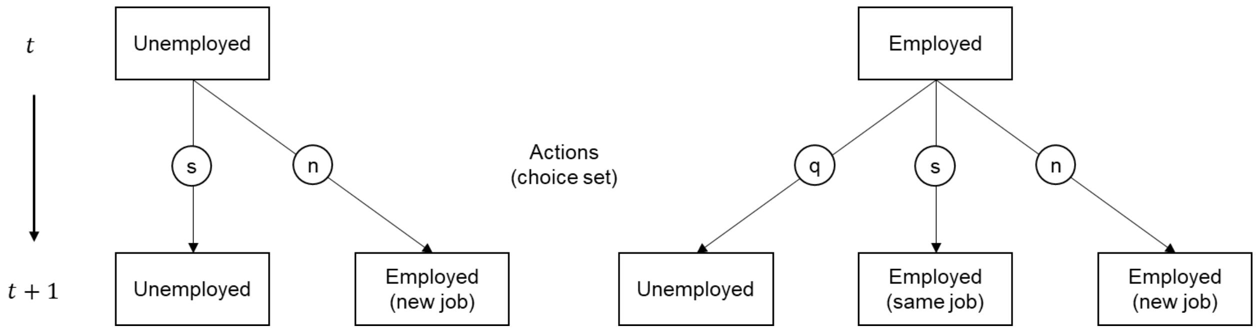
$$\omega = \begin{cases} w & if\ employed \\ u & if\ unemployed \end{cases} \tag{C.4}$$

$$c = \begin{cases} 1 & if\ switched\ jobs\ since\ last\ period \\ 0 & otherwise \end{cases} \tag{C.5}$$

β_c represents the inertia that must be overcome when switching jobs. Note that utility is measured in dollar-terms. At the end of a period, each person chooses to take one of three actions $A = \{q, s, n\}$:

- quit q (provided they were employed);
- stay s (employed or unemployed), or
- accept a new job n (provided at least one new job opportunity presents itself).

Figure C-1: Dynamics of the job search model



Each action is associated with an unobserved utility component ε. We call the total utility a person derives in a period the reward and denote it R_t.

$$R_t(\omega, c) = \bar{R}_t(\omega, c) + \varepsilon \tag{C.6}$$

We assume that people are utility-maximizers. However, rather than simply maximizing the reward R_t, i.e. the utility derived in the period immediately following a choice, we assume instead that people have some foresight, which we approximate with a discounted lifetime utility. For example, an unemployed person with a job offer $w > u$ may choose to stay unemployed. This decision is rational if they expect that remaining unemployed allows them to find a better job with sufficiently high wages to make up for the period(s) of foregone earnings. In other words, the optimal decision now depends on the expectation of actions and events at every future t. Solving this problem by brute force methods becomes infeasible even for small T. Instead we adopt a dynamic programming approach. Let us describe the remaining lifetime utility at time t with a recursively defined value function V_t,

$$V_t(\omega, c) = R_t(\omega, c) + \delta E\left[\max\left(\widetilde{\mathrm{V}_{t+1}}\right)\right] \tag{C.7}$$

where δ is the per-period discount factor and $E\left[\max\left(\widetilde{\mathrm{V}_{t+1}}\right)\right]$ is the expected maximum value, i.e. remaining lifetime utility, at time $t+1$ given the available actions at time t. Stated differently, the remaining lifetime utility at the state defined by (ω, c) at time t is the sum of the immediate reward at t and the expected maximum remaining lifetime utility at $t+1$. To evaluate the expected maximum value, let ε, the random unobserved component associated with each action, be independent and identically distributed (IID) extreme-value with parameters $(0, \mu)$, and let $\bar{V}_t$ denote the value function without this term.

$$\bar{V}_t\,(\omega, c) = \bar{R}_t(\omega, c) + \ \delta E\left[\max\left(\widetilde{\mathrm{V}_{\mathrm{t+1}}}\right)\right] = \omega + \beta_c c + \delta E\left[\max\left(\widetilde{\mathrm{V}_{\mathrm{t+1}}}\right)\right] \tag{C.8}$$

Then we can express the expected maximum value analytically.

$$E\left[\max\left(\widetilde{\mathrm{V}_{\mathrm{t+1}}}\right)\right] = \frac{1}{\mu}\log\left(\sum_{A}\exp(\mu\bar{V}_{t+1}^{A})\right) + \frac{\gamma}{\mu} \tag{C.9}$$

where γ is the Euler-Mascheroni constant. We determine the value function for every period t by backwards induction. In the final period T, the value functions and rewards must be equal, since there are no subsequent periods with rewards. With V_T, we can evaluate the value functions in the second-to-last time period. Repeating these steps with decreasing t we determine the value functions that serve as systematic utilities for the choice problem.

$$\begin{gathered}
V_T(\omega, c) = R_T(\omega, c) \\
V_{T-1}(\omega, c) = \mathrm{R}_{T-1}(\omega, c) + \delta E\left[\max\left(\widetilde{\mathrm{V}_{\mathrm{T}}}\right)\right] \\
\vdots \\
V_1(\omega, c) = \mathrm{R}_1(\omega, c) + \delta E\left[\max\left(\widetilde{\mathrm{V}_2}\right)\right]
\end{gathered} \tag{C.10}$$

Finally, the choice probability in time period t becomes a simple multinomial logit.

$$P(a_t) = \frac{\exp(\mu\bar{V}_{t+1}^{a_t}(\omega, c))}{\sum_{A_t}\exp(\mu\bar{V}_{t+1}^{A_t}(\omega, c))} \tag{C.11}$$

Data

To estimate the model described above, we require detailed longitudinal data from both individuals and the land use transport system. Specifically, we need workers' employment status, wages, and locational information. The section below describes a survey that I will conduct to elicit these. In addition to person-specific data, I will use origin-destination travel impedance matrices and job location data provided to the Future Urban Mobility (FM) research group at the Singapore-MIT Alliance for Research and Technology (SMART) by the Singapore Land Transport Authority (LTA). These describe the land use and transport systems with remarkable spatial detail, dividing Singapore in more than a thousand zones. The median zone area is 0.34 km^2 with sizes generally smaller in areas with dense build-up. This spatial granularity more than suffices for our purposes. On the other hand, the temporal granularity potentially comprises a source of error. Over the past decade, we have received from government agencies historical and periodically updated land use and transportation data. However, matching these to the time steps of the model will undoubtedly still necessitate some interpolation.

Survey of employment histories and wages

The sampling frame for the survey consists of two distinct groups of individuals:

- High school graduates; employed in the accommodation and food service activities sector, corresponding to section I in the Singapore Standard Industrial Classification (SSIC); graduated high school (if female) or completed mandatory military service (if male) in 2008.
- College graduates; employed in the professional, scientific, and technical activities sector, corresponding to section M in the SSIC; graduated college in 2008.

The two groups, representing low and high skill sectors, respectively, will be used for comparison. Constraining the sampling frame by the year survey participants completed their education or military service serves two purposes. First, it ensures that the initial state for the dynamic programming problem is exogenously determined. Additionally, it controls for the overall macro-economic conditions over surveyed period, although conditions can differ between the sectors.

The survey consists of five sections: demographics, education, residential history, employment history, and commuting preferences. Table C-1 presents an initial outline of the survey questionnaire. I plan to conduct the survey on an online platform, such as Qualtrics, starting with a pilot in the early winter of 2019/20. Please refer to the workplan in Section 5 for a more detailed timeline of the survey.

Table C-1: Survey outline

<table>
<tr><th>Section</th><th>Information to elicit</th></tr>
<tr><td>Demographics</td><td>Gender
Race
Birthday
Citizenship / residential status</td></tr>
<tr><td>Education</td><td>Highest degree obtained
Graduation date</td></tr>
<tr><td>Residential history (since graduation)</td><td>For each residential location
<ul style="list-style:none">
Location
If renting
Monthly rent
If owning
Property value
</td></tr>
<tr><td>Employment history (since graduation)</td><td>For each job/unemployment period
<ul style="list-style:none">
If switched jobs
Reason for change
Expected annual raise if stayed at previous job
Start and end dates
Gross monthly income history (incl. promotions and raises)
Occupation
Sector
Full-time or part-time
If part-time
Reason for working part-time
If unemployed
Reason for unemployment
Financial support while unemployed
</td></tr>
<tr><td>Commuting preferences</td><td>Current commute mode
<ul style="list-style:none">
If drive
Max one-way driving commute duration
If transit
Max one-way transit commute duration
If able (license and vehicle availability) and willing to drive
Max one-way driving commute duration
If walk
Max one-way walking commute duration
If able (license and vehicle availability) and willing to drive
Max one-way driving commute duration
If willing to take transit
Max one-way transit commute duration
</td></tr>
</table>

Model Specification

In the survey, we only observe new job opportunities when they are accepted. Unfortunately, this is a limitation for most practical dataset, considering the loose and somewhat subjective definition of what constitutes a new accessible job opportunity. Appropriately handling these opportunities, when they are unobserved, is the primary challenge in this study. First, let us assume that each job opportunity is associated with a wage offer w_{offer}, consisting of a linear-in-parameters terms and two random unobserved components, ν and η. ν represents person-specific unobserved attributes, such as skill, ambition, etc. while η captures unobserved offer-specific heterogeneity. Let both random terms be normal-distributed with mean zero and standard deviations to be estimated.

$$w_{offer} = \beta_{const} + \beta_x x + \nu + \eta \tag{C.12}$$

where x denotes experience. Including an experience term means that the incentive for workers to accept a new job, i.e. take action n, increases the longer they stay in the same position without receiving raises. Additional linear-in-parameters terms, e.g. to account for education, sector, gender, etc., could be introduced to equation (C.*12)* if these were not controlled for in the dataset. For any given job, the wage offer is drawn from the normal distribution $\mathcal{N}\left(\beta_{const} + \beta_x x + \nu, \sigma_\eta^2\right)$. This approach of assuming a probability distribution for the unobserved wage offers is described by Ljungqvist & Sargent (2012). It is also similar in principle to a Heckman correction (Heckman, 1979), however the applications of course differ.

At this point, the attentive reader may recall that wage is a state variable in our dynamic programming problem, whose value function was given in (C.*7)*. A continuous distribution of wages results in an infinitely large state space. Hence, for the sake of tractability, let us instead consider a discrete approximation to the normal distribution H with equally spaced grid points, cumulative distribution function $F_H(\eta)$, and mass function $f_H(\eta)$. Furthermore, we also need to expand the state space to account for individuals' experience x. Specifically, the state variables are those that we require to evaluate the value function and can change over time. Equation (C.*13)* presents the updated value function with wage offers following $f_H(\eta)$.

$$V_t(\omega, c, x) = R_t(\omega, c) + \delta \sum_{\eta \in H} E\left[\max\left(\widetilde{\mathrm{V_{t+1}}}\right)\right] f_H(\eta) \tag{C.13}$$

If the wage offer is unobserved, i.e. $a \in \{q, s\}$, then the likelihood of the observed action must be considered over the all the possible values of H. On the other hand, if the person takes a new job, i.e. $a \in \{n\}$, we observe the new wage, and thus need only consider the joint likelihood of the action taken and the observed wage offer η.

$$L(a_{it}) = \begin{cases} \sum_{\eta \in H} \frac{\exp\left(\mu \bar{V}_{t+1}^{a_{it}}(\omega, c, x)\right)}{\sum_{A_t} \exp\left(\mu \bar{V}_{t+1}^{A_{it}}(\omega, c, x)\right)} f_H(\eta) & if\ a_{it} \in \{q, s\} \\ \frac{\exp\left(\mu \bar{V}_{t+1}^{a_{it}}(\omega, c, x)\right)}{\sum_{A_t} \exp\left(\mu \bar{V}_{t+1}^{A_{it}}(\omega, c, x)\right)} f_H(\eta) & if\ a_{it} \in \{n\} \end{cases} \tag{C.14}$$

Note that when the person chooses to take a new job, $a \in \{n\}$, we may have to consider the person's expectations of raises and promotions at the job they just left. I request this information in a separate question in the survey as it is not available in the employment histories. Additionally, equation (C.*14)* includes the subscript i to account for the individual-specific heterogeneity, which affects the value functions and by extension the choice probabilities. Finally, there are a few decisions related to job search that we do not model. Specifically, I exclude part-time workers from the dataset. The decision to work part time is in many cases determined by outside factors that we do not model here. As such, modelling this decision would not contribute to the study while unnecessarily increasing the complexity of the model. Furthermore, I consider being fired a random exogenous event that occurs to everyone with the same probability.

We now have a fully specified base model without any spatial effects. In the following sections, I propose ways to extend the model which allow us to test our hypotheses regarding various spatial and agglomeration effects.

Thick markets

For the first extension to the base model, we explicitly consider thick market effects. Given heterogeneity in job opportunities, having access to more offers generally results in better matches by allowing the job-seeker to choose the best offer from a larger pool. We incorporate this effect by doing away with the assumption that people only consider a single new job opportunity in each period as was the case in the base model. Instead, I hypothesize that the number of new job opportunities revealed to a person in a time period, denoted N, is monotonously increasing with accessibility z_{home} but with diminishing marginal return, e.g. a natural logarithm function.

$$N = \alpha_h \ln(z_{home}) \tag{C.15}$$

The parameter α represents search effort in face of frictions and constraints, e.g. time to write applications, fatigue, etc. Here, we assume that α_h remains unchanged regardless of employment status or wage. However, one can imagine that those unemployed have more time and incentive to search or that those who feel undervalued at their current job more eagerly look for new opportunities. We can capture such effects by having distinct α's by employment status or by letting α be a function of the difference between an individual's actual wage and their expected wage, i.e. mean of the normal distribution from which we draw wage offers. I will test such specifications empirically. For our purpose, relatively simple measures of accessibility, such as such as number of jobs within some travel time or distance from home, will suffice. We use it as a measure of market thickness as opposed to consumer surplus, which would have called for more comprehensive accessibility measures. The N new opportunities revealed to an individual in a given time period, $w_{offer}^{(1)}, w_{offer}^{(2)}, \dots, w_{offer}^{(N)}$, are IID draws from $\mathcal{N}\left(\beta_{const} + \beta_x x + \nu, \sigma_\eta^2\right)$, i.e. the same normal distribution that we used for wage offers in the base model. Once again, we consider a discrete approximation to this distribution for tractability with cumulative distribution function $F_H(\eta)$ and mass function $f_H(\eta)$. Since we assume that people are utility-maximizers, they only consider the best offer m to determine whether or not to accept a new job at a given time.

$$m = \max\left(w_{offer}^{(1)}, w_{offer}^{(2)}, \dots, w_{offer}^{(N)}\right) \tag{C.16}$$

The probability mass function is then,

$$f_M(m) = 1 - F_H\left(\underline{m}\right)^N - (1 - F_H(m)^N) \tag{C.17}$$

where $\underline{m}$ is the discrete wage value of subsequent rank following m, i.e. the largest possible wage offer that is still smaller than m. In equation (C.*25)*, the $F_H(\underline{m})^N$-term is the probability that all offers are smaller than m and the $(1 - F_H(m)^N)$-term is the probability that any offer is greater m. As the number of offers N and number of grid points in the discretization increase, f_M approaches an extreme value distribution.

I assume that residential relocation choices are exogenous to the model. In reality, job access often plays a role in residential location decisions. However, it is typically one of many factors, meaning that modelling these decisions without omitting factors from the reward function would require us to considerably expand the state space and likely increase computation time by several orders of magnitude. Thus, we keep the state space unchanged, however we still need to update our value and likelihood functions to account for the new distribution of wage offers.

$$V_t(\omega, c, x) = R_t(\omega, c) + \delta \sum_{m \in M} E\left[\max\left(\widetilde{\mathrm{V_{t+1}}}\right)\right] f_M(m) \tag{C.69}$$

$$L(a_{it}) = \begin{cases} \displaystyle\sum_{m \in M} \frac{\exp\left(\mu \bar{V}_{t+1}^{a_{it}}(\omega, c, x)\right)}{\sum_{A_t} \exp\left(\mu \bar{V}_{t+1}^{A_{it}}(\omega, c, x)\right)} f_M(m) & if\ a_{it} \in \{q, s\} \\ \displaystyle\frac{\exp\left(\mu \bar{V}_{t+1}^{a_{it}}(\omega, c, x)\right)}{\sum_{A_t} \exp\left(\mu \bar{V}_{t+1}^{A_{it}}(\omega, c, x)\right)} f_M(m) & if\ a_{it} \in \{n\} \end{cases} \tag{C.70}$$

Deliberate search or serendipitous encounters

In the previous section, we considered only accessibility from home locations. The underlying assumption is that commute time or distance constrain the set of opportunities people consider feasible and as such people limit their search accordingly. However, serendipitous encounters in professional settings are one of the central benefits associated with urban agglomeration and potentially an important avenue for finding new jobs. In practice, we account for this by letting the number of new job opportunities N be a function of both accessibility from home z_{home} and work z_{work}, as shown in equation (C.*25)* below.

$$N = \alpha_h \ln(z_{home}) + \alpha_w \ln(z_{work}) \tag{C.71}$$

Given that we include accessibility from work to capture the benefits of serendipitous encounters in professional networks, it seems sensible to measure z_{work} using a small, e.g. 10 minute, walking time radius. This is in contrast to the z_{home}, which should be measured by a radius reflecting commute willingness, e.g. 60 minutes by transit or car depending on vehicle availability. Again, I will test these hypotheses empirically.

Including accessibility from work in our model introduces a further complication vis-à-vis the dynamic programming problem. Since job location now affects the number of new opportunities, it should be included as a state variable. Unfortunately, this requires that we know the location of each job offer, including those rejected. In essence, the problem is the same as that for unobserved wage offers, which we dealt with by assuming that offers were drawn from a normal distribution. However, conjuring a spatial distribution of job offers is less straightforward, as it should ideally account for job availability within the relevant sector and anchoring effects of the current job location. I discuss some ideas on ways to address this in 0 Potential Research Avenues. For the first iteration of this model extension, we assume that people do not account for changes in job accessibility in their career decisions. In other words, we use equation (C.*71)* for determining the number of opportunities but assume that people make decisions believing that the current number of opportunities will persist in the following time period regardless of where they work. Under this assumption, the value and likelihood functions still look like (C.*69)* and (C.*70)* from the previous section.

Better matching or better learning

Another aspect of urban agglomeration benefits, discussed in the background section, are the learning advantages. Our models thus far would likely have confounded the benefits from learning with those from matching. To capture the learning benefits of dense professional networks, we consider an additional experience term χ in the equation for wage offers.

$$w_{offer} = \beta_{const} + \beta_x x + \beta_\chi \chi + \nu + \eta \qquad \text{(C.72)}$$

where χ, similarly to experience x, accumulates over time but the gain in each period is a function of accessibility from work z_{work}.

$$\chi_{t+1} = \begin{cases} \chi_t + \ln(z_{work}) & if\ employed \\ \chi_t & if\ unemployed \end{cases} \tag{C.73}$$

Importantly, these learning benefits persist across unemployment periods. On the other hand, matching benefits, while also cumulative, reset upon unemployment. For this extension, we need to expand the state space of the dynamic programming problem, which requires a reasonable discretization of the χ variable. Equations (C.*74)* and (C.*75)* present the updated value and likelihood functions.

$$V_t(\omega, c, x, \chi) = R_t(\omega, c) + \delta \sum_{m \in M} E\left[\max\left(\widetilde{\mathrm{V_{t+1}}}\right)\right] f_M(m) \tag{C.74}$$

$$L(a_{it}) = \begin{cases} \sum_{m \in M} \dfrac{\exp\left(\mu \bar{V}_{t+1}^{a_{it}}(\omega, c, x, \chi)\right)}{\sum_{A_t} \exp\left(\mu \bar{V}_{t+1}^{A_{it}}(\omega, c, x, \chi)\right)} f_M(m) & if\ a_{it} \in \{q, s\} \\ \dfrac{\exp\left(\mu \bar{V}_{t+1}^{a_{it}}(\omega, c, x, \chi)\right)}{\sum_{A_t} \exp\left(\mu \bar{V}_{t+1}^{A_{it}}(\omega, c, x, \chi)\right)} f_M(m) & if\ a_{it} \in \{n\} \end{cases} \tag{C.75}$$

Estimation

First, some useful notation: the dataset consists of employment histories from I individuals, $i = 1, \ldots, I$. Let the set of observed actions for individual i be denoted $y_i = \langle a_{i1}, \ldots, a_{it}, \ldots, a_{iT} \rangle$, and the complete set of observed actions $Y = \langle y_1, \ldots, y_i, \ldots, y_I \rangle$. For proof-of-concept testing, I coded the simple version of the model, generated synthetic data following the hypothesized decision-making process with reasonable parameters, and then attempted to estimate the model to see if I could recover the underlying parameters. For these initial tests, I used maximum likelihood estimation in a so-called nested fixed-point algorithm, described by Rust (1994). In short, an outer loop iterates over and optimizes the parameters to be estimated, while an inner loop evaluates the likelihood of the observed actions *(C.25)* by solving the dynamic programming problem for each individual in the sample assuming the parameters given by the outer loop.

$$L(Y|\beta) = \prod_i \prod_t L(a_{it}|\beta) \tag{C.25}$$

Although occasionally successful, the estimation was unstable and highly sensitive to initial conditions due to local extrema. This behaviour is not too surprising considering the complexity of the likelihood function, which involves the recursively solved dynamic programming problem for each individual. Instead, I will attempt to estimate the model using a Bayesian estimation procedure, which I describe below.

Bayesian estimation procedure

By using a Bayesian procedure, we can avoid having to maximize the likelihood function. Instead, we assume that our model parameters are normal-distributed and that the variation in the data we observe reflects this. The Bayesian estimation procedure essentially samples from these distributions conditional on the observed data and any other prior information we have available. Train (2003) provides a more complete description of Bayesian estimation methods.

Let β be the vector of parameters we are looking to estimate. The likelihood of an individual i's employment history given β is

$$L(y_i|\beta) = \prod_t L(a_{it}) \tag{C.77}$$

However, we do not know the vector of parameters β. Instead, as mentioned previously, we assume that β is normal-distributed with means b and standard deviations W, both vectors themselves. Now, we can integrate over the distribution of β to get the likelihood of y_i not conditional on β.

$$L(y_i|b, W) = \int L(y_i|\beta)\phi(\beta|b, W)d\beta \tag{C.78}$$

where ϕ is the normal probability density function. By Bayes' theorem (C.*79)*, we can get the distributions of b and W conditional on the observed data Y. These are the so-called posterior distributions, which we denote $K(b, W|Y)$.

$$K(b, W|Y) = \frac{L(Y|b, W)k(b, W)}{L(Y)} = \prod_i \frac{L(y_i|b, W)k(b, W)}{L(y_i)} \tag{C.79}$$

Here, $k(b, W)$ are the prior distributions, representing our best guess for the distributions of b and W prior to the observations. Since we do not have any useful information about these beyond the data, we assume diffuse, i.e. uninformative, prior distributions for both. Specifically, we assume for b a normal distribution with unboundedly large variance and for W an Inverse

Wishart distribution with V degrees of freedom, where V is the number of parameters we are estimating, and scale matrix J, a V-dimensional identity matrix. In equation (C.*79)*, the denominator on the right-hand side $L(Y)$ is independent of the parameters b and W. Thus, we can simplify and replace the equality with a proportionality to get

$$K(b, W|Y) \propto \prod_i L(y_i|b, W)k(b, W) \tag{C.80}$$

Sampling directly from (C.*80)* using a Metropolis-Hastings (MH) algorithm is theoretically possible but still computationally difficult due to the non-analytical integral (C.*78)*, which must be solved at each iteration. However, we can avoid this by considering β_i, the realizations of the model parameters for each individual i, parameters in the Bayesian estimation procedure. In this case, the posterior distribution becomes

$$K(b, W, \beta_i \forall i|Y) = \prod_i L(y_i|\beta_i)\phi(\beta_i|b, W)k(b|W) \tag{C.81}$$

Now we can sample from this joint posterior distribution by repeating the following three-step Gibbs sampling algorithm:

1. Draw β_i conditional on b, W, and the observed data for individual.

$$K(\beta_i|b, W, y_i) \propto \prod_t P(a_{it}|\beta_i)\, \phi(\beta_i|b, W) \quad \forall i \tag{C.82}$$

   This step is the bottleneck of the sampler because evaluating $P(a_{it}|\beta_i)$ requires solving the dynamic programming problem with new parameters β_i for each person at each iteration. Furthermore, drawing from the posterior distribution (C.*82)* requires an MH algorithm, though fortunately without any non-analytical integrals.

2. Draw b conditional on W and β_i.

$$K(b|W, \beta_i \forall i) = N\left(\frac{\sum_i \beta_i}{I}, \frac{W}{I}\right) \tag{C.83}$$

   This is a draw from a simple normal distribution where I is the sample size.

3. Draw W conditional on b and β_i.

$$K(W|b, \beta_i \forall i) = IW\left(V + I, \frac{VJ + I\bar{S}}{V + I}\right) \tag{C.84}$$

This is a draw from the Inverse Wishart distribution where $\bar{S} = \frac{\sum_i (\beta_i - b)(\beta_i - b)'}{I}$, V is the number of parameters to estimated, i.e. the length of each β_i, and J is a V-dimensional identity matrix.

For illustrative purposes, I ran speed tests using the base model with 40 time steps, representing approximately the number of years between college graduation and retirement. Solving the dynamic programming problem with this model 100,000 times takes approximately five minutes on a single core on a typical laptop computer. Thus, drawing, say, 10,000 times from the posterior for a sample size I of 1,000 on the same quad-core laptop would take in the order of 2 hours. While somewhat cumbersome, this falls within the feasible range, provided that convergence is not too illusive. Furthermore, I will have access to several computing clusters at SMART FM, which can reduce the computation time by an order of magnitude, making the more complex model extensions feasible.

Challenges

Identification of the models relies fundamentally on job changes. Hence, it is crucial that the collected dataset covers, not just a sufficiently large number of individuals, but also follows them for long enough to observe changes in employment. The average monthly recruitment and resignation rates in the accommodation and food service sector since 2008 are 3.9% and 3.5%, respectively. For the professional, scientific, and technical services sector, the rates are 2.5% and 1.8% for recruitment and resignation, respectively. These rates are averages across firms without accounting for firm size, so they do not translate directly to individuals. However, even if 1.8% of people change employment per month, our 12-year dataset will capture an average of 2.6 changes per person. Furthermore, recent entrants to the labour market typically change jobs more frequently.

It is crucial that we subject our model to a series of sensitivity and robustness checks. Structural models are highly simplified versions of reality, relying heavily on parametric assumptions. In our case, this pertains in particular to the assumption about normal-distributed wage offers. Ideally this would be validated empirically. However, if that is infeasible I will test the sensitivity of the model to that assumption by considering alternative wage offer distributions. More broadly, we also need to consider the validity of the model's structural assumptions by devising tests for confounding factors and omitted variables.

Potential Research Avenues

This section describes several potential extensions to the spatial job search model proposed here that I find intellectually compelling and warrant additional consideration.

The cost of commuting

Travel costs are fundamental to any spatial equilibrium model. Their importance is also well-recognized in studies of urban transportation equity. However, the proposed models only consider travel impedances for determining accessibility as a measure of agglomeration; they do not offset wages. To incorporate the latter effect in a job search model, we need information the spatial distribution of the new job opportunities revealed individuals in each time period. Unfortunately, we generally do not observe these as discussed in Model Specification section. Hence, in practice we have to make a reasonable assumption about this distribution – analogous to the assumption about the wage offer distribution – and the interaction between the two. In other words, we would no longer draw wage offers from a normal distribution but from the joint normal distribution of wages and spatial distribution of job opportunity locations.

Information and communications technology (ICT)

The paradox of ICT is that it appears to simultaneously increase and decrease the importance of spatial impedance. It severely reduces the cost of communication and eliminates the need for face-to-face interactions in some context, while facilitating new trade linkages and collaborations that would not otherwise have existed. Examining the effect that ICT has had on the local nature of job search is technically straightforward but empirically difficult, since we require a comparable pre-ICT dataset.

Competition

The proposed models consider individuals in isolation. In other words, we do not account for competition for jobs or supply-demand interactions. Of course, this means that we must be careful in applying the findings to contexts with notably different labour supply or demand. However, more fundamentally, it also means that the proposed approach is not appropriate for project evaluation. For example, investment in a new subway line may improve employment outcomes for those living near it. However, their improved matches with employers likely happen at the expense of those who would have matched otherwise. Thus, while the improved access is a net positive, since more options improves total surplus in theory, neglecting competition means we do not accurately account for the external costs. We can potentially address this by formulating the number of offers each person receives in each time period as a function of job demand in addition to supply, see for example Petrongolo and Pissarides (2001) for inspiration on matching function specifications, or by using a measure of accessibility that accounts for competition.

APPENDIX D: COMMUTING MODE CHOICE MODEL

For determining the generalized cost of travel c_{ij} used in the agglomeration measures in equation (4.1), I estimate a commuting mode choice model and calculate the expected maximum utility (also known as logsum) between each zone pair. This simultaneously captures and summarizes the attributes of all modes and are thus an appropriate measure generalized cost measure. The mode choice model is based on the Massachusetts Travel Survey 2010-11. The dataset consists of 12030 respondents. Each respondent indicated their “usual mode to work” field and provide information about the location of both their home and workplace at the blockgroup level. Travel times and costs were extracted from the OD travel time matrices from the CTPS. The mode choice model itself is a multinomial logit with three choice alternatives: active modes, drive, and transit. Equation (D.1) shows the utility function. U_m is the total utility of mode choice m, comprising the systematic component V_m and the iid Gumbel-distributed error term ε. Equation (D.2) shows the resulting multinomial logit choice probability, i.e. the probability of choosing mode m given the choice set M.

$$U_m = V_m + \varepsilon \tag{D.1}$$

$$P(m) = \frac{e^{V_m}}{\sum_{m'}^{M} e^{V_{m'}}} \tag{D.2}$$

Equations (D.3) present the systematic utility function specification for each of the modes.

$$\begin{aligned} V_{active} &= \beta_0 \cdot distance \\ V_{drive} &= \beta_1 + \beta_2 \cdot time \\ V_{transit} &= \beta_3 + \beta_4 \cdot insystem\ time + \beta_5 \cdot access\ time + \beta_6 \cdot fare \end{aligned} \tag{D.3}$$

Table D-1 shows the estimated coefficients and their t-statistics. The model’s McFadden pseudo R^2 value is 0.22.

Table D-1: Mode choice model estimation results

Variable description	β	t
Active modes: Distance	-0.59	-66.26 ***
Drive: Alternative-specific constant	0.49	9.58 ***
Drive: Time	-0.13	-54.17 ***
Transit: Alternative-specific constant	3.43	28.45 ***
Transit: In-system travel time	-0.04	-14.52 ***
Transit: Access time	-0.18	-32.98 ***
Transit: Fare	-0.29	-5.88 ***

p-value: 0.00 *** 0.001 ** 0.01 * 0.05 ‘ 0.1

With these coefficients, I calculate logsums by equation (D.4).

$$EMU_{ij} = \ln \sum_{m}^{M} e^{V_m} \tag{D.4}$$

APPENDIX E: CONVERGENCE DIAGNOSTICS

Heidelberger and Welch (1983) propose a diagnostic to test the stationarity of a Markov chain. Specifically, they use the Cramér-von Mises test statistic to see if they can reject the null hypothesis that the chain is from a stationary distribution. The Cramér-von Mises criterion is defined as

$$\omega^2 = \int_{-\infty}^{\infty} [F_N(x) - F(x)]^2 dF(x) \tag{E.1}$$

where F_N is an empirical distribution function and F is a CDF to which it is being compared. For a set of observed values in ascending order $x_1, x_2, \ldots, x_n$, the Cramér-von Mises test statistic can be calculated as

$$T = n\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(\frac{2i-1}{2n} - F(x_i) \right)^2 \tag{E.2}$$

If T is greater than the corresponding tabulated value in Anderson (1962), the null hypothesis that the data came from the distribution F is rejected.

The Heidelberger and Welch diagnostic should be repeated for each parameter and typically comprises the following steps:

1. Choose a confidence level.
2. Generate a chain of N iterations.
3. Calculate the Cramer-von Mises test statistic for the whole chain with the chosen confidence level.
4. If the null hypothesis is rejected, remove the first 10% of the observations and recalculate the test statistic for the remainder of the chain.

5. Repeat 4 until either the null hypothesis cannot be rejected or 50% of the chain has been discarded. In the latter case, the chain has failed the Heidelberger and Welch diagnostic and requires more iterations.

For the purposes of this study, I apply the Cramer-von Mises test to the final 100,000 iterations of each chain (using every 100 observation), rather than iteratively removing 10% of the chain. If the null hypothesis can be rejected with at least 95% confidence, I continue the chain for longer. This approach is less dependent on the starting values. The final p-values are presented in the tables below.

Table E-1: P-values from Heidelberger-Welch diagnostic

Variable	Manufacturing	Professional	Food & Accommodation
TFP constant β_2	0.41 (0.64)	0.66 (0.23)	0.14 (0.54)
Employment size β_0	0.19 (0.20)	0.19 (0.28)	0.13 (0.34)
Land β_1	0.07 (0.30)	0.72 (0.65)	0.06 (0.07)
GDP β_3	0.93 (0.17)	0.92 (0.11)	0.37 (0.40)
Productivity β_4		0.05 (0.21)	0.26 (0.28)
Education β_5	0.25 (0.13)	0.14 (0.05)	0.17 (0.33)
Fixed adj. cost, hiring μ_0^+	0.40 (0.08)	0.19 (0.15)	0.10 (0.51)
Fixed adj. cost, firing μ_0^-	0.13 (0.05)	0.28 (0.39)	0.50 (0.23)
Linear adj. cost, hiring μ_1^+	0.25 (0.32)	0.44 (0.45)	0.46 (0.85)
Linear adj. cost, firing μ_1^-	0.24 (0.11)	0.17 (0.17)	0.68 (0.31)
Quadratic adj. cost, hiring μ_2^+	0.07 (0.14)	0.32 (0.46)	0.30 (0.20)
Quadratic adj. cost, firing μ_2^-	0.13 (0.24)	0.40 (0.20)	0.16 (0.33)
Zero-inflation ζ	0.99 (0.17)	0.91 (0.72)	0.53 (0.41)
Error st. dev. σ	0.51	0.92	0.76

Table E-2: P-values from Heidelberger-Welch diagnostic for manufacturing agglomeration parameters

Agglomeration measure (η)	**Manufacturing**
Specialization (walk only, 2.0)	0.37 (0.57)
Specialization (0.5)	0.11 (0.34)
Population (2.0)	0.75 (0.17)
Specialization (walk only, 2.0)	0.07 (0.35)

Table E-3: P-values from Heidelberger-Welch diagnostic for professional services agglomeration parameters

Agglomeration measure (η)	**Professional**
Population (0.25)	0.07 (0.19)
Employment (5.0)	0.06 (0.10)
Population (0.25)	0.23 (0.06)
Specialization (5.0)	0.06 (0.87)

Table E-4: P-values from Heidelberger-Welch diagnostic for food and accommodation services agglomeration parameters

Agglomeration measure (η)	**Food and accommodation**
Population (1.0)	0.25 (0.23)
Employment (walk only, 5.0)	0.11 (0.05)
Specialization (walk only, 5.0)	0.26 (0.12)
Population (1.0)	0.06 (0.21)

A second diagnostic, the Gelman-Rubin convergence diagnostic, ensures that multiple chains of the same variable converge to the same value by comparing within and between chain variances. The mean within-chain variance W is

$$W = \frac{1}{m}\sum_{j=1}^{m} s_j^2 \tag{E.3}$$

where s_j^2 is the within-chain variance of each chain j, calculated as

$$s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\theta_{ij} - \bar{\theta}_j\right)^2 \tag{E.4}$$

The between chain variance B is

$$B = \frac{n}{m-1}\sum_{j=1}^{m}\left(\bar{\theta}_j - \bar{\bar{\theta}}\right)^2 \tag{E.5}$$

where $\bar{\bar{\theta}}$ is the mean across all chains.

$$\bar{\bar{\theta}} = \frac{1}{m}\sum_{j=1}^{m} \bar{\theta}_j \tag{E.6}$$

The statistic of interest is called the potential scale reduction factor $\hat{R}$, and is calculated as

$$\hat{R} = \sqrt{\frac{\left(1 - \frac{1}{n}\right) W + \frac{1}{n} B}{W}} \tag{E.7}$$

If this value is large, say greater than 1.2, the chains require more iterations for convergence (Gelman & Rubin, 1992).

Table E-5: $\hat{R}$-values from Gelman-Rubin diagnostic

Variable	**Manufacturing**	**Professional**	**Food & Accommodation**
TFP constant β_2	1.18 (1.12)	1.09 (1.01)	1.12 (1.13)
Employment size β_0	1.02 (1.03)	1.09 (1.01)	1.06 (1.20)
Land β_1	1.06 (1.01)	1.03 (1.07)	1.07 (1.13)
GDP β_3	1.06 (1.16)	1.03 (1.12)	1.02 (1.10)
Productivity β_4		1.10 (1.14)	1.01 (1.04)
Education β_5	1.02 (1.07)	1.01 (1.10)	1.00 (1.01)
Fixed adj. cost, hiring μ_0^+	1.14 (1.13)	1.14 (1.11)	1.10 (1.17)
Fixed adj. cost, firing μ_0^-	1.17 (1.18)	1.12 (1.10)	1.13 (1.17)
Linear adj. cost, hiring μ_1^+	1.11 (1.15)	1.09 (1.19)	1.05 (1.16)
Linear adj. cost, firing μ_1^-	1.10 (1.11)	1.09 (1.12)	1.08 (1.10)
Quadratic adj. cost, hiring μ_2^+	1.09 (1.13)	1.07 (1.09)	1.17 (1.09)
Quadratic adj. cost, firing μ_2^-	1.08 (1.08)	1.09 (1.05)	1.01 (1.01)
Zero-inflation ζ	1.03 (1.02)	1.00 (1.11)	1.00 (1.04)
Error st. dev. σ	1.00	1.01	1.05

Table E-6: $\hat{R}$-values from Gelman-Rubin diagnostic for manufacturing firm agglomeration parameters

Agglomeration measure (η)	**Manufacturing**
Specialization (walk only, 2.0)	1.05 (1.03)
Specialization (0.5)	1.02 (1.04)
Population (2.0)	1.00 (1.09)
Specialization (walk only, 2.0)	1.12 (1.12)

Table E-7: $\hat{R}$-values from Gelman-Rubin diagnostic for professional services firm agglomeration parameters

Agglomeration measure (η)	**Professional**
Population (0.25)	1.16 (1.09)
Employment (5.0)	1.04 (1.04)
Population (0.25)	1.02 (1.04)
Specialization (5.0)	1.06 (1.03)

Table E-8: $\hat{R}$-values from Gelman-Rubin diagnostic for food and accommodation services firm agglomeration

Agglomeration measure (η)	**Food and accommodation**
Population (1.0)	1.16 (1.14)
Employment (walk only, 5.0)	1.19 (1.04)
Specialization (walk only, 5.0)	1.02 (1.07)
Population (1.0)	1.10 (1.02)

BIBLIOGRAPHY

Acheampong, R. A., & Silva, E. (2015). Land use–transport interaction modeling: A review of the literature and future research directions. *Journal of Transport and Land Use*, *8*(3). https://doi.org/http://dx.doi.org/10.5198/jtlu.2015.806

Adnan, M., Pereira, F. C., Miguel, C., Azevedo, L., Basak, K., Lovric, M., … Ben-akiva, M. E. (2015). SimMobility : A Multi-Scale Integrated Agent-based Simulation Platform. In *Transportation Research Board (TRB) 95th Annual Meeting*. 1–18.

Aguirregabiria, V., & Mira, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, *156*(1), 38–67. https://doi.org/10.1016/j.jeconom.2009.09.007

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2015). The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, *83*(6), 2127–2189. https://doi.org/10.3982/ecta10876

Allied. (2012). *2012 Allied Workforce Mobility Survey: Onboarding and Retention*.

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent*.

Cambridge, MA and London, England: Harvard University Press.

https://doi.org/10.4159/harvard.9780674730854

Anas, A. (1983). Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research Part B*, *17*(1), 13–23. https://doi.org/10.1016/0191-2615(83)90023-1

Anas, A., & Liu, Y. (2007). A regional economy, land use, and transportation model (RELU-TRAN©): Formulation, algorithm design, and testing. *Journal of Regional Science*, *47*(3), 415–455. https://doi.org/10.1111/j.1467-9787.2007.00515.x

Anderson, T. W. (1962). On the Distribution of the Two-Sample Cramer-von Mises Criterion. *The Annals of Mathematical Statistics*, *33*(3), 1148–1159. https://doi.org/10.1214/aoms/1177704477

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics*. Princeton University Press.

Arzaghi, M., & Henderson, J. V. (2008). Networking off Madison Avenue. *Review of Economic Studies*, *75*(4), 1011–1038. https://doi.org/10.1111/j.1467-937X.2008.00499.x

Asparouhov, T., & Muthén, B. (2010). Bayesian analysis of latent variables models using Mplus. *Mplus Web Notes*, 1–60.

Baum-Snow, N., & Pavan, R. (2012). Understanding the city size wage gap. *Review of Economic Studies*, *79*(1), 88–127. https://doi.org/10.1093/restud/rdr022

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*. MIT Press.

Bertaud, A. (2014). Cities as labor markets. Marron Institute on Cities and the Urban Environment, New York University.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(17), 7301–7306. https://doi.org/10.1073/pnas.0610172104

Borysenko, K. (2014). The Cost of Employee Turnovers. Retrieved from https://www.linkedin.com/pulse/20140701121556-17497251-the-cost-of-employee-turnover/

Buchanan, J. M. (1965). An Economic Theory of Clubs. *Economica*, *32*(125), 1–14. Retrieved from http://www.jstor.org/stable/2548575

Carlino, G., & Kerr, W. R. (2015). *Agglomeration and Innovation*. *Handbook of Regional and Urban Economics* (Vol. 5). Elsevier. https://doi.org/10.1016/B978-0-444-59517-1.00006-4

Ciccone, A., & Hall, R. E. (1993). Production and density of economic activity. *NBER Working Paper Series*.

Central Transportation Planning Staff. (2007) *OD Travel Time and Cost Skims* [Data set].

Central Transportation Planning Staff. (2012). *Massachusetts Travel Survey 2010-11* [Data set].

Chow, Y. S., Robbins, H., & Siegmund, D. Great Expectations: The Theory of Optimal Stopping. 1971.

Coles, M. G., & Smith, E. (1998). Marketplaces and Matching. *International Economic Review*, *39*(1), 239. https://doi.org/10.2307/2527239

Combes, P. P., & Gobillon, L. (2015). The Empirics of Agglomeration Economies. *Handbook of Regional and Urban Economics* (Vol. 5, pp. 247-348). Elsevier. https://doi.org/10.1016/B978-0-444-59517-1.00005-2

Damodaran, A. (2021). *Cost of equity and capital (US)*. [Data set]. Damodaran Online.

http://people.stern.nyu.edu/adamodar/New_Home_Page/datafile/wacc.html

Davidson, N. M., & Infranca, J. J. (2016). The Sharing Economy as an Urban Phenomenon. *Yale Law & Policy Review*, *34*(2), 215–280.

de Bok, M. (2009). Estimation and validation of a microscopic model for spatial economic effects of transport infrastructure. *Transportation Research Part A: Policy and Practice*, *43*(1), 44–59. https://doi.org/10.1016/j.tra.2008.06.002

de la Roca, J., & Puga, D. (2017). Learning by working in big cities. *Review of Economic Studies*, *84*(1), 106–142. https://doi.org/10.1093/restud/rdw031

Di Addario, S., & Patacchini, E. (2008). Wages and the City. Evidence from Italy. *Labour Economics*, *15*(5), 1040–1061. https://doi.org/10.1016/j.labeco.2007.09.003

Diamond, P. A., & Maskin, E. (1965). *An equilibrium analysis of search and breach of contract*. *The British Journal of Psychiatry* (Vol. 111). https://doi.org/10.1192/bjp.111.479.1009-a

Duranton, G., & Puga, D. (2001). Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products. *American Economic Review*, *91*(5), 1454–1477.

Graaff, T. D., & Zondag, B. (2013). A Population-Employment Interaction Model as Labour Module in TIGRIS XL. In Employment Location in Cities and Regions (pp. 57-77). Springer, Berlin, Heidelberg.

de la Barra, T. (2011). TRANUS : Integrated Land Use and Transport Modeling System, 23. Retrieved from http://www.tranus.com/

Duranton, G., & Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, *91*(5), 1454–1477. https://doi.org/10.1257/aer.91.5.1454

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies.

Handbook of regional and urban economics (Vol. 4, pp. 2063-2117). Elsevier.

Durlauf, S. N. (2004). Neighborhood effects. *Handbook of Regional and Urban Economics* (Vol. 4). Elsevier Inc. https://doi.org/10.1016/S1574-0080(04)80007-5

Echenique, M. H., Flowerdew, A. D. J., Hunt, J. D., Mayo, T. R., Skidmore, I. J., & Simmonds, D. C. (1990). The Meplan models of bilbao, Leeds and Dortmund: Foreign summaries. *Transport Reviews*, *10*(4), 309–322. https://doi.org/10.1080/01441649008716764

Ellison, G., Glaeser, E., & Kerr, W. (2010). What Causes Industry Agglomeration? *American Economic Review*, *100*(3), 1195–1213.

Engelberg, D., He, H., Le, D. T., & Zegras, P. C. (2021). Accessibility, land use models, and modeling. *Urban Form and Accessibility*, 379-409. Elsevier.

Faggian, A. (2014). Job Search Theory. In M. M. Fischer & P. Nijkamp (Eds.), Handbook of Regional Science (pp. 59–73). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-23430-9_8

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–511.

Glaeser, E. L. (1998). Are Cities Dying? *Journal of Economic Perspectives*, *12*(2), 139–160. https:// Glaeser, E. L. (1998). Are Cities Dying? *Journal of Economic Perspectives*, *12*(2), 139–160. https://doi.org/10.1257/jep.12.2.139

Glaeser, E. L. (1999). Learning in Cities. *Journal of Urban Economics*. https://doi.org/10.1006/juec.1998.2121

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, *100*(6), 1126–1152. Retrieved from http://www.nber.org/papers/w3787

Glaeser, E. L., & Maré, D. C. (1994). *Cities and Skills* (No. 4728). *NBER Working paper series*.

Glaeser, E. L., Laibson, D., and Sacerdote, B. (2002). An economic approach to social capital. *Economic Journal*, *112*(483), 437–458. https://doi.org/10.1111/1468-0297.00078

Gould, E. D. (2007). Cities, workers, and wages: A structural analysis of the urban wage premium. *Review of Economic Studies*, *74*(2), 477–506. https://doi.org/10.1111/j.1467-937X.2007.00428.x

Hansen, W. G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, *25*(2), 73–76. https://doi.org/10.1080/01944365908978307

Harris, C. D. (1954). The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers*, *44*(4), 315–348. https://doi.org/10.1080/00045605409352140

Heckman, J. (1979). Sample selection bias as a specification error. *Applied Econometrics*, *47*(1), 153–161. https://doi.org/10.2307/1912352

Helsley, R. W. (1990). Knowledge and production in the CBD. *Journal of Urban Economics*, *28*(3), 391–403. https://doi.org/10.1016/0094-1190(90)90035-L

Helsley, R. W., & Strange, W. C. (1990). Matching and agglomeration economies in a system of cities. *Regional Science and Urban Economics*, *20*(2), 189–212. https://doi.org/10.1016/0166-0462(90)90004-M

Henderson, V., Kuncoro, A., & Turner, M. A. (1995). *Industrial Development in Cities*. *NBER Working paper series*. https://doi.org/10.1038/142699c0

Heidelberger, P., & Welch, P. D. (1983). Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research*, *31*(6), 1109–1144. https://doi.org/10.1287/opre.31.6.1109

Holl, A. (2012). Market potential and firm-level productivity in Spain. *Journal of Economic Geography*, *12*(6), 1191–1215. https://doi.org/10.1093/jeg/lbr030

Holmes, T. J., & Sieg, H. (2015). Structural Estimation in Urban Economics. *Handbook of Regional and Urban Economics* (Vol. 5). Elsevier. https://doi.org/10.1016/B978-0-444-59517-1.00002-7

Hunt, J. D., & Abraham, J. . (2005). Design and Implementation of PECAS: A Generalised System for Allocating Economic Production, Exchange and Consumption Quantities. *Integrated Land-Use and Transportation Models*, 253–273. https://doi.org/10.1108/9781786359520-011

Hunt, J. D., Kriger, D. S., & Miller, E. J. (2005). Current operational urban land-use-transport modelling frameworks: A review. *Transport Reviews*, *25*(3), 329–376. https://doi.org/10.1080/0144164052000336470

Iacono, M., Levinson, D., & El-Geneidy, A. (2008). Models of transportation and land use change: A guide to the territory. *Journal of Planning Literature*, *22*(4), 323–340. https://doi.org/10.1177/0885412207314010

Immergluck, D. (1998). *Job proximity and the urban employment problem: Do suitable nearby jobs improve neighbourhood employment rates? Urban Studies* (Vol. 35). https://doi.org/10.1080/0042098983917

InfoGroup. (2014). *ReferenceUSA Business Historical Data Files* (V8) [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/GW2P3G

Jacobs, J. (1961). *The Death and Life of Great American Cities*.

Jin, Y., & Wegener, M. (2013). Beyond equilibrium. Environment and Planning B: Planning and Design, 40(6), 951–954. https://doi.org/10.1068/b4006ge

Karlstrom, A., Palme, M., & Svensson, I. (2004). A dynamic programming approach to model the retirement behaviour of blue-collar workers in Sweden. *Journal of Applied Econometrics*, *19*(6), 795–807. https://doi.org/10.1002/jae.798

Keane, M. P., & Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, *105*(3), 473–522. https://doi.org/10.1086/262080

Kennan, J. (1979). The Estimation of Partial Adjustment Models with Rational Expectations. *Econometrica*, *47*(6), 1441. https://doi.org/10.2307/1914011

King, R. G., & Thomas, J. K. (2006). Partial adjustment without apology. *International Economic Review*, *47*(3), 779–809. https://doi.org/10.1111/j.1468-2354.2006.00396.x

Krugman, P. R. (1991). *Geography and Trade*. MIT Press.

Leontief, W. (1966). *Input-output economics*. Oxford University Press.

Ljungqvist, L., & Sargent, T. J. (2012). *Recursive Macroeconomic Theory*. Cambridge, Massachusetts: MIT Press. https://doi.org/10.1192/bjp.112.483.211-a

Lowry, I. S. (1964). *A Model of Metropolis*. *Concepts and Techniques in Urban Analysis*. Santa Monica: The RAND Corportation. https://doi.org/10.4324/9781315105352-12

Marshall, A. (1890). *Principles of Economics*. New York: Random House.

Massachusetts Bureau of Geographic Information. (2019). *Massachusetts Land Parcel Database (Metro Boston Region)* [Data set]

Matano, A., & Naticchioni, P. (2012). Wage distribution and the spatial sorting of workers. *Journal of Economic Geography*, *12*(2), 379–408. https://doi.org/10.1093/jeg/lbr013

Melo, P. C., Graham, D. J., & Noland, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, *39*(3), 332–342. https://doi.org/10.1016/j.regsciurbeco.2008.12.002

Metropolitan Area Planning Council. (2009). *MAPC MetroFuture Population and Employment Projections* [Data set].

Metro Research Center. (2016). *MetroScope 4.0 Generalized Design and Equation Specifications as Implemented at Portland Metro*.

Miller, E. J. (2018). Integrated urban modeling: Past, present, and future. *Journal of Transport and Land Use*, *11*(1), 387–399. https://doi.org/10.5198/jtlu.2018.1273

Miller, R. A. (1984). *Job Matching and Occupational Choice*. *The Journal of Political Economy* (Vol. 92). Retrieved from http://www.comlabgames.com/ramiller/jobMatchandOccupChoic.pdf

Moeckel, R. (2009). Simulation of firms as a planning support system to limit urban sprawl of jobs. *Environment and Planning B: Planning and Design*, *36*(5), 883–905. https://doi.org/10.1068/b34110

Moretti, E. (2011). Local labor markets. *Handbook of Labor Economics*, *4*(PART B), 1237–1313. https://doi.org/10.1016/S0169-7218(11)02412-9

Mortensen, D. T. (1978). *Specific Capital and Labor Turnover*. *The Bell Journal of Economics* (Vol. 9).

Mortensen, D. T. (1986). Job search and labor market analysis. *Handbook of Labor Economics*, *2*(C), 849–919. https://doi.org/10.1016/S1573-4463(86)02005-9

Mostafa, T. S. (2017). *A microsimulation platform of firm evolution processes*. University of Toronto.

Ottaviano, G., & Thisse, J.-F. (2004). Agglomeration and Economic Geography. *Handbook of regional and urban economics* (Vol. 4, pp. 2563-2608). Elsevier. https://doi.org/10.1016/S1574-0080(04)80015-4

Pakes, A. (1986). Patents as options: Some estimates of the value of holding European patent stocks. *Econometrica*, (54), 755–784.

Petrongolo, B., & Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, *39*(2), 390–431. https://doi.org/10.1257/jel.39.2.390

Petrongolo, B., & Pissarides, C. A. (2003). Scale effects in markets with search. *Economic Journal*, (508), 21–44. https://doi.org/10.1111/j.1468-0297.2006.01046.x

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, *50*(1), 203–219. https://doi.org/10.1111/j.1467-9787.2009.00657.x

Ravulaparthy, S. K., Selby, B., Kuppam, A., Jeon, K., Nippani, S., & Livshits, V. (2017). Spatial Pissarides, C. A. (1979). *Job Matchings with State Employment Agencies and Random Search*. *Source: The Economic Journal* (Vol. 89).

firm demographic microsimulator: Development and validation for phoenix and Tucson, Arizona, Megaregion. *Transportation Research Record*, *2671*, 59–70. https://doi.org/10.3141/2671-07

Rosen, S. (1983). Specialization and Human Capital. *Journal of Labor Economics*, *1*(1), 43–49. https://doi.org/10.2307/j.ctv1kz4h99.11

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. *Handbook of Regional and Urban Economics*, (Vol. 4, pp. 2119–2171). Elsevier. https://doi.org/10.1016/S0169-7218(04)07049-2

Rosenthal, S. S., & Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, *64*(2), 373–389. https://doi.org/10.1016/j.jue.2008.02.006

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, *55*(5), 999. https://doi.org/10.2307/1911259

Rust, J. (1994). Structural estimation of Markov decision processes. In *Handbook of Econometrics* (Vol. 4). Elsevier.

Rust, J., & Phelan, C. (1997). How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets. *Econometrica*, *65*(4), 781. https://doi.org/10.2307/2171940

Schuurman, N. K., Grasman, R. P. P. P., & Hamaker, E. L. (2016). A Comparison of Inverse-Wishart Prior Specifications for Covariance Matrices in Multilevel Autoregressive Models. *Multivariate Behavioral Research*. https://doi.org/10.1080/00273171.2015.1065398

Scotchmer, S. (2002). Local public goods and clubs. *Handbook of Public Economics*, *4*, 1997–2042. https://doi.org/10.1016/S1573-4420(02)80008-8

SHRM. (2016). *2016 Human Capital Benchmarking Report*. Retrieved from https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Documents/2016-Human-Capital-Report.pdf

Simmonds, D., Waddell, P., & Wegener, M. (2013). Equilibrium versus dynamics in urban modelling. Environment and Planning B: Planning and Design, 40(6), 1051–1070. https://doi.org/10.1068/b38208

Sobel, J. (2000). Economists' Models of Learning. *Journal of Economic Theory*, *94*, 241–261.

Starrett, D. (1978). Market allocations of location choice in a model with free mobility. *Journal of Economic Theory*, *17*(1), 21–37. https://doi.org/10.1016/0022-0531(78)90120-5

Stigler, G. J. (1962). Information in the Labor Market. *Journal of Political Economy*, *70*(5), 95–105.

Timmermans, H. J. P. (2003). The saga of integrated land use-transport modeling: how many more dreams before we wake up? In *10th International Conference on Travel Behaviour Research*. Lucerne. Retrieved from https://edit.ethz.ch/ivt/news/archive/20030810_IATBR/timmermans.pdf

Train, K. E. (2003). Discrete choice methods with simulation. Discrete Choice Methods with Simulation (Vol. 9780521816). https://doi.org/10.1017/CBO9780511753930

Topa, G., & Zenou, Y. (2015). Neighborhood and Network Effects. *Handbook of Regional and Urban Economics* (Vol. 5). Elsevier. https://doi.org/10.1016/B978-0-444-59517-1.00009-X

U.S. Bureau of Economic Analysis. (2021a). *CAEMP25N Total Full-Time and Part-Time Employment by NAICS Industry.* https://apps.bea.gov/iTable/iTable.cfm?reqid=70&step=30&isuri=1&year_end=-1&acrdn=5&classification=naics&state=5&yearbegin=-1&unit_of_measure=levels&major_area=5&area=14460&year=-1&tableid=33&category=733&area_type=4&statistic=-1

U.S. Bureau of Economic Analysis. (2021b). *CAGDP9 Real GDP by country and metropolitan area.* https://apps.bea.gov/iTable/iTable.cfm?reqid=70&step=30&isuri=1&year_end=-1&acrdn=5&classification=naics&state=5&yearbegin=-1&unit_of_measure=levels&major_area=5&area=14460&year=-1&tableid=503&category=2503&area_type=5&statistic=-1

U.S. Bureau of Economic Analysis. (2021c). *CAINC6N Compensation of Employees by NAICS Industry.* https://apps.bea.gov/iTable/iTable.cfm?reqid=70&step=30&isuri=1&year_end=-1&acrdn=5&classification=naics&state=5&yearbegin=-1&unit_of_measure=levels&major_area=5&area=14460&year=-1&tableid=55&category=755&area_type=4&statistic=-1

U.S. Bureau of Economic Analysis. (2021d). *Table 6.5D. Full-Time Equivalent Employees by Industry.* https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=3&isuri=1&select_all_years=0&nipa_table_list=197&series=q&first_year=2003&last_year=2013&scale=-99&categories=survey&thetable=

U.S. Bureau of Economic Analysis. (2021e). *Real Value Added by Industry.*
https://apps.bea.gov/iTable/iTable.cfm?reqid=150&step=3&isuri=1&table_list=10&series=a&fir

st_year=2003&columns=ii&scale=-

9&last_year=2013&categories=gdpxind&thetable=&rows=22r

U.S Bureau of Labor Statistics. (2022). *Business Employment Dynamics.* https://data.bls.gov/PDQWeb/bd

van Wissen, L. (2000). A micro-simulation model of firms: Applications of concepts of the demography of the firm. *Papers in Regional Science*, *79*(2), 111–134. https://doi.org/10.1007/s101100050039

von Thünen, J. H. (1826). *The Isolated State*.

Wegener, M. (2004). Overview of Land-use Transport Models. In D. A. Hensher & K. Button (Eds.), *Transport Geography and Spatial Systems* (pp. 127–146). Kidlington, UK: Pergamon/Elservier. https://doi.org/10.1038/nrc1306

Wegener, M. (2011). *The IRPUD Model*. *Spiekermann & Wegener Urban and Regional Research Working Papers* (Vol. 11). https://doi.org/10.1016/S0032-5910(99)00010-8

Wegener, M., Gnad, F., & Vannahme, M. (1986). The time scale of urban change. Advances in Urban Systems Modelling, 175–197.

Wikipedia. (2022). *File:MBTA Green Line.svg.* https://en.wikipedia.org/wiki/File:MBTA_Green_Line.svg

Wolpin, K. I. (1984). An Estimable Dynamic Stochastic Model of Fertility and Child Mortality. *Journal of Political Economy*, *92*(5), 852–874. https://doi.org/10.1086/261262

Zhuge, C., & Shao, C. (2019). Agent-based modelling of office market for a land use and transport model. *Transportmetrica B*. https://doi.org/10.1080/21680566.2019.1589599

Zondag, B., de Bok, M., Geurs, K. T., & Molenwijk, E. (2015). Accessibility modeling and evaluation: The TIGRIS XL land-use and transport interaction model for the Netherlands. *Computers*, *Environment and Urban Systems*, *49*, 115–125. https://doi.org/10.1016/j.compenvurbsys.2014.06.001