

Opportunity &  
Inclusive Growth  
INSTITUTE



FEDERAL RESERVE BANK OF MINNEAPOLIS



SYSTEM WORKING PAPER  
No. 20-08

# The Geography of Business Dynamism and Skill Biased Technical Change

August 2020

---

**Hannah Rubinton**

*Federal Reserve Bank of St. Louis*

The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

This paper was originally published as Working Paper 2020-020A by the Federal Reserve Bank of St. Louis. This paper may be revised. The most current version is available at <https://doi.org/10.20955/wp.2020.020>.



# RESEARCH DIVISION

*Working Paper Series*

## **The Geography of Business Dynamism and Skill Biased Technical Change**

**Hannah Rubinton**

Working Paper 2020-020A  
<https://doi.org/10.20955/wp.2020.020>

July 2020

**FEDERAL RESERVE BANK OF ST. LOUIS**

Research Division

P.O. Box 442

St. Louis, MO 63166

---

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Federal Reserve Bank of St. Louis Working Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.

# The Geography of Business Dynamism and Skill-Biased Technical Change\*

Hannah Rubinton<sup>†</sup>

July 21, 2020

*[Click here for the latest version](#)*

## Abstract

This paper seeks to explain three key components of the growing regional disparities in the U.S. since 1980, referred to as the Great Divergence by [Moretti \(2012\)](#). Namely, big cities saw a larger increase in the relative wages of skilled workers, a larger increase in the relative supply of skilled workers, and a smaller decline in business dynamism. These trends can be explained by differences across cities in the extent to which firms adopt new skill-biased technologies. In response to the introduction of a new skill-biased, high fixed cost but low marginal cost technology, firms endogenously adopt more in big cities, in cities that offer abundant amenities for high-skilled workers and in cities that are more productive in using high-skilled labor. The differences in adoption can account for the increasing relationship between skill intensity and city size, the divergence of the city size wage premium by skill group and the changing cross-sectional patterns of business dynamism. I document a new fact that firms in big cities invest more in Information and Communication Technology per employee than firms in small cities, consistent with patterns of technology adoption in the model.

---

\*I am extremely grateful to my advisor Richard Rogerson and to my committee, Ezra Oberfield, Teresa Fort and Stephen Redding, for their invaluable help and guidance. For helpful comments, I thank Esteban Rossi-Hansberg, Gene Grossman, Eduardo Morales, Kamran Bilir, Mark Aguiar, Gianluca Violante, Oleg Itskhoki, Nobu Kiyotaki, Gregor Jarosch, Jonathan Dingel, Rebecca Diamond, Chris Tonetti, Benjamin Pugsley, Andrew Bernard, David Chor, Peter Schott, Ethan Lewis, Bob Staiger, Nina Pavcnik, Treb Allen, Laura Castillo-Martinez, Elisa Giannone, Fabian Eckert, Gideon Bornstein, and other visitors to and participants of the IES Student Seminar Series. Thank you to my classmates for their continuous support and encouragement. This research benefited from financial support from the International Economics Section (IES) and the Simpson Center for the Study of Macroeconomics. Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1975. (CBDRB-FY20-332)"

<sup>†</sup>Federal Reserve Bank of St. Louis; Email: [hannah.rubinton@stls.frb.org](mailto:hannah.rubinton@stls.frb.org)

# 1 Introduction

Since 1980, economic growth in the United States has been concentrated in large, urban areas while small rural areas have fallen further behind. [Moretti \(2012\)](#) refers to this growing gap as the Great Divergence, which has become an increasing concern for policy makers as they have learned of its implications for gaps in health outcomes, disintegrating social connections, economic mobility, and political tension<sup>1</sup>. For policies to effectively address these trends, it is necessary to understand the underlying causes.

This paper seeks to explain three components of the Great Divergence. First, big cities saw a larger increase in the relative wages of skilled workers. Second, big cities saw a larger increase in the relative supply of skilled workers. Third, big cities saw a smaller decline in business dynamism, as measured by the rates of establishment entry and exit and the rates of job creation and destruction<sup>2</sup>. While big and small cities had similar rates of business dynamism in 1980, big cities are more dynamic than small cities today. The simultaneous increase in high-skilled wage growth and the abundance of high-skilled workers is typically explained at the aggregate level by skill-biased technical change (SBTC) ([Katz and Murphy, 1992](#)). But, this aggregate story does not address the question of why SBTC would have occurred differently across cities.

To explain these facts I develop a spatial model in which the amount of SBTC will endogenously vary across cities. I consider the introduction of a new technology that lowers the marginal cost for a firm but has a higher fixed cost and uses high-skilled labor more intensively. The key insight from the model is that firms face different incentives to adopt a new technology based on the characteristics of the city in which they are located. Firms will adopt more in cities that are big, cities with amenities attractive to high-skilled workers, and cities that are more productive in using high-skilled labor. The extent of adoption, and therefore, SBTC endogenously varies across cities, explaining the changing relationships between wages and skill intensity with city size.

Differences in adoption also affect rates of business dynamism across cities. In cities with high adoption rates, small and unproductive firms are less profitable, increasing their probability of exit (a selection effect similar to [Melitz, 2003](#)), which increases the equilibrium rate of turnover. Since these less-productive firms are using the old technology, which places more weight on low-skilled labor, the increase in selection amplifies the differences across cities in the share of firms adopting the new technology, and therefore, SBTC.

The changing relationship between these economic variables with respect to city-size has important implications for welfare inequality, economic mobility, and declining regional convergence ([Diamond \(2016\)](#), [Moretti \(2013\)](#), [Giannone \(2017\)](#), and [Autor \(2019\)](#)). Understanding the drivers of these trends is key to policy makers who wish to influence them. I make progress by showing that differences in technology adoption across cities will amplify existing geographic inequalities along

---

<sup>1</sup>For papers that discuss these implications see [Austin et al. \(2018\)](#), [Chetty and Hendren \(2018\)](#), [Chetty et al. \(2016\)](#), [Autor et al. \(2019\)](#), and [Autor et al. \(2016\)](#).

<sup>2</sup>While facts one and two have been previously documented by the literature ([Autor, 2019](#); [Baum-Snow et al., 2018](#); [Giannone, 2017](#)), the third is a new fact that I introduce to the literature on the Great Divergence.

the dimensions for which a new technology is most suitable.

My analysis proceeds in three steps. First, I document empirical changes in the distribution of economic activity across cities since 1980. I focus on three cross-sectional relationships of interest. First, I introduce a new fact to the literature on the Great Divergence, the changing relationship between business dynamism and city-size since 1980. In 1980, these measures were similar in big and small cities. By 2014, big cities exhibited much faster rates of dynamism than small cities. Second, I document the changing correlation between average wages and city-size, referred to as the city-size wage premium, by skill group. While the city-size wage premium was similar for high- and low-skilled workers in 1980, by 2014, the city-size wage premium for high-skilled workers was almost twice that of low-skilled workers. This divergence was driven by both an increasing city-size wage premium for high-skilled workers and a decreasing city-size wage premium for low-skilled workers. Third, I document that the relationship between skill intensity, or the ratio of high- to low-skilled workers, and city-size has increased since 1980.

My second contribution is to build a model that matches the salient features of the data in 1980. I embed a rich model of firm dynamics into an otherwise standard spatial equilibrium model with high- and low-skilled workers, allowing a joint consideration of the geographic distribution of relative wage inequality and firm dynamics. The model features two types of workers, high- and low-skilled, who are freely mobile across space. Workers pay rent to live and work in a city and receive an amenity specific to the city where they live and their skill level. Workers have idiosyncratic preferences for each city and choose to live in the city that gives them the highest utility. Within each city there is a continuum of monopolistically competitive firms that use high- and low-skilled labor to produce non-tradable intermediate goods. Firms pay a fixed cost in units of high- and low-skilled labor, and they pay rent on a unit of floor space in the city where they produce. Firms receive idiosyncratic productivity shocks and, therefore, make dynamic entry and exit decisions. Entrants choose the city where they want to enter and produce. In equilibrium, firms are indifferent between entering in different cities. I calibrate the model steady state to the data in 1980 and show that it can match the key features of the data.

My third contribution is to use the model to analyze the diffusion of a new technology that favors skilled workers. I consider the introduction of a new technology that has an absolute productivity advantage but is more skill-biased in that the marginal productivity of high-skilled labor is higher than that of the old technology. Firms can choose to adopt the new technology, and, depending on the factor costs in their city, adopting may lower their marginal cost. However, the new technology requires a higher fixed cost. Even though the new technology is available everywhere, firms that are ex-ante similar will make different technology adoption decisions depending on the environment in their city.

After introducing the new skill-biased technology and allowing the aggregate supply of high- and low-skilled labor to adjust as it has in the data, I solve for a new steady-state equilibrium in the model. I compare the model steady states before and after the introduction of the new technology and show that the model can match the key changes in the data, including the changing relationship

of skill intensity and firm dynamics to city-size and the divergence of the city-size wage premium by skill group.

A key feature of the response to the new technology is that adoption is higher in big cities. The literature has emphasized Information and Communication Technologies (ICT) as being intimately related to skill-biased technologies (Krueger, 1993; Autor et al., 1998). Thus, as further evidence for differences in technology adoption across cities, I document a new cross-sectional fact. Using data on ICT investment between 2003 and 2013, I show that firms in big cities have more ICT-related expenditures per employee than firms in small cities, and firms in big cities spend a higher share of their total investment budget on ICT, consistent with the patterns of technology adoption in the model.

In the model, there are four channels that drive the higher rates of technology adoption in the big city. The first channel is the market size effect. For a given productivity, firms in big cities receive a higher volume of sales and are more willing to pay the additional fixed cost to achieve the marginal cost saving from adopting the new technology. The second channel driving differences in adoption is ex-ante differences in technology between the cities. In 1980, firms in big cities were already more productive at using high-skilled labor than firms in small cities. Even though the new technology keeps the relative skill intensities of the old and new production functions constant across cities, firms will be more likely to adopt in cities that are, ex-ante, more productive in using high-skilled labor. The third channel that drives differences in adoption are differences in amenities. If a city is rich in amenities for high-skilled workers, then all else equal, there will be a lower relative wage for high-skilled workers, which further increases the return to adoption. Finally, a fourth channel affecting adoption rates is selection. In cities where firms adopt more, the exit threshold below which firms exit the market shifts up, meaning that selection becomes tougher. Small, less-productive firms are better at using low-skilled labor than big firms that have adopted the new technology. When selection becomes tougher, the small firms exit the market amplifying the amount of SBTC that the city experiences. This is the same force that drives up equilibrium dynamism rates in big cities relative to small cities.

Finally, I use the model for a series of counterfactuals. The first counterfactual examines the effect of the increased sorting of high-skilled workers to big cities. Specifically, I shut down the migration channel in the model and look at what would happen if cities maintained their 1980 share of high-skilled labor.

Second, I look at a counterfactual in which congestion forces are relaxed in big frontier cities. Several recent papers have identified housing supply constraints as an important factor in driving misallocation across space and in changing patterns of migration by high- and low-skilled workers (Hsieh and Moretti, 2019; Herkenhoff et al., 2018; Ganong and Shoag, 2017). Using data from Saiz (2010), I find that big cities have more inelastic housing supplies than small cities. I examine the effect of equalizing housing supply elasticities across cities.

These counterfactuals are motivated by opposing views of these trends. The first view is that small cities are being hurt by “brain drain”, or the out-migration of their high-skilled workforce. The

second view is that big cities are too constrained in their housing supply, and therefore, no allowing enough migration of both high- and low-skilled workers to the big city. For each counterfactual, I consider the effect on the aggregate rate of technology adoption and, therefore, technical change, in addition to the usual considerations of welfare. I find that there is a tradeoff between technology adoption and welfare. While a policy of no migration is unambiguously bad for welfare, it slightly increases aggregate adoption rates and, therefore, may be good for technological progress. On the other hand, equalizing housing supply elasticities is unambiguously good for welfare, but it decreases aggregate technology adoption.

## Literature Review

There is a large recent literature devoted to documenting and explaining what [Moretti \(2012\)](#) calls the Great Divergence, or the growing economic disparities across cities in the United States since 1980. However, the divergence of the city-size wage premium by skill group has largely gone unrecognized until [Autor \(2019\)](#) brought the fact to the forefront in his 2019 Ely Lecture (the main exception being [Baum-Snow et al. \(2018\)](#)<sup>3</sup>). Furthermore, to the best of my knowledge, this paper is the first to document the changing relationship of business dynamism with city-size and that the well-documented aggregate decline in dynamism ([Decker et al., 2016](#); [Karahan et al., 2019](#); [Pugsley and Sahin, 2018](#)) was more severe in small cities, and this is the first paper to document the relationship between ICT investment and city size. Thus, an important contribution of this paper is documenting these facts and showing that they are robust to controlling for industry composition and changing demographics.

As in [Giannone \(2017\)](#), I build on the long literature on aggregate SBTC ([Autor et al., 2008](#); [Katz and Murphy, 1992](#); [Krusell et al., 2000](#)) and assess the extent to which differences in SBTC across cities can account for the Great Divergence. However, instead of calibrating the differences in SBTC to match the data, I allow the extent of SBTC to be endogenous across cities by considering the decision of heterogeneous firms to adopt a new skill-biased technology. This allows me to address the question of *why* the extent SBTC would have been different across locations.

While there is a lot of work on documenting patterns of technology adoption ([Fort, 2017](#); [Fort et al., 2018](#)) and some models of a skill biased technology choice where firms trade a high fixed cost for a low marginal cost ([Bustos, 2011](#); [Hsieh and Rossi-Hansberg, 2019](#); [Yeaple, 2005](#)), [Beaudry et al. \(2010\)](#) were the first to recognize that the incentives firms face to adopt new technologies will vary across cities and that the extent of SBTC will endogenously vary as a result. They provide reduced-form estimates of the causal impact of shocks to the relative supply of skilled labor on computer purchases. I build on this insight by embedding this firm decision into a full general equilibrium spatial model in which several characteristics of the location, in addition to labor supply, affect the firm adoption decision.

Two recent papers, [Davis et al. \(2019\)](#) and [Eeckhout et al. \(2019\)](#), have a similar goal in that

---

<sup>3</sup>In addition, closely related facts have been documented by [Diamond \(2016\)](#); [Eckert \(2019\)](#); [Ganong and Shoag \(2017\)](#); [Giannone \(2017\)](#); [Moretti \(2013\)](#).

they show an aggregate technical change shock will endogenously have different effects in different cities, but both focus on a polarization shock rather than SBTC. They show that firms' incentives to purchase capital that is substitutable with middle-skill jobs will be higher in big cities. Neither has a rich model of firm heterogeneity or considers the decision of heterogeneous firms to adopt new technologies. Further, they do not consider how the adoption decisions will affect the city level dynamism rates or that selection may amplify the differences in adoption.

Furthermore, while [Davis et al. \(2019\)](#), [Eeckhout et al. \(2019\)](#) and [Beaudry et al. \(2010\)](#) all show that spatial differences in the relative price of high- and low-skilled (and medium-skilled) labor will be important drivers of spatial differences in adoption, this paper identifies three additional city characteristics that will drive adoption differences. Besides differences in factor prices driven by differences in amenities, this paper identifies market size, selection and ex-ante differences in the productivity of using high-skilled labor as important drivers of spatial differences in technology adoption.

A large set of papers in the spatial economics literature considers a similar set of facts related to the sorting of high- and low-skilled workers across space and the growing relationship between city-size and the skill premium. Several papers ([Baum-Snow et al. 2018](#); [Rossi-Hansberg et al. 2019](#); [Michaels et al. 2013](#); [Davis and Dingel 2014, 2019](#)) argue that these facts can be explained by agglomeration economies that are biased towards high-skilled workers. [Eckert \(2019\)](#) and [Jiao and Tian \(2019\)](#) argue that decreasing communication or trade costs will cause increased geographic concentration of the high-skilled business services or high-skilled managers in large frontier cities amplifying initial productivity differences. [Diamond \(2016\)](#) shows that the effect of a SBTC shock that is asymmetric across space will be amplified if amenities respond endogenously to an increase in the supply of skilled labor. However, while these theories can explain an increase in the high-skilled city-size wage premium and an increase in sorting, they would also suggest an increase in the city-size wage premium for low-skilled workers. Instead, I propose that these trends can be explained by differences in adoption of a new skill-biased technology. While all of these mechanisms can be complementary, differences in SBTC across locations is crucial for a joint explanation of the empirical facts.

A relatively small strand of literature considers differences in business dynamism across cities. [Nocke \(2006\)](#), [Asplund and Nocke \(2006\)](#) and [Gaubert \(2014\)](#) build models in which establishment turnover will be higher in big cities. [Asplund and Nocke \(2006\)](#) confirm these patterns for Swedish hair salons and [Gaubert \(2014\)](#) in the French micro-data. I am the first to document the relationship of business dynamism to city-size in the United States, and I am the first to show that these patterns changed over time<sup>4</sup>. Despite previous literature documenting a positive correlation between dynamism and city-size in other countries, it is not a stylized fact that big cities are more dynamic; in the 1980s, big cities were *not* more dynamic than small cities.

---

<sup>4</sup>In simultaneous work, [Walsh \(2020\)](#) documents negligible differences in exit rates with city size in the pooled cohort of firms born between 1980 and 1995, consistent with my finding of no relationship or even a slightly negative relationship between the establishment exit rate and city size in 1980, he does not discuss how these patterns have changed over time.



The rest of this paper is organized as follows. In Section 2, I describe the data and document changes in economic activity across cities. In Section 3, I present a spatial equilibrium model with firm dynamics and show that it can match the relevant features of the data in 1980. Then in Section 4, I describe the SBTC shock and the calibration of the new skill-biased technology. In Section 5, I show the extent to which a SBTC shock can account for the changes in economic activity across cities since 1980. In Section 6, I discuss four channels that drive differences across cities in technology adoption in the model and validate these mechanisms using data on ICT spending. Finally in Section 7, I use the model to undertake counterfactuals before concluding in Section 8.

## 2 Descriptive Facts

In this section, I document several facts related to the changing relationship of economic activity with city-size over time. First, the correlation between average wages and city-size, known as the city-size wage premium, diverged by skill group. Second, the relationship between skill intensity and city-size has grown steeper over time. In 1980, big and small cities had similar ratios of high- to low-skilled workers, but by 2014, big cities had become much more skill intensive than small cities. Third, firms in big cities spend more on ICT per employee than firms in small cities, and a higher share of their total investment is spent on ICT. This relationship holds even when controlling for characteristics of the firms such as industry, size and age. Finally, I document the changing relationship between business dynamism and city-size. In 1980, big and small cities exhibited similar rates of business dynamics, while today, big cities are more dynamic than small cities. Relatedly, I show that the well-documented decline in the aggregate start-up rate (Decker et al., 2016; Pugsley and Sahin, 2018) was more pronounced in small cities than big cities.

### 2.1 Data description

To document facts on the changing relationship of economic activity with city-size over time, I use several micro datasets on individuals and establishments from the U.S. Census Bureau. Throughout, I use Core Based Statistical Areas (CBSAs) as the definition of cities.<sup>5</sup> As a measure of city-size, I use working age population (ages 20-64), though all results are robust to using total population. Estimates of county-level working-age population are available from the Census Bureau's Intercensal Population Estimates, which I aggregate to the CBSA level. In the main results, I limit the sample of cities to those with non-missing values of population, wages and measures of business dynamism. Measures of business dynamism are only available publicly in the Business Dynamic Statistics for Metropolitan Statistical Areas (MSAs, cities with a population of greater than 50,000). The final sample includes 366 MSAs. Where possible, I provide results using the full sample of CBSAs, which includes 942 cities.

After aggregating to the CBSA level, I further aggregate the cities into 30 bins based on 1980

---

<sup>5</sup>I use the 2009 definitions following the public use Business Dynamic Statistics.

city-size with 12 or 13 cities per bin. Binning by city-size has several advantages. First, I can calibrate and solve my model with 30 bins, which would be computationally infeasible with the full set of cities. Second, the analysis with equally sized bins gives equal weight to cities across the size distribution. Alternatively, performing the analysis at the city level would give more weight to small cities since there are more small cities than big cities, while weighting by population would give more weight to big cities since the biggest cities are home to a disproportionate share of workers<sup>6</sup>.

To document facts on the city-size wage premium and skill intensity by city-size, I use data from the 1980, 1990, and 2000 Decennial Censuses and the American Community Surveys (ACS) from 2001 to 2014<sup>7</sup> compiled by IPUMS (Ruggles et al., 2019). The finest geographic unit available in the ACS is the Public Use Micro Area (PUMA), which, unfortunately, does not map uniquely to a CBSA. To deal with cases where there is no unique mapping, I follow the method of Autor and Dorn (2013) and Baum-Snow et al. (2018): When an individual resides in a PUMA that crosses CBSA lines, I include the individual in both CBSAs and adjust their weights for the probability that they reside in each using the Geographic Equivalency Files from the Census Bureau.

I limit the sample to civilian workers aged 16 to 64 who work at least 30 hours a week. I use the occupation codes from Autor and Dorn (2013) and the 1990 industry codes provided by IPUMS. I group the detailed race categories into four groups that are consistent over time: white non-hispanic, hispanic, black and asian. Wages include all wage and salary income including cash tips, commissions and bonuses collected in the previous calendar year. High-skilled workers are those with at least four years of college, while low-skilled workers are those with less than four years of college. In Appendix A, I show the main facts are robust to alternative definitions of skill. For data on rents and housing prices, I follow Ganong and Shoag (2017) and calculate a housing composite measure, which is 12 times the monthly rent for renters and 5 percent of the value of the house for homeowners. I use measures of housing supply elasticity from Saiz (2010).

To document facts on business dynamics and technology adoption I use several different sources of firm and establishment micro data from the U.S. Census. The primary dataset is the Longitudinal Business Database (LBD), which is a panel dataset of all U.S. establishments with at least one employee. The LBD includes information on establishment county, employment, industry, and payroll from 1976 to 2014. Establishments are linked to their parent firm using a firm ID. Firm age is imputed from the first time its oldest establishment is observed in the LBD. Because the LBD starts in 1976, I only have meaningful age information starting in 1987, at which point I know if a firm is age 0 to 10 or 11+. For firms with establishments across multiple sectors, I assign a firm industry from the establishment-level industry codes using the industry with the firm's highest share of employment. I assign the two-digit NAICS code first and then the three digit code with the most employment that is consistent with the two digit codes, and so on, up to six digits. For a consistent measure of industry codes going back to 1976, I use the NAICS codes available for the LBD from Fort and Klimek (2018). I also use the public Business Dynamic Statistics (BDS), which

---

<sup>6</sup>The top 10 cities contain 29.6 of the population in 1980.

<sup>7</sup>I stop my analysis in 2014 because that is the last year the BDS is available.

is available at the MSA level and provides information on the number of firms and establishments, as well as job creation and destruction by firm and establishment age.

I supplement the LBD with establishment-level sales data from the Economic Censuses. There are 9 Economic Censuses<sup>8</sup> available in years ending in two and seven (1977, 1982, etc.). However, the coverage varies by sector. As such, when using data on sales, I use data only from sectors that were available in 1977<sup>9</sup>. This ensures the results are not driven by changes in sample selection.

For data on ICT expenditures, I use data from the Annual Capital Expenditure Survey (ACES), which has an annual ICT supplement starting in 2003. The survey is collected at the firm level and, when a firm is sampled into the ICT supplement, contains annual ICT expenditures. A new sample of firms is chosen each year giving a series of repeated cross-sections. The data include capitalized and non-capitalized expenditures, as well as rental, lease and maintenance payments for software, computer equipment, and other peripheral ICT equipment. Software expenses include prepackaged, customized or in-house built software, including payroll related to development. Where possible, I supplement the ACES data with information on employment, payroll, sales and industry from the LBD and the Economic Censuses. Unfortunately, only 95.5% of records contain the firm identifiers necessary to match to the LBD. Appendix A.3 provides information on the characteristics of matchers versus non-matchers. Because the ICT supplement is only available in later years, I interpret the data as being informative about the period in my model in which the new skill-biased technology has already become available.

For the ACES data, which are at the firm level, I assign a firm location first using the location of the establishment with a NAICS code of 55 (being devoted to the management of companies). If a firm has more than one establishment with NAICS code 55 or if it has no establishment with NAICS 55, I assign the location using the location of the firm's establishment with the highest payroll per employee. I use payroll per employee instead of the location with the highest share of employees because the number of employees is likely to be spuriously correlated with the size of the city. For example, I do not want to assign the location of Starbucks as New York City just because it has a location on every block. Further, an establishment with higher-than-average pay per employee is likely to be associated with high-paying management activities.

## 2.2 Demographic facts

I document two sets of facts about changing outcomes for skilled and unskilled workers by city-size. First, I show that the city-size wage premium has diverged by skill group. Second, I show that the relationship between skill intensity and city-size has grown steeper over time. Closely related facts have been previously documented by researchers (including Autor (2019); Baum-Snow et al. (2018); Diamond (2016); Giannone (2017)), and I reproduce them in my sample as they will be an

---

<sup>8</sup>Census of Construction Industries, Census of Services, Census of Retail Trade, Census of Manufacturers, Census of Wholesale Trade, Census of Auxiliary Establishments, Census of Finance, Census of Mining, and the Census of Utilities.

<sup>9</sup>All of the Economic Census were available in 1977 except the Census of Finance, Census of Mining and the Census of Utilities.

important input in my quantitative exercise.

*Fact 1: The city-size wage premium diverged by skill group since 1980.*

Figure I shows the relationship between wages and city-size by skill group. In 1980, the elasticity of wages with respect to city-size was about .056 and .047 for high- and low-skilled workers, respectively, meaning that high-skilled workers earned about 4.0 percent more for working in cities twice as large and low-skilled workers earned about 3.3 percent more<sup>10</sup>. By 2014, the wage premium had substantially diverged. The city-size wage premium increased to .072 percent for high-skilled workers, while the city-size wage premium fell to just .035 percent for low-skilled workers.

The divergence of the city-size wage premium by skill group is robust to controlling for demographics, occupation, and industry of the workers. To control for demographics, I estimate a Mincer regression in each year and for each skill group separately,

$$\log(w_{it}) = \alpha_t + \beta_t X_{it} + \epsilon_{it}.$$

In the first set of regressions I include standard demographic controls typical in a Mincer regression: age, age squared, a gender dummy and full set of race fixed effects. In the second set of regressions, I add a full set of industry and occupation fixed effects. I then use the residuals to compute the average adjusted wages in a city,  $\bar{w}_{c,t} = \sum_{i \in c} \exp(\epsilon_{it})$ <sup>11</sup>. I fit a separate Mincer regression for each skill group and in each year to allow the coefficients to change over time and to differ by skill. This ensures that the divergence in the city-size wage premium is not driven by differences in demographic composition between big and small cities. In Appendix A, Tables A.I reproduces these figures using more education groups, and Figure A.I reproduces them using the full set of cities instead of city bins.

In panels (a) and (b) of Figure II, I show the city-size wage premium adjusted for standard demographic controls. The divergence of the city-size wage premium by skill group is unchanged. For both groups the city-size wage premium is about .045 percent in 1980. By 2014, high-skilled workers are earning a city-size wage premium of .074 percent while for low-skilled workers the city-size wage premium fell to .038 percent.

Finally, in panels (c) and (d) of Figure II, I show the city-size wage premium controlling for occupation and industry fixed effects in addition to the standard demographic controls. Including industry and occupation fixed effects decreases the level of the city-size wage premium substantially, suggesting that workers were already engaged in different activities in big and small cities in 1980.

<sup>10</sup>The increase in earnings will be equal to  $y'/y = (\exp(\beta * \ln(x'/x)) - 1)$ . If  $x'$  is twice the size of city  $x$ , then  $x'/x$  will be 2.

<sup>11</sup>I take the exponential of the residual because the Mincer regression uses log wages and the residual should therefore be interpreted as the adjusted log wage. Instead one could compute the average of the residuals directly, but this would be interpreted as taking the average of the  $\log(w_{it})$  in a city. In practice both methods produce similar results. However, in theory, the difference between the mean of the log and the log of the mean will depend on variance of the wage distribution. Baum-Snow and Pavan (2013) and Santamaria (2018) show that inequality is systematically higher in big cities suggesting that using the mean of the log could bias the estimate of the city-size wage premium.

The occupation and industry controls also explain some, but not all, of the divergence. In 1980, both groups earn a wage premium of approximately .025%. By 2014, it is .052% for high-skilled workers and .030% for low-skilled workers. The city-size wage premium for low-skilled workers is no longer lower in 2014 than in 1980, but the divergence between the high- and low-skilled city-size wage premium still holds. The fact that the occupation and industry controls can explain some of the divergence means that workers are shifting into higher-paid occupations and industries more in big cities than in small cities. This is consistent with firms in big cities increasing their demand for high-skilled occupations more than firms in small cities.

*Fact 2: The relationship between skill intensity and city-size has grown steeper over time.*

Figure III shows the relationship between skill intensity and city-size. Skill intensity is the ratio of high-skilled workers (those with 4 or more years of college) to low-skilled workers (those with less than 4 years of college). In 1980, the semi-elasticity of skill intensity with respect to city-size was .026 percentage points. By 2014, the semi-elasticity increased to .074. Cities that were twice as large in 1980 were, on average, 1.80 percentage points more skill intensive. By 2014, the relationship between skill intensity and city-size had steepened; cities twice as large were 5.13 percentage points more skill intensive. In Figure A.II of Appendix A.1, I show the results are robust to using alternative definitions of high- and low-skilled workers and to using the full set of cities.

### 2.3 Firm facts

*Fact 3: Today, firms in big cities invest more in ICT.*

Table I shows the results from a regression of log of ICT expenses per employee and ICT share of investment on city-size and industry fixed effects. The elasticity of ICT spending per employee with respect to city-size is .054 percent, meaning that firms in cities that are twice as large spend 3.8 percent more per employee on ICT investment than firms in small cities. In column 2, I add controls for the age and size of the firm with little effect on the correlation between city-size and ICT spending per employee. Similarly, columns 3 and 4 present the same regressions using the ICT share of investment as the dependent variable. Firms in cities twice as large spend, on average, 1 percentage point more of their total investment budget on ICT-related expenses such as computer equipment and software development. Controlling for size and age has little effect on the coefficient. I run these regressions on pooled data across the years for which the ICT data is available (2003-2013) and include yearly fixed effects in the regression. Because the data are only available for recent years, I interpret these results as being informative about the second steady state in my model in which the new technology is available.

*Fact 4: Big cities today are more dynamic than small cities. This was not true in 1980.*

Throughout the analysis, I use several different measures of dynamism including the firm and establishment start-up and exit rate and the job creation and destruction rates. Following Davis

et al. (1996), the start-up rate in time  $t$  is defined as  $sr_t = \frac{\text{age 0 firms}_t}{(\text{firms}_t + \text{firms}_{t-1})/2}$ ; the exit rate is defined as  $er_t = \frac{\text{firm deaths}_t}{(\text{firms}_t + \text{firms}_{t-1})/2}$ ; the establishment start-up rate and exit rates at time  $t$  are defined analogously as  $esr_t = \frac{\text{age 0 establishments}_t}{(\text{establishments}_t + \text{establishments}_{t-1})/2}$  and  $eer_t = \frac{\text{establishments exits}_t}{(\text{establishments}_t + \text{establishments}_{t-1})/2}$ ; and job creation and destruction are employment increases and decreases at expanding or contracting firms, respectively. Specifically, job creation is

$$jcr_t = \frac{\sum_i (emp_{ti} - emp_{t-1,i}) \mathbb{I}[emp_{ti} > emp_{t-1,i}]}{(emp_t + emp_{t-1})/2}$$

and

$$jdr_t = \frac{\sum_i (emp_{t-1,i} - emp_{ti}) \mathbb{I}[emp_{ti} < emp_{t-1,i}]}{(emp_t + emp_{t-1})/2}.$$

I take three-year moving averages of all dynamism rates to smooth out cyclical variation.

Figure IV shows the relationship between dynamism and city-size for 1980 and 2014, using data from the public Business Dynamic Statistics. In 1980, there is no relationship between measures of dynamism and city-size, while by 2014, big cities are more dynamic than small cities. In 2014, the semi-elasticity of the establishment start-up rate with respect to city-size is .45 percentage points, meaning that the establishment start-up rate is .31 percentage points higher in cities twice as large.

In Appendix A.2, I show that these results are robust to controlling for city-level industry composition. I present the correlation between dynamism and city-size over time, controlling for industry composition at the 3-digit NAICS level. To do this, I compute dynamism measures within a city-industry cell and then estimate the following specification:

$$D_{ict} = \alpha_{it} + \lambda_t \times \sum_t \beta_t \log(pop_{ct}) + \epsilon_{ict}$$

where  $D_{ict}$  is the measure of dynamism and  $\alpha_{it}$  is a full set of industry-year fixed effects. The population variable is interacted with a full set of year dummies,  $\lambda_t$ , giving an estimate of the cross-sectional relationship between dynamism and city-size,  $\beta_t$ , for each year. Because city-industry bins become very small for detailed industry categories, I also perform an alternative analysis by grouping cities into 7 size categories<sup>12</sup>. I then calculate the dynamism measures within a city-size category-industry bin, using 4-digit NAICS codes.

Next, I show that cities that were small in 1980 exhibited larger declines in dynamism than did big cities. This ensures that the changing relationship of dynamism to city-size is not just driven by dynamic cities moving up the city-size distribution. Figure V shows the relationship between the decline in dynamism versus 1980 city-size in the public BDS data. To control for industry

<sup>12</sup>Specifically, city-size categories are based on population percentiles. City-size category 1 is cities below the 50th percentile, category 2 is the 50th to 75th percentile, 3 is the 75th to 90th percentile, 4 is the 90th to 95th percentile, 5 is 95th percentile to 98th percentile, 6 is the 98th to 99th percentile and finally category 7 is cities above the 99th percentile.

composition, I estimate the following specification:

$$\Delta_{1980,2014}D_{ic} = \alpha_i + \beta \log(\text{pop}_{c1980}) + \epsilon_{ic}$$

where  $\Delta_{2014,1980}D_{ic}$  is the change in dynamism between 1980 and 2014 in city  $c$  and industry  $i$ ,  $\alpha_i$  is a full set of industry fixed effects and  $\text{pop}_{c1980}$  is a city’s 1980 working age population. I estimate this at the city-NAICS 3 level. The results are presented in Table II. In Appendix A.2, I group cities into size bins and control for more detailed industry composition.

In cities twice as large, the establishment start-up rate fell .5 percentage points less than in cities half the size. In practice, this amounts to a sizable difference with the smallest cities experiencing a decline in dynamism twice as severe as the largest cities.

### 3 A Spatial Equilibrium Model with Firm Dynamics

A key objective of the model is to explain the 1980 relationships of wages, skill intensity and dynamism with city-size, as documented in the previous section. As such, the model embeds a standard model of firm dynamics (Hopenhayn, 1992; Luttmer, 2007) into a benchmark spatial equilibrium framework (Roback, 1982), allowing a joint consideration of firm dynamics and relative wage inequality.

There are two types of workers, high- and low-skilled, who are freely mobile between  $J$  cities. Workers have idiosyncratic preferences for each city, and they consume a freely traded final good and housing. In each city there is a final goods producer that aggregates a continuum of local intermediate varieties. The perfectly competitive final good is freely traded on the national market, and the price of the final good is normalized to one.

In each city there is a continuum of monopolistically competitive firms that produce using a CES bundle of high- and low-skilled labor. They sell their output to a local final goods producer, who produces a non-differentiated perfectly tradable final good. Intermediate firms pay a fixed cost of production in units of high- and low-skilled labor weighted according to their CES cost shares. Firms must also purchase a fixed unit of floor space in their city. Time is continuous and firms are subject to a stochastically varying productivity. Firms exit when their productivity falls below a threshold, which varies endogenously across cities. Firms choose a city in which to enter and pay an entry cost denominated in units of labor. Firms cannot move once they enter. In the 1980 steady state, all firms within a city use the same technology to produce. In Section 3.7, I discuss the calibration of the initial steady state.

#### 3.1 Final good producer

In each city,  $j$ , a final goods producer uses a continuum of local intermediate varieties, indexed by  $\omega$ . The final goods producer aggregates the intermediates according to a standard CES demand

function

$$Y_j = \left( \int_{\Omega_j} q_j(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}.$$

Solving the profit maximization problem of the final goods producer, an intermediate firm faces the demand function

$$q_j(\omega) = Y_j P_j^\sigma p(\omega)^{-\sigma} = R_j P_j^{\sigma-1} p(\omega)^{-\sigma},$$

where  $P_j = \left( \int_{\Omega_j} p_j(\omega)^{1-\sigma} d\omega \right)^{\frac{1}{1-\sigma}}$  is the price of the final good. The final good is freely traded in a national market and the price is normalized to 1.

### 3.2 Intermediate producer

The intermediate firm's problem can be separated into a static component and a dynamic component. In the static profit maximization problem, firms choose their optimal bundle of high- and low-skilled labor to minimize their unit cost function and then choose price and output to maximize profits given demand from the local final goods producer. In the dynamic problem, incumbent firms choose the optimal exit threshold given their stochastically varying productivity and a continuum of potential entrants choose whether or not to enter.

#### Static problem

All intermediate firms in a city  $j$  produce using the same CES bundle of high- and low-skilled labor

$$q_{jt} = \psi_j z_t(\omega) \left( (1 - \gamma_j) l_{jt}^{\frac{\epsilon-1}{\epsilon}} + \gamma_j h_{jt}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}},$$

where  $z_t(\omega)$  is a firm-specific stochastically varying productivity term and  $\psi_j$  is a city-specific productivity term. The weight placed on high-skilled labor  $\gamma_j$  is city specific.

First, firms choose the bundle of high- and low-skilled labor that minimizes their unit cost function,

$$c_j(z) = \min_{l_{jt}, h_{jt}} w_{l_{jt}} l_{jt} + w_{h_{jt}} h_{jt}$$

such that

$$\psi_j z_t(\omega) \left( (1 - \gamma_j) l_{jt}^{\frac{\epsilon-1}{\epsilon}} + \gamma_j h_{jt}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \geq 1.$$

Solving the minimization problem gives the Hicksian demand function for labor (dropping time subscripts for ease of notation)

$$l_j^d(z) = \frac{w_{l_j}^{-\epsilon} (1 - \gamma_j)^\epsilon}{\psi_j z (\gamma_j^\epsilon w_{h_j}^{1-\epsilon} + (1 - \gamma_j)^\epsilon w_{l_j}^{1-\epsilon})^{\frac{\epsilon}{\epsilon-1}}},$$

$$h_j^d(z) = \frac{w_{h_j}^{-\epsilon} \gamma_j^\epsilon}{\psi_j z (\gamma_j^\epsilon w_{h_j}^{1-\epsilon} + (1 - \gamma_j)^\epsilon w_{l_j}^{1-\epsilon})^{\frac{\epsilon}{\epsilon-1}}},$$



and the unit cost function

$$c_j(z) = (\psi_j z)^{-1} (\gamma_j^\epsilon w_{hj}^{1-\epsilon} + (1 - \gamma_j)^\epsilon w_{lj}^{1-\epsilon})^{\frac{1}{1-\epsilon}}$$

for a firm in city  $j$  with productivity  $z$ . The share of high-skilled and low-skilled labor that a firm uses to produce one unit of output is given by  $\theta_j^h = \frac{h_j^d}{l_j^d + h_j^d}$  and  $\theta_j^l = \frac{l_j^d}{l_j^d + h_j^d}$ , respectively.

Next, the intermediate firm chooses output and price to maximize profits subject to their downward sloping demand curve from the final goods producer:

$$\pi_j(z) = \max_{p_j(\omega), q_j(\omega)} p_j(\omega) q_j(\omega) - q_j(\omega) c_j(z) - (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c - r_j f_j^b$$

such that

$$q_j(\omega) = Y_j P_j^\sigma p_j(\omega)^{-\sigma},$$

where  $f_j^c$  is the fixed cost paid by firms in units of high- and low-skilled labor and  $f_j^b$  is the fixed units of building space firms need to produce in city  $j$ . Firms choose their price  $p_j(z)$  as a mark-up over their marginal cost (which, since the production function exhibits constant returns to scale, is the same as the unit cost function  $c_j(z)$  solved for in the previous step)

$$p_j(z) = c_j(z) \frac{\sigma}{\sigma - 1}.$$

Variable profits are thus

$$\pi_j^v(z) = \left( c_j(z) \frac{\sigma}{\sigma - 1} \right)^{1-\sigma} Y_j P_j^\sigma \frac{1}{\sigma}.$$

### Dynamic problem: exit

The productivity of intermediate firms varies according to a geometric Brownian motion, and firms discount future profits at an interest rate  $\rho$ .<sup>13</sup> The dynamic problem of the firm is as follows: Firms choose the optimal exit threshold  $z_x$ , below which they will exit the market

$$V(z) = \max_{z_{xj} \leq z} \mathbb{E}_z \int_0^{T(z_{xj})} e^{-\rho t} \left( \pi_j(z(t)) - (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c - r_j f_j^b \right) dt$$

$$d \log(z_t) = \mu dt + \Psi dW(t),$$

where  $T(z_{xj})$  is the time at exit. Defining

$$s = \ln z^{\sigma-1}$$

---

<sup>13</sup>Household consumption is linear in wages, so the interest rate will be pinned down by the household's discount rate.

and using a change of variable, the value function can be rewritten as

$$V(s) = \max_{s_{xj} \leq s} \mathbb{E}_s \int_0^{T(s_{xj})} e^{-\rho t} \left[ Z_j e^s - (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c - r_j f_j^b \right] dt$$

$$ds_t = \tilde{\mu} dt + \tilde{\Psi} dW(t),$$

where  $Z_j = \left( (\psi_j)^{-1} (\gamma_{jh}^\epsilon w_{jh}^{1-\epsilon} + \gamma_{jl}^\epsilon w_{jl}^{1-\epsilon})^{\frac{1-\sigma}{\sigma-1}} Y_j P_j^\sigma \frac{1}{\sigma} \right)^{1-\sigma}$  and  $\tilde{\mu} = (\sigma - 1) \left( \mu - \frac{1}{2} \Psi^2 \right)$  and  $\tilde{\Psi} = \Psi(\sigma - 1)$ . The stochastic process for  $s$  is derived using Ito's Lemma<sup>14</sup> and follows a standard Brownian motion.

The solution to the firm's optimal stopping time problem solves the HJB equation

$$\rho V(s) = [Z_j e^s - (w_{hj} + w_{lj}) f_j^c - r f_j^b] + \tilde{\mu} V'(s) + \frac{1}{2} \tilde{\Psi}^2 V''(s)$$

and the following boundary conditions

$$\text{no bubble: } \lim_{s \rightarrow \infty} V(s) - V_p(s) = 0$$

$$\text{value matching: } V(s_{xj}) = 0$$

$$\text{smooth pasting: } V'(s_{xj}) = 0.$$

An advantage of using continuous time is that it allows a closed form solution for the value function.

I derive the solution to the value function in Appendix B.1:

$$V(s) = \frac{Z_j}{\rho - \tilde{\mu} - \frac{1}{2} \tilde{\Psi}^2} e^s + \frac{1}{\rho} [-(w_{hj} + w_{lj}) f_j^c - r_j f_j^b] \dots \\ + \left[ \frac{-Z_j}{\rho - \tilde{\mu} - \frac{1}{2} \tilde{\Psi}^2} e^{s_x - \xi^- s_x} - \frac{1}{\rho} [-(\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c - r_j f_j^b] e^{-\xi^- s_x} \right] e^{\xi^- s},$$

and the exit threshold

$$s_{xj} = \ln \left[ \frac{1}{Z_j} \left( \frac{-\xi^-}{1 - \xi^-} \right) \left( \frac{\rho - \tilde{\mu} - \frac{1}{2} \tilde{\Psi}^2}{\rho} \right) \left( (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c + r_j f_j^b \right) \right]$$

where  $\xi^- = -\tilde{\mu} - \frac{\sqrt{\tilde{\mu}^2 + 2\tilde{\Psi}^2 \rho}}{\tilde{\Psi}^2}$ .

<sup>14</sup>Ito's lemma says that if  $x$  is a Brownian motion  $dx = \mu(x)dt + \sigma(x)dW$  then  $y = f(x)$  is also a Brownian motion of the form  $dy = df(x) = (\mu(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x)) dt + \sigma(x)f'(x)dW$ . Thus,

$$ds = df(z) = \frac{1}{a}(\sigma - 1) \left( \mu z \frac{1}{z} - \frac{1}{2} \psi^2 z^2 \frac{1}{z^2} \right) dt + \frac{1}{a}(\sigma - 1) \psi dW(t) \\ = (\sigma - 1) \left( \mu - \frac{1}{2} \psi^2 \right) dt + (\sigma - 1) \psi dW(t) = \tilde{\mu} dt + \tilde{\psi} dW(t)$$

### Dynamic problem: entry

There is an unbounded mass of potential entrants. Entrants pay a fixed entry cost in units of high- and low-skilled labor, weighted by the CES shares of high- and low-skilled labor, and start with an initial productivity of  $z_e$ . Firms choose the city in which to enter that will maximize their expected value. Once they start in a city, they cannot move. As a result, a free entry condition holds in each market

$$V_j(z_e) = (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^e,$$

such that entrants are indifferent between potential entry locations.

### 3.3 Firm size distribution

The distribution of firm size evolves according to the Kolmogorov Forward Equation

$$\frac{\partial g_j(s)}{\partial t} = -\tilde{\mu} \frac{dg_j(s)}{ds} + \frac{1}{2} \tilde{\psi}^2 \frac{d^2 g_j(s)}{ds^2}.$$

In steady state, the distribution must be unchanging,  $\frac{\partial g_j(s)}{\partial t} = 0$ , which implies the stationary distribution must satisfy  $-\delta \frac{dg_j(s)}{ds} = \frac{d^2 g_j(s)}{ds^2}$  where  $\delta = \frac{-\tilde{\mu}}{\tilde{\psi}^2/2}$ . In Appendix B.2 I show that the distribution can be solved for in closed form and is described by the following equation:

$$g_j(s|s_e) = \begin{cases} \frac{E_j}{\tilde{\mu}} (e^{-\delta(s-s_{xj})} - 1) & s \in (s_{xj}, s_e) \\ \frac{E_j}{\tilde{\mu}} (e^{\delta s_{xj}} - e^{\delta s_e}) e^{-\delta s} & s > s_e. \end{cases} \quad (1)$$

Integrating over the distribution of firms in city  $j$  gives the mass of firms  $M_j$ :

$$M_j = \int_{s_{xj}}^{\infty} g_j(s) = \frac{E_j}{\tilde{\mu}} (s_{xj} - s_e).$$

Rearranging the expression for the mass of firms and noting that in steady state, the mass of exitors,  $E_j$ , must be equal to the mass of entrants,  $N_j$ , the start-up rate in city  $j$  is given by

$$sr_j = \frac{N_j}{M_j} = \frac{-\tilde{\mu}}{s_{ej} - s_{xj}}.$$

The start-up rate in a city is a function of the drift in firm size,  $\tilde{\mu}$ , (note that  $\tilde{\mu} < 0$ ) and the difference between the entry and exit thresholds. As the exit threshold increases, the start-up rate rises. This is because as the exit threshold increases, the probability that a firm exits increases. Since the entry rate and exit rate must be equal in the steady state, an increase in the probability of exit implies an increase in the start-up rate.

### 3.4 Households

Labor is perfectly mobile across cities. High- and low-skilled workers provide one unit of labor and consume the final good and residential land. The utility of worker  $i$  of type  $\tau$  in city  $j$  is given by

$$V_{i\tau j}(w_{\tau j}, q_j) = \max_{c,b} \log(c_{\tau j}^{\beta} d_{\tau j}^{1-\beta}) + \log(A_{\tau j}) + \zeta_{ij}$$

subject to their budget constraint

$$w_{\tau j} = c_{\tau j} + q_j d_{\tau j},$$

where  $q_j$  is the price of residential floor space and  $\zeta_{ij}$  is an idiosyncratic location preference drawn from a type-I extreme value distribution with shape parameter  $\nu$ , which determines the elasticity of labor supply with respect to the real wage.  $A_{\tau j}$  is a type specific amenity received for living in city  $j$ . The indirect utility function is given by

$$\bar{V}_{i\tau j}(w_{\tau j}, q_j) = \log\left(\beta^{\beta}(1-\beta)^{(1-\beta)}w_{\tau j}q_j^{\beta-1}\right) + \log(A_{\tau j}) + \zeta_{ij}.$$

Define  $\bar{V}_{\tau j} \equiv \log\left(\beta^{\beta}(1-\beta)^{(1-\beta)}w_{\tau j}q_j^{\beta-1}\right) + \log(A_{\tau j})$ . Then the probability that city  $j$  provides the highest utility to worker  $i$  will follow a multinomial logit

$$\pi_{\tau j} = P(V_{i\tau j} > V_{i\tau k}, \forall k \neq j) = \frac{\exp(\nu \bar{V}_{\tau j})}{\sum_j \exp(\nu \bar{V}_{\tau j})}.$$

### 3.5 Land

A landlord builds floorspace and rents commercial space to firms and residential housing to workers. He chooses the amount of floor space,  $B_j$ , to build and the share of it that is rented to firms,  $\Theta_j$ , to solve

$$\max_{B_j, \theta_j} r_j \Theta_j B_j + q_j (1 - \Theta_j) B_j - \frac{B_j^{\eta_j}}{b_j^{\eta_j-1} \eta_j},$$

where  $r_j$  is the rental price of a commercial floor space,  $q_j$  is the rental price of residential housing, and  $\frac{B_j^{\eta_j}}{b_j^{\eta_j-1} \eta_j}$  is the cost of producing  $B_j$  units of floor space. In steady state, a city cannot exist without strictly positive amounts of commercial and residential space. A no arbitrage condition requires that rents for both uses of floor space are equal,  $r_j = q_j$ , otherwise landlords would only rent to the users that pay the highest rent.

This gives an equation for the supply of buildings

$$B_j = b_j r_j^{\frac{1}{\eta_j-1}},$$

where  $\frac{1}{\eta_j-1}$  is the elasticity of building supply and  $b_j$  is the productivity of the building sector. Profits of the landlord are  $\Pi_j^B = (b_j r_j)^{\frac{\eta_j}{\eta_j-1}} \frac{1}{b_j} \left(\frac{\eta_j-1}{\eta_j}\right)$  and their costs, which are paid in units of the

final good, are  $c_j^B = \frac{1}{b_j \eta_j} (b_j r_j)^{\frac{\eta_j}{\eta_j - 1}}$ . The market clearing condition for the building market is

$$M_j f^b + (1 - \beta) H_j \frac{w_{hj}}{r_j} + (1 - \beta) L_j \frac{w_{lj}}{r_j} = B_j^s$$

where the left-hand side gives demand for buildings from firms and workers, respectively.

### 3.6 Equilibrium

Profits are made by the landlords and by the owners of the firms. I assume that profiteers consume units of the final good in their city, but do not consume housing. Thus, even though the final good is freely traded, the final goods market will clear in each city since the final good is not differentiated and each city makes only enough to satisfy local demand. The final goods market clearing condition in each city is given by

$$\begin{aligned} \beta w_{Hj} H_j + \beta w_{Lj} L_j + \left( \Pi_j^v - (\theta_j^h w_{hj} + \theta_j^l w_{lj}) M_j f^c - M_j r_j f^b - (\theta_j^h w_{hj} + \theta_j^l w_{lj}) N_j f_j^e \right) \\ \dots + \Pi_j^B + c_j^B = Y_j. \end{aligned}$$

A labor market clearing condition holds for each type of labor

$$\begin{aligned} \pi_{lj} L &= M_j \int_z l_j^d(z) q(z) g(z) dz + \theta_j^h M_j f_j^c + \theta_j^l N_j f_j^e \\ \pi_{Hj} H &= M_j \int_z h_j^d(z) q(z) g(z) dz + \theta_j^h M_j f_j^c + \theta_j^l N_j f_j^e, \end{aligned}$$

where  $\pi_{\tau j}$  is the share of labor type  $\tau$  that chooses to go to city  $j$  and the right hand side aggregates over the distribution of firms to get city level labor demand.

A steady-state equilibrium is a set of prices, labor allocations, output, the mass of firms and entrants, and exit thresholds,  $\{w_{lj}, w_{hj}, r_j, L_j, H_j, Y_j, M_j, N_j, z_{xj}\}$ , for each city, such that:

1. there is an invariant distribution of firms,  $g_j(s)$ , that satisfies the Kolmogorov Forward Equation,
2. there is a constant mass of firms and entrants,  $M_{jt}$  and  $N_{jt}$ ,
3. exit policies,  $z_{xj}$ , satisfy the firms HJB equation and the free entry condition holds in each city,
4. firms and landlords maximize profits,
5. workers choose the city that maximizes utility and the share of their income to spend on consumption and housing,
6. the final goods market, the land market, and labor markets clear in each city.

### 3.7 Calibration

To calibrate the model, I first calibrate a set of parameters that are constant across cities. Some of these parameters are standard values that I take from the literature, and others are calibrated to match moments in the data. Next, I recover the city-level fundamentals that can rationalize the cross-sectional patterns by city-size in 1980. I show that conditional on the aggregate parameters, there is a unique set of city fundamentals that are consistent with the data in 1980 being a steady-state equilibrium of the model.

Table III gives the aggregate parameters and the corresponding moments or sources that I use to calibrate them. The mass of low-skilled workers is normalized to 1, and the mass of high-skilled workers is calculated using the aggregate share of workers with 4 or more years of college in 1980.  $\beta$  gives the share of income spent on rent, calibrated to be .4 following Monte et al. (2018). The elasticity of substitution between varieties is also a standard parameter taken from the literature. I set it to 6.8, in line with the estimates from Broda and Weinstein (2006). Finally, I take the scale parameter of the Gumbel distribution, which governs the migration elasticity, from Allen et al. (2018) and set it to 3.

In Table IV, I show the calibration and model fit of the firm productivity parameters and the firm production function. These four parameters are jointly calibrated using a method-of-moments routine. A standard value for the aggregate elasticity of substitution between high- and low-skilled labor is 1.62, as estimated by Autor et al. (2008). However,  $\epsilon$  in my model is a firm level elasticity of substitution for variable factors of production (excluding the fixed costs). I calibrate the firm level  $\epsilon$  to match an aggregate elasticity of substitution of 1.59. I calibrate the parameters of the firm’s stochastic process for productivity, the drift  $\mu$  and the variance  $\Psi$ , using moments calculated in the establishment micro-data. The establishment size distribution will have a Pareto tail with tail parameter  $\delta$ . In the LBD, I calculate that  $\delta$  is equal to 1.26 and the standard deviation of employment growth is 42%<sup>15</sup>. I calibrate the fixed cost that firms pay in units of building space,  $f_b$ , so that the share of the firm’s costs spent on the fixed unit of building space is 3.5%. The fixed building space should be interpreted as the minimum amount of real estate a firm needs to produce in the city.<sup>16</sup>

Next, I calibrate the parameters that vary across cities. Let  $\mathbb{P}^c = \{\gamma, \psi, A_h, A_l, f^e, f^c, b, \eta\}$  be the vector of city level fundamentals, that is the weight on high-skilled labor,  $\gamma$ , the city specific productivity term,  $\psi$ , high- and low-skilled amenities,  $A_h$  and  $A_l$ , entry costs,  $f_e$ , fixed costs,  $f_c$ , the productivity of the building sector  $b$ , and the elasticity of building supply,  $\eta$ . Table V gives the moments from the data used to identify each of the fundamentals. In

<sup>15</sup>Specifically, this is the standard deviation of the symmetric growth rate for all establishments weighted by employment. This is larger than the typically used 25% number used by ? who calculate the standard deviation for large firms only.

<sup>16</sup>I could not find estimates from the literature to pin down this parameter. The share of costs paid in rent for the minimum building space cannot be much higher than 4% without the value of entering falling below the cost of entering in which case there is no entry in equilibrium. Thus, within the feasible range of this target (between 1 and 4 percent of costs), it makes little difference to the results.

Appendix C.1, I show that there are unique values of the fundamentals that rationalize the data in 1980 as being an equilibrium of the model, and I describe the steps to back out the fundamentals from the data.

## 4 The Arrival of a New Technology

In this section, I consider what happens when a new technology is introduced into the economy in Section 3. The new technology has an absolute productivity advantage but uses high-skilled labor more intensely, and so it represents SBTC. Firms can choose to adopt the new technology but will pay an additional fixed cost in units of high- and low-skilled labor. The adoption choice is irreversible. Even though the new technology is available across all cities, firms that are otherwise the same will potentially make different adoption choices based on the characteristics of the city in which they produce. I first describe the new technology and the technology adoption choice. Then I define the equilibrium in a model with endogenous technology adoption and describe the calibration of the new technology. Finally, I describe the mechanisms that drive differences in the technology adoption decisions across cities.

### 4.1 The new technology

The new technology uses a CES bundle of high- and low-skilled labor

$$q_{jt} = \Gamma_\psi \psi_j z_t(\omega) \left( (1 - \Gamma_\gamma \gamma_j) l_{jt}^{\frac{\epsilon-1}{\epsilon}} + \Gamma_\gamma \gamma_j h_{jt}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}},$$

which has an absolute productivity advantage,  $\Gamma_\psi > 1$ , but places a higher weight on the use of high-skilled labor than the old technology,  $\Gamma_\gamma > 1$ . As in Section 3.2, firms choose the optimal bundle of high-skilled and low-skilled labor to minimize their unit cost function. The unit cost function for using the new technology is given by

$$c_j^a(z) = (\Gamma_\psi \psi_j z)^{-1} \left( (\Gamma_\gamma \gamma_j)^\epsilon w_{hj}^{1-\epsilon} + (1 - \Gamma_\gamma \gamma_j)^\epsilon w_{lj}^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}}.$$

Depending on the wages in a city, adopting the new technology may deliver marginal cost savings to the firm. Conditional on adopting, the firm's profit maximization problem is the same as the static profit maximization problem under the old technology described in Section 3.2, except that now it has a lower marginal cost.

In order to use the new technology, firms pay an additional fixed cost in units of high-skilled labor denoted as  $\Gamma_{fc}$ , which is constant across cities and paid according to the cost shares of the new CES technology.

### 4.2 The firm's adoption decision

As long as adopting lowers the marginal cost,  $c_j^a(z) < c_j(z)$ , there will always be a productivity threshold,  $z_a$ , above which it is optimal to adopt the new technology. If wages are such that the

new technology does not lower the firm's marginal cost,  $c_j^a(z) > c_j(z)$ , no firm will adopt the new technology.

The dynamic optimization problem of the firm involves solving for three thresholds. First, firms that have not yet adopted choose an exit threshold below which they will exit the market,  $z_x^a$ . Second, conditional on the value of adopting the new technology, firms that have not yet adopted choose the technology adoption threshold,  $z_a$ , above which firms will adopt the new technology and third, they choose an exit threshold,  $z_x$ , below which they exit the market. Because the adoption choice is irreversible, firms that have already adopted no longer make a technology choice and only choose when to exit. I first describe the dynamic exit decision of firms that have already made the irreversible technology adoption choice before describing the dynamic problem of a firm that has not yet adopted.

### Adopters

Firms that have adopted the new technology solve the static profit maximization problem, described in Section 4.1, and an optimal stopping time problem analogous to the dynamic problem of a firm in the economy with no technology adoption choice, described in Section 3.2. Their value function is given by:

$$V^a(z) = \max_{z_x^a \leq z} \mathbb{E}_z \int_0^{T(z_x^a)} e^{-\rho t} (\pi_j^{av}(z(t)) - (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c - (\theta_j^{ha} w_{hj} + \theta_j^{la} w_{lj}) \Gamma_{fc} - r_j f^b) dt$$

$$d \log(z_t) = \mu dt + \Psi dW(t)$$

where  $\pi_j^{av}(z)$  is variable profits of a firm in city  $j$  with productivity  $z$  that has adopted the new technology. As before, I use a change of variable  $s(z) = \ln z^{\sigma-1}$ . The value function takes the same form as in the case without adoption:

$$V^a(s) = \frac{Z_j^a}{\rho - \tilde{\mu} - \frac{1}{2} \tilde{\Psi}^2} e^s + \frac{1}{\rho} [ -(\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c - w_{hj} \Gamma_{fc} - r_j f^b ] \dots$$

$$+ \left[ \frac{-Z_j^a}{\rho - \tilde{\mu} - \frac{1}{2} \tilde{\Psi}^2} e^{s_x^a - \xi^- s_x^a} - \frac{1}{\rho} [ -(\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c - (\theta_j^{ha} w_{hj} + \theta_j^{la} w_{lj}) \Gamma_{fc} - r_j f^b ] e^{-\xi^- s_x^a} \right] e^{\xi^- s}$$

and the exit threshold for firms that have adopted is

$$s_x^a = \ln \left[ \frac{1}{Z_j^a} \left( \frac{-\xi^-}{1 - \xi^-} \right) \left( \frac{\rho - \tilde{\mu} - \frac{1}{2} \tilde{\Psi}^2}{\rho} \right) ((\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c + (\theta_j^{ha} w_{hj} + \theta_j^{la} w_{lj}) \Gamma_{fc} + r_j f^b) \right]$$

where  $Z_j^a = \left( (\Gamma_\psi \psi_j)^{-1} \left( (\Gamma_\gamma \gamma_j)^\epsilon w_{jh}^{1-\epsilon} + (1 - \Gamma_\gamma \gamma_j)^\epsilon w_{jl}^{1-\epsilon} \right) \frac{1}{1-\epsilon} \frac{\sigma}{\sigma-1} \right)^{1-\sigma} Y_j P_j^\sigma \frac{1}{\sigma}$ .

### Non-Adopters

Firms that have not yet adopted solve for two thresholds: an adoption threshold  $z_a$  and an exit threshold  $z_x$ . The value function for the firm is the expected discounted stream of profits plus the



expected discounted value of adopting,  $v^a(s(z_a))$ , times the probability that the firm reaches the adoption threshold  $z_a$  before the exit threshold  $z_x$ . Let  $T = T(z_x) \wedge T(z_a)$  be the stopping time, the discounted stream of profits is given by:

$$F_j(z, z_x, z_a) = \begin{cases} \mathbb{E}_z[\int_0^T e^{-\rho t} (\pi_j(z(t)) - (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c - r_j f^b) dt \dots & , z \in (z_x, z_a) \\ + [e^{-\rho T} |z(T) = z_{aj}] Pr [z(T) = z_{aj}] V_j^a(z_{aj}) & \\ V^a(z_{aj}) & z > z_{aj} \end{cases}$$

$$d \log(z_t) = \mu dt + \Psi dW(t).$$

As above, I use a change of variable  $s(z) = \ln z^{\sigma-1}$  to rewrite the expected profit function as

$$F(s, s_x, s_a) = \begin{cases} \mathbb{E}_s[\int_0^T e^{-\rho t} \left( \frac{z_j}{\rho - \tilde{\mu} - \frac{1}{2} \tilde{\Psi}^2} e^s - (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c - r_j f^b \right) dt \dots & , s \in (s_x, s_a) \\ + [e^{-\rho T} |s(T) = s_a] Pr [s(T) = s_a] (V_j^a(s_a)) & \\ V^a(s_a) & s > s_a \end{cases}$$

$$ds_t = \tilde{\mu} dt + \tilde{\Psi} dW(t).$$

The value function is defined as

$$V(s) = \max_{s_x, s_a} F(s, s_x, s_a),$$

i.e. the discounted stream of profits when choosing the optimal exit and adoption thresholds. The value function will satisfy the HJB equation

$$\rho V(s) = \begin{cases} \left[ \pi_j^v(s) - (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^c - r_j f^b \right] + \tilde{\mu} V'(s) + \frac{1}{2} \sigma^2 V''(s), & s \in (s_x, s_a) \\ \rho [V^a(s_a)], & s \geq s_a \end{cases}$$

and the conditions:

$$\text{value matching: } V(s_x) = 0; \quad V(s_a) = V^a(s_a)$$

$$\text{smooth pasting: } V'(s_x) = 0; \quad \lim_{z \rightarrow z_a} V'(s_a) - V^{a'}(s_a) = 0$$

There is no closed form solution for the value function and the exit and adoption thresholds. In Appendix B.3, I derive the value function as a function of  $s_x$  and  $s_a$  and provide the two smooth pasting conditions that give the optimal thresholds when solved numerically.

### 4.3 The distribution of firms with adoption

I solve for the distribution of adopters and the distribution of non-adopters separately. As in Section 3.3, the stochastic process for firm size is  $\frac{ds}{dt} = \tilde{\mu} dt + \tilde{\Psi} dW(t)$  and the distribution of firms satisfies

the Kolmogorov Forward Equation

$$\frac{\partial g_j(s)}{\partial t} = -\tilde{\mu} \frac{dg_j(s)}{ds} + \frac{1}{2} \tilde{\psi}^2 \frac{d^2 g_j(s)}{ds^2}.$$

In steady state  $\frac{\partial g_j(s)}{\partial t} = 0$  and  $\delta = \frac{-\tilde{\mu}}{\tilde{\psi}^2/2}$ . In Appendix B.4, I show that the distribution of adopters is given by<sup>17</sup>

$$f_j^a(z) = \begin{cases} \frac{E_j^a}{\tilde{\mu}} \left( z^{\delta(1-\sigma)} e^{\delta s_{xj}} - 1 \right) \left( \frac{\sigma-1}{z} \right) & z \in (z_{xj}^a, z_{aj}) \\ \frac{E_j^a}{\tilde{\mu}} \left( e^{\delta s_{xj}^a} - e^{\delta s_a} \right) z^{\delta(1-\sigma)} \left( \frac{\sigma-1}{z} \right) & z > z_{aj} \end{cases}$$

and the distribution of non-adopters is given by

$$f_j^n(z) = \begin{cases} \frac{E_j^n}{\tilde{\mu}} \left( z^{\delta(1-\sigma)} e^{\delta s_{xj}} - 1 \right) \left( \frac{\sigma-1}{z} \right) & z \in (z_{xj}, z_e) \\ \frac{E_j^n}{\tilde{\mu}} \left( \frac{e^{-\delta(s_e - s_{xj})} - 1}{e^{-\delta s_e} - e^{-\delta s_{aj}}} \right) \left( z^{\delta(1-\sigma)} - e^{-\delta s_{aj}} \right) \left( \frac{\sigma-1}{z} \right) & z \in (z_e, z_{aj}) \end{cases}.$$

Integrating both of these over  $z$  to get the mass of adopters and non-adopters and rearranging gives

$$sr_j^a = \frac{E_j^a}{M_j^a} = \frac{-\tilde{\mu}}{s_{aj} - s_{xj}^a}$$

and

$$sr_j^n = \frac{E_j^n}{M_j^n} = \frac{-\tilde{\mu}}{(s_e - s_{xj}) + \left( \frac{e^{-\delta(s_e - s_{xj})} - 1}{e^{-\delta(s_e - s_{aj})} - 1} \right) (s_{aj} - s_e)}$$

where  $E_j^a$  and  $E_j^n$  are the mass of adopters who exit and non-adopters that exit, respectively. Finally in Appendix B.4, I show that  $E_j^n = -E_j^a \left( \frac{e^{-\delta(s_e - s_{xj})} - 1}{e^{-\delta s_e} - e^{-\delta s_{aj}}} \right) e^{-\delta s_{aj}}$  which means both distributions can be written as a function of  $E_j^a$ , and the mass of firms  $M_j = M_j^a + M_j^n$  can be written as a function of  $E_j^a$ .

In steady state, the mass of new adopters,  $A$ , must be equal to the mass of adopters that exit,  $E^a$  and the mass of entrants,  $N$ , must be equal to the mass of exitors,  $N_j = E_j^n + E_j^a$ . Finally, the share of adopters is given by

$$\begin{aligned} \mu_j &= \frac{M_j^a}{M_j^a + M_j^n} \dots \\ &= \frac{(s_{xj}^a - s_{aj})}{(s_{xj}^a - s_{aj}) + \left( \frac{e^{-\delta(s_e - s_{xj})} - 1}{e^{-\delta s_e} - e^{-\delta s_{aj}}} \right) e^{-\delta s_{aj}} \left[ (s_e - s_{xj}) + \left( \frac{e^{-\delta(s_e - s_{xj})} - 1}{e^{-\delta s_e} - e^{-\delta s_{aj}}} \right) e^{-\delta s_{aj}} (s_{aj} - s_e) \right]}. \end{aligned}$$

Thus, the share of adopters, the start-up rate for adopters and the start-up rate for non-adopters can all be characterized by three objects: the difference between the entry threshold and exit threshold for non-adopters, the difference between the adoption threshold and the exit threshold for adopters, and the difference between the entry threshold and the adoption threshold. In Appendix D.4, I discuss how the churn rates and the share of adopters move with each threshold, and I discuss how

<sup>17</sup>Note that here I am giving the distribution over productivities  $f(z)$  instead of the distribution over sizes  $g(s)$ . In the appendix, I derive the distribution over size  $g(s)$  and then use a change of basis to find the distribution over  $z$ .

each threshold moves with endogenous objects such as prices and market size.

#### 4.4 Steady-state equilibrium with new technology

The equilibrium in the model with the technology adoption decision is very similar to the model without adoption, where the only difference with the model without adoption is the inclusion of the additional fixed cost for adopters. The final goods market clearing condition simplifies to

$$Y_j = w_{H_j}H_j + w_{L_j}L_j + \Pi_j^v - (\theta_j^h w_{h_j} + \theta_j^l w_{l_j})M_j^n f_j^c - (\theta_j^{ha} w_{h_j} + \theta_j^{la} w_{l_j})M_j^a \Gamma^{fc} - (\theta_j^h w_{h_j} + \theta_j^l w_{l_j})N_j f_j^e.$$

A steady-state equilibrium is a set of prices, labor allocations, output and the mass of adopters, non-adopters, entrants and new adopters, and exit and adoption thresholds

$\{w_{l_j}, w_{h_j}, r_j, L_j, H_j, Y_j, M_j^a, M_j^n, N_j, A_j, z_{x_j}, z_{x_j}^a, z_{a_j}\}$ , for each city, such that

1. in each city, there is an invariant distribution of adopters and non-adopters that satisfies the Kolmogorov Forward Equations with  $\frac{\partial g_j^a(s)}{\partial t} = 0$  and  $\frac{\partial g_j^n(s)}{\partial t} = 0$ ,
2. there is a constant mass of adopters, non-adopters, entrants and new adopters,  $M_j^a$ ,  $M_j^n$ ,  $N_j$  and  $A_j$ ,
3. exit and adoption policies satisfy the firms HJB equations, and the free entry conditions hold in each city,
4. firms and landlords maximize profits,
5. workers choose the city that maximizes utility and the share of their income to spend on consumption and housing, and
6. the final goods market, the housing market, and labor markets clear in each city.

#### 4.5 Calibration of the new technology

There are three parameters of the new technology that need to be calibrated:  $\Gamma_\gamma$ , which scales the weight on high-skilled labor;  $\Gamma_\psi$ , which scales the absolute productivity advantage of the new technology; and  $\Gamma_{fc}$ , the additional fixed cost paid in units of high-skilled labor. I calibrate the parameters to match moments on the aggregate growth in high- and low-skilled wages between 1980 and 2014 and on the change in average firm size. In addition, [Katz and Murphy \(1992\)](#) show that in order to understand the extent of SBTC, it is important to consider changes in the relative supply of high-skilled workers. I change the aggregate quantities of high- and low-skilled labor,  $H$  and  $L$ , to match the growth in high- and low-skilled labor between 1980 and 2014.

The five parameters and the corresponding targets are listed in [Table VI](#). The model does a good job of matching wage growth for high-skilled workers, but understates wage growth for low-skilled workers. This is because I am limiting the change in technology to the introduction of a new technology that favors high-skilled workers with no change in the old technology. In [Appendix C.2](#),

I show how the targeted moments change with the parameters of the new technology and show that the main results of the paper are robust to variations in these parameters.

## 5 Results

### 5.1 Model predictions versus the data

Figure VI shows that the introduction of the skill-biased technology can reproduce the patterns in the data described in Section 2. In 1980, the model perfectly matches the relationship in the data between wages, skill intensity, and dynamism with city-size. I then show these relationships in the 2014 data and in the second steady state in the model. In the second steady state, the only things that changed are the aggregate quantities of high- and low-skilled labor and the availability of the new technology.

In the data, the city-size wage premium for high-skilled workers increased from .056 percent to .072 percent. In the model, the high-skilled city-size wage premium increases from .056 to .065 percent. Thus, the model can explain 51 percent of the increase in the city-size wage premium for high-skilled workers. Similarly, the city-size wage premium for low-skilled workers decreases from .047 to .035 percent in the data, while in the model it decreases from .047 to .031 percent. Thus, the model slightly over-explains the decline. The model is successful in matching the divergence by skill group; the city-size wage premium rises for high-skilled workers while it falls for low-skilled workers.

The model also explains a substantial share of the changing relationship between the skill premium and city-size. In 1980, the skill premium is, on average, 1.00 percentage points higher in cities twice as large. By 2014, the skill premium was 5.12 percentage points higher in cities twice as large. In the model, the relationship between the skill premium and city-size increases from 1.00 to 4.64 percentage points. Thus the model explains 94 percent of the increase in the slope.

The model also matches the changing relationship of skill intensity, or the ratio of high to low-skilled workers, with city-size. In 1980, the semi-elasticity between skill intensity and city-size was .026 percentage points. By 2014, it had increased to .074 percentage points. The model matches this fact well. The relationship between skill intensity and city-size increases from .026 in the initial steady state to .085 in the final steady state. Thus, the model over explains the changing relationship between city-size and skill intensity.

Turning to the establishment start-up rate, in the data the semi-elasticity between the establishment start-up rate and city-size increased from  $-.322$  to .45 percentage points, while in the model it increased from  $-.322$  to .074 percentage points. Thus, the model can explain 51 percent of the changing slope. Next, I discuss the mechanisms that are important for driving the changing relationship between dynamism and city-size.

One concern is that the success of the model in generating the changing cross-sectional patterns is driven by the persistence of variables such as population and wages. In Appendix D.1, I further test the model in changes. For each of the four main variables, high-skilled wages, low-skilled

wages, skill intensity and business dynamism, I show the relationship between the change in the data between 1980 and 2014 and the change in the model between the first and second steady state. The correlation between the change in the data and the change in the model is positive for each of the four main variables.

## 5.2 The changing relationship between dynamism and city-size

The start-up rate in city  $j$  is given by

$$sr_j = (1 - \mu_j)sr_j^{na} + \mu_jsr_j^a,$$

a weighted average of the start-up rate for non-adopters and adopters where the weight is the share of firms that have adopted,  $\mu_j$ . In Appendix D.4, I characterize the three objects in this equation  $sr_j^{na}$ ,  $sr_j^a$  and  $\mu_j$ , and I discuss in detail how they are determined by three thresholds: the exit threshold for adopters, the exit threshold for non-adopters and the adoption threshold. I show how each threshold and, therefore, the churn rates are affected by endogenous objects such as wages, rents, and market size and the parameter  $\gamma$ , which gives the weight on high-skilled labor in the initial equilibrium.

In this section, I discuss which factors are important for driving the changing equilibrium relationship of dynamism with city-size. In Figure VII, I plot the counterfactual change in dynamism rates versus city-size between the 1980 and 2014 steady states, holding each of the endogenous objects fixed across cities (or in the case of market size I consider what would have happened if the growth rate of market size had been equal across cities).<sup>18</sup>

Panel (a) shows what the change in dynamism would have been if rent had remained constant at the 1980 level. In the baseline model, the change in dynamism is about 2 percentage points higher in the biggest city relative to the smallest cities. Holding rent constant, the difference across cities largely disappears. Rent increases affect the fixed costs paid by the firm. When fixed costs are high, firms need a higher volume of sales to justify staying in the market. The exit threshold shifts up meaning that for any given productivity, firms are more likely to exit the market. In steady state, entry and exit rates must be equal so an increase in the exit threshold means that both entry and exit rates will increase. If there had been no endogenous response of rent to the availability and adoption of the new technology, the start-up rate would not increase more in big cities relative to small cities. Therefore, the increases in rent, which are larger in big cities than small cities and in turn drive changes in selection, are an important driver of the changing cross sectional pattern of dynamism and city-size. In Appendix D.3, I show that the model does a good job of matching the changes in rent that occurred in the data.

Panel (b) of Figure VII shows what would have happened had every city experienced equal

---

<sup>18</sup>I do this because holding market size at the 1980 level, the exit threshold will rise above the entry threshold and there will be no firms in the market.

growth in the market size,  $Y_j$ . Though market size grows everywhere, it actually grows more in small cities than in large cities. This is because of the inelastic housing supply in big cities, which limits their growth. In Appendix D.2, I show that the model matches the relationship between population growth and initial city-size in the data. An increase in market size increases profits for a firm and makes firms more willing to stay in the market for any given productivity. This lowers the exit threshold and decreases the city level churn rate. Thus if market size had increased the same everywhere, the change in dynamism rates would have been smaller in big cities since the growth in  $Y$  was bigger in small cities than in big cities. Thus, the endogenous response of  $Y$  contributes to the increasing relationship of dynamism and city-size in the second steady state.

Panels (c) and (d) look at the effect of wages on the change in dynamism with city-size. In panel (c), I hold both high- and low-skilled wages fixed at their 1980 level, and in panel (d), I allow low-skilled wages to grow as they did in the second steady state but adjust high-skilled wages so that the skill premium is held at its 1980 level. Lower wages imply weaker selection - firms are more willing to stay in the market for any given productivity level. The two figures have similar effects on the cross sectional pattern of dynamism, implying most of the effect on dynamism comes from the change in high-skilled wages<sup>19</sup>. In both counterfactuals, the differences across cities in the change in dynamism would have been attenuated - big and small cities would have seen similar changes in dynamism rates without the equilibrium response of wages, particularly high-skilled wages, to the introduction of the new technology.

## 6 Technology Adoption

A key result in the quantitative exercise is that technology adoption is higher in larger cities. This happens despite the fact that all firms in all cities have access to the new technology. The goal of this section is to determine what factors drive differences in adoption. Section 6.1 examines which city characteristics are important for driving the differences Section 6.2 uses data on ICT spending to validate these mechanisms in the data.

### 6.1 The drivers of technology adoption

Here, I discuss four channels driving differences in adoption rates across cities: market size, initial technology, selection, and amenities. In Figure VIII, I show how the share of firms that adopt the new technology will change as each channel is strengthened or weakened. I do this for one city in the calibrated model as an illustration.

The first channel affecting adoption rates across cities is the market size effect. In cities that are large, firms will be more willing to pay the additional fixed cost associated with adopting the new

---

<sup>19</sup>In other words, both figures have the same 1980 skill premium, but panel (c) keeps low-skilled and high-skilled wages at the 1980 level while panel (d) allows only low-skilled wages to adjust to the 2014 level. Since we see most of the change in dynamism rates when we only hold the skill premium constant (panel d), this implies most of the effect is coming from the change in high-skilled wages.

technology because the marginal cost savings are more valuable when there is a larger volume of sales. Panel (a) of Figure VIII shows what happens to the share of firms that adopt and the three thresholds governing the firm size distribution as the market size, given by  $Y$ , changes. The share of firms that adopt the new technology increases with market size.

Panel (b) shows the change in the share of firms adopting as the skill intensity of the city-level technology changes. As  $\gamma_j$  increases, the share of firms who adopt the new technology increases. This is true even though the weight on high-skilled labor in the new technology is proportional to the weight in the old technology. This is driven by two factors. First, as  $\gamma_j$  increases, non-adopters use high-skilled labor more intensively and have more expensive labor costs, but do not get the productivity advantage associated with the new technology. As a result, the exit threshold increases, and the non-adopters exit the market. The market share of adopters increases, which increases the returns to adopting and lowers the exit threshold for adopters. Eventually the exit threshold for non-adopters rises above the entry threshold and there will be no firms in the market (the lines end at this point). Second, as  $\gamma_j$  increases the difference between the labor cost bundle for adopters and non-adopters increases even though the weight on high-skilled labor for the new technology is scaled proportionally. As the difference between the cost bundles grows, so does the benefit to adopting and more firms become adopters.

The third factor affecting adoption rates is selection. This can be seen in panel (c) of Figure VIII which shows what happens to adoption rates and the exit thresholds as rent increases. As rent increases, the fixed costs firms need to pay to stay in the market increase. Non-adopters are hurt especially hard by the increase in rent since they have a lower volume of sales to cover the fixed cost. The exit threshold for non-adopters increases, meaning that less of them stay in the market, and the share of firms that have adopted the new technology increases.

Finally, a fourth factor affecting adoption rates is amenities for high-skilled workers, which lower the relative wages paid to high-skilled workers. If wages for high skilled workers are high, the returns to adopting will be lower since the new technology uses high skilled labor more intensively. Panel (d) plots what happens to the share of firms that adopt the new technology and the adoption and exit thresholds as high skilled wages increase. As high skilled wages increase, labor becomes more expensive for both adopters and non-adopters and both types of firms are more likely to exit the market. Thus, the change in adoption rates is primarily drive by the change in the adoption threshold. As high skilled wages increase, the adoption threshold increases and the share of adopters declines.

## 6.2 Model validation

As discussed in the previous section, there are four channels that influence adoption rates: market size, initial technology, relative wages and selection. In Table VII, I validate the mechanisms in the model by showing the relationship of ICT spending per employee today versus 1980 city characteristics, and that the same relationships hold for the share of firms adopting the new technology in the model. While ICT spending is not a binary measure of firm technology adoption, it does

capture firms' investments decisions in new technologies, and thus provides suggestive evidence on the model's mechanisms. In Appendix E.1, Table A.III, I show that the same relationships hold using ICT share of investment, instead of ICT spending per employee, as the dependent variable.

As a measure of market size, I use 1980 working age population. The elasticity between ICT spending per employee today and 1980 population is 5.65 percent, meaning that ICT spending per employee is 3.99 percent higher today in markets that were twice as large in 1980. Similarly, in the model, the share of adopters is 1.16 percentage points higher in cities that were initially twice as large. This provides suggestive evidence that market size is an important factor in a firms decision to adopt new technologies.

Two additional factors influencing the technology adoption decision of firms are the weight on high-skilled labor in the initial technology used in the city ( $\gamma_j$ ) and the amenities for high-skilled labor. Firms that are in locations which are already good at using high skilled labor will be more likely to adopt the new technology. In such locations we expect to see high initial skill intensity and a high initial skill premium. This would drive a positive correlation between initial skill intensity and subsequent adoption and a positive correlation between the initial skill premium and subsequent adoption.

On the other hand, locations that have abundant amenities for high-skilled labor will, all else equal, have a larger supply of high-skilled labor and a lower skill premium, increasing the return to adoption. This would drive a positive correlation between initial skill intensity and subsequent adoption and a negative correlation between the initial skill premium and subsequent adoption.

The initial technology ( $\gamma_j$ ) and the amenities will jointly determine the initial skill mix of the city and the initial skill premium. Thus, we expect a positive relationship between the initial skill intensity and subsequent adoption, driven by both technology and amenities. On the other hand, the expected relationship between the initial skill premium and subsequent adoption is ambiguous. A higher weight on high-skilled labor will imply a high initial skill premium and higher subsequent adoption, while abundant amenities for high-skilled workers will imply a lower initial skill premium and higher subsequent adoption.

I examine the relationship between both measures, the initial skill intensity and the initial skill premium, and ICT spending. There is a positive correlation between ICT spending and initial skill intensity. Cities with a one percentage point higher skill intensity in 1980 have about one percent more ICT spending per employee today. This provides suggestive evidence for both the initial technology mechanism and the labor supply mechanism. Both abundant amenities for high skilled workers and an initial technology that is more skill intensive will make the initial population more high-skilled. The data show a positive relationship between ICT spending and initial skill premium, which provides suggestive evidence for the importance of initial technology on subsequent technology adoption. Though, consistent with the ex-ante ambiguity of the two mechanisms, controlling for skill intensity decreases the size and significance of the relationship between ICT spending and initial skill-premium. The model weakly matches this relationship when looking at the correlation between the share of adopters and the initial skill premium, but after controlling for initial skill



intensity the partial correlation is negative.

## 7 Counterfactuals

In this section, I use the model to undertake two counterfactuals. In the first I ask what would happen if there were no migration. The first counterfactual is motivated by policies that have the goal of attracting high-skilled workers and firms to small cities. I ask what would happen if a policy maker could do this perfectly, and keep a small city’s share of the labor force constant at its 1980 level.

In the second counterfactual, I ask what would happen if housing supply elasticities were equalized across cities. This counterfactual is motivated by recent papers ([Ganong and Shoag, 2017](#); [Herkenhoff et al., 2018](#); [Hsieh and Moretti, 2019](#)) that show housing supply constraints keep workers from moving to cities that have experienced high TFP growth. Using data on housing supply elasticities from [Saiz, 2010](#), I find that big cities are more constrained in their housing supply than small cities. I ask what would happen if housing supply became more elastic in big frontier cities.

The two counterfactuals are motivated by opposite views on using mobility to affect the geographic distribution of welfare. The first has the view that small cities have been hurt by out-migration of high-skilled workers (“brain drain”) and limits mobility to keep high-skilled workers and firms in small cities. The second has the view that low-skilled workers are hurt by being kept out of the big productive cities and seeks to reallocate more people to big cities even at the expense of small cities.

In examining the counterfactuals it is important to understand the effect on the aggregate rate of technology adoption and, therefore, technical change in addition to the usual considerations of welfare. I find that this tradeoff is relevant when considering these two policies. While a policy of no migration is unambiguously bad for welfare, it slightly increases aggregate adoption rates and, therefore, may be good for technological progress. On the other hand, equalizing housing supply elasticities is unambiguously good for welfare but it decreases aggregate technology adoption.

### 7.1 No migration

Motivated by the view that small cities are suffering from “brain drain,” or the out migration of their high-skilled workers, the first counterfactual I undertake is to ask what would happen if there were no migration in the model. Specifically, I change the aggregate quantities of high- and low-skilled labor as in the data, but I keep each city’s share of the population of high- and low-skilled workers constant at its 1980 level. Panel (a) of [Figure IX](#) shows the relationship between skill intensity and city size in the second steady state of the model and the counterfactual world with no migration. Without migration, skill-intensity declines in the big cities while it increases in the small cities.

I compare the welfare of a worker who stays in the same city  $j$  between a world with no migration and the 2014 steady state with free mobility. Specifically, I look at the indirect utility of a worker

of type  $\tau$  in city  $j$  with idiosyncratic preference  $\zeta_{ij}$

$$V_{\tau j i}^{2014} = A_{\tau j} (c_{\tau j}(w_{\tau j}^{2014}, q_j^{2014}))^\beta (d_{\tau j}(w_{\tau j}^{2014}, q_j^{2014}))^{1-\beta} e^{\zeta_{ij}},$$

and I ask how much consumption a worker would give up in order to avoid living in the counterfactual world with no migration. The consumption equivalent,  $\Delta_c$ , solves

$$A_{\tau j} (\Delta_c c_{\tau j}(w_{\tau j}^{2014}, q_j^{2014}))^\beta (d_{\tau j}(w_{\tau j}^{2014}, q_j^{2014}))^{1-\beta} e^{\zeta_{ij}} = V_{\tau j}^{\text{no migration}}. \quad (2)$$

Only workers with a particularly high preference for a certain city will remain there once they are allowed to migrate, but since their preferences are fixed, they do not affect the calculation of their consumption equivalent. However, the welfare calculation ignores the welfare gains of the worker who was allowed to move cities.

Table VIII shows what happens to the aggregate share of firms that adopt the new technology, aggregate wages and the aggregate consumption equivalent, and Figure IX shows the relationship of these measures with city-size.

Average wages increase slightly for both high- and low-skilled workers in the model with no migration versus the 2014 steady state. However, looking at the average consumption equivalent measure, the policy makes both types of workers worse off. For high-skilled workers, the average consumption equivalent is  $-.29$  meaning that high-skilled workers would, on average, give up .29 percent of their consumption in the 2014 steady state to avoid living in the world with no migration. The welfare costs for low-skilled workers are much higher. They are, on average, willing to give up 1.55 percent of their consumption to remain in the 2014 steady state with free mobility.

Panel (b) of Figure IX plots the consumption equivalent for high- and low-skilled workers versus city-size in the 2014 steady state. For high-skilled workers, the consumption equivalent is negative in the small cities while positive in the largest city, meaning the policy benefits high-skilled workers in the largest city only. This is not surprising since the high-skilled workers are forced to stay in the small cities where there is less adoption. High-skilled workers in the big city benefit since there is no in-migration that puts downward pressure on their wages. For low-skilled workers the welfare costs are larger in the big city. This is because of the effect of the no migration policy on rents. Wages for low-skilled workers are actually slightly higher in the big city in the no-migration steady state, but rents are higher as well. This is because population is slightly higher in the big city when low-skilled workers are not allowed to migrate to smaller, less congested cities<sup>20</sup>.

Looking at panel (c), the no-migration policy increases the adoption rates in the small city and decreases adoption in the big city since high-skilled workers have become more abundant in small cities and less abundant in big cities. However, the increase in adoption in the small city does

---

<sup>20</sup>Though counterintuitive, in Appendix A.X, I show that there is a negative relationship between population growth and initial city-size in the model and in the data. This is driven by the lower housing supply elasticities in big cities that keep their population from growing even though they are becoming more productive. This is consistent with evidence from Rappaport (2018).

not increase high-skilled labor demand enough to offset the increase in supply. As a result, wages for high-skilled workers in small cities fall. Panel (d) shows that wages for high-skilled workers in the small city fall in the world with no migration. Overall, there is a slight increase in aggregate adoption rates, but even though adoption rates increase in the small cities, this does not result in wage increases in the small cities. Panel (d) shows that wages, on average, decreased in small cities for both high- and low-skilled workers even though adoption rates went up.

## 7.2 Elasticity of housing supply

Recent work from Ganong and Shoag (2017), Herkenhoff et al. (2018), and Hsieh and Moretti (2019) identifies housing supply constraints as an important drag on aggregate growth. In this counterfactual, I ask what would happen if we could equalize the housing supply elasticities across cities. Using data on housing supply elasticities from Saiz (2010), panel (a) of Figure X shows that big cities have a lower housing supply elasticity than small cities.

Housing supply in a city is given by

$$B_j = b_j r_j^{\frac{1}{\eta_j - 1}}.$$

In the counterfactual with equalized housing supply elasticities, I set  $\eta_j = \frac{1}{J} \sum \eta_j \forall j$ . I also adjust the productivity of the building sector in each city,  $b_j$ , to keep the building supply constant in the 1980 steady state.

Panel (b) of Figure X plots the consumption equivalence measures, defined in equation 2, for high- and low-skilled workers versus 2014 city-size. For both high- and low-skilled workers, the consumption equivalence measure is increasing in city-size. Both worker types benefit from relaxing the inelastic housing supply in big cities and they are hurt by a more inelastic supply in small cities. However, as shown in Table VIII, the average consumption equivalents are resoundingly positive for both types of workers meaning workers prefer the steady state with equalized housing supply elasticities. High- and low-skilled workers would need to be compensated with 8.6 and 6.1 percent more consumption, respectively, in the 2014 steady state to be indifferent between the 2014 steady state and the one with equalized housing supply elasticities.

Both types are hurt by the policy in the smallest cities. This is because of the equilibrium effect on market size. When the housing supply elasticities are equalized, big cities get bigger and small cities get smaller. There is an agglomeration force in the model through the CES specification of the final-goods producer. When the small cities get smaller, they become less productive and wages go down for both types of workers. Since housing supply became more inelastic in the small city, rents do not fall to compensate the workers for the fall in wages. Panel (d) shows what happens to rents in the model with equalized housing supply elasticities. Even though the small cities are smaller than in the 2014 steady state, their rents increase because their housing supplies became less elastic.

Table VIII shows what happens to aggregate adoption rates in the steady state with equal

housing supply elasticities, and Panel (c) of Figure X shows what happens to adoption versus city-size. Adoption rates go down in the big city and up in small cities. In the aggregate, the adoption rate falls slightly. This is because rents (shown in panel d) in the counterfactual world are much lower than in the second steady state. Lower rents decrease the fixed costs paid by firms who, as a result, are more willing to stay in the market even if they are less-productive non-adopters. Thus, the tougher selection in big cities, stemming from their inelastic housing supply, amplifies the differences in adoption rates and, therefore, SBTC across cities.

Finally, panel (e) shows the effect on the cross-sectional pattern of dynamism rates from equalizing housing supply elasticities. As shown in Section 5.2, changes in selection driven by the equilibrium response of rent is an important driver of the cross sectional changes in dynamism. Thus, it is not surprising that a policy of equalizing housing supply elasticities is partially effective in restoring the cross-sectional pattern of dynamism rates to its 1980 level. Between the initial and final steady state, the relationship between dynamism and city-size increases from  $-.322$  to  $.074$ . In the model with equalized housing supply elasticities the relationship between dynamism and city-size is  $-.05$ . Thus, the increase in the relationship between dynamism and city-size is about two thirds of the increase in the baseline model.

## 8 Conclusion

I document that, since 1980, big and small cities have diverged on several important dimensions. In 1980, big cities paid a similar wage premium to both high- and low-skilled workers. By 2014, the city-size wage premium for high-skilled workers was twice that of low-skilled workers. Furthermore, the relationship between skill intensity and city-size and between dynamism and city-size both increased over time. While facts about the divergence of the city-size wage premium and skill intensity and city-size have been previously documented in the literature, the changing relationship of dynamism and city-size is a new fact that I introduce.

I build a spatial equilibrium model that includes a model of firm dynamics embedded in each city. Adding a rich model of firm dynamics to an otherwise benchmark spatial equilibrium model allows the joint consideration of relative wage inequality and business dynamism across cities. I calibrate the model to the 1980 steady state and show that it can match the salient features of the 1980 data.

I then use the model to consider the introduction of a new skill-biased technology. The new technology has an absolute productivity advantage but uses skill more intensively and incurs a higher fixed cost. Firms that are otherwise the same will make different adoption decisions based on the characteristics of the city in which they are located. Firms adopt more when they are in a city that is big, one that is better at using high-skilled labor and one that has abundant amenities for high-skilled labor. Cities in which there is more adoption become more skill intensive, more competitive, and more congested, driving an increase in high-skilled wages and rents. Smaller, less-productive firms that do not find it profitable to adopt the new technology find it harder to compete

and exit the market, which amplifies the differences in adoption rates across cities and changes the cross-sectional relationship between city size and business dynamism. Furthermore, the selection force puts additional downward pressure on low-skilled wages since these firms use low-skilled labor more productively.

Even though the new technology is available everywhere, the introduction of the technology favors specific kinds of labor and market characteristics, amplifying existing differences across cities. The introduction of the new skill-biased technology that is adopted differently across cities can quantitatively account for the divergence between big and small cities. Specifically, it can account for the changes I document across cities in the divergence of the city-size wage premium by skill group, the increasing relationship between skill intensity and city-size, and the increasing relationship between dynamism and city-size.

As evidence for this mechanism I use data on ICT spending to document a novel fact: firms in big cities spend more per worker on ICT investment than firms in small cities and they devote a higher share of their investment budget to ICT expenses.

Firms in big cities adopt more, and as they do so, they increase their demand for skilled labor relative to unskilled labor, driving an increase in the city-size wage premium for high-skilled workers, a decrease in the city-size wage premium for low-skilled workers and an increasing relationship between skill intensity and city-size. As big cities become more competitive and more congested, it becomes harder for small, less-productive firms to stay in the market, which increases the equilibrium rate of turnover in big cities relative to small cities. Small firms are better at using low-skilled labor than big firms that have adopted the new technology. As small firms exit the market, it amplifies the differences across cities in the extent of SBTC.

## Works Cited

- ALLEN, T., C. DE CASTRO DOBBIN, AND M. MORTEN (2018): “Border Walls,” NBER Working Papers 25267, National Bureau of Economic Research, Inc.
- ASPLUND, M. AND V. NOCKE (2006): “Firm Turnover in Imperfectly Competitive Markets,” *Review of Economic Studies*, 73, 295–327.
- AUSTIN, B., E. GLAESER, AND L. SUMMERS (2018): “Jobs for the Heartland: Place-Based Policies in 21st-Century America,” *Brookings Papers on Economic Activity*, 49, 151–255.
- AUTOR, D. (2019): “Work of the Past, Work of the Future,” NBER Working Papers 25588, National Bureau of Economic Research, Inc.
- AUTOR, D., D. DORN, AND G. HANSON (2019): “When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men,” *American Economic Review: Insights*, 1, 161–178.
- AUTOR, D., D. DORN, G. HANSON, AND K. MAJLESI (2016): “Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure,” NBER Working Papers 22637, National Bureau of Economic Research, Inc.
- AUTOR, D. H. AND D. DORN (2013): “The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market,” *American Economic Review*, 103, 1553–1597.
- AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2008): “Trends in U.S. Wage Inequality: Revising the Revisionists,” *The Review of Economics and Statistics*, 90, 300–323.
- AUTOR, D. H., L. F. KATZ, AND A. B. KRUEGER (1998): “Computing Inequality: Have Computers Changed the Labor Market?” *The Quarterly Journal of Economics*, 113, 1169–1213.
- BAUM-SNOW, N., M. FREEDMAN, AND R. PAVAN (2018): “Why Has Urban Inequality Increased?” *American Economic Journal: Applied Economics*, 10, 1–42.
- BAUM-SNOW, N. AND R. PAVAN (2013): “Inequality and City Size,” *The Review of Economics and Statistics*, 95, 1535–1548.
- BEAUDRY, P., M. DOMS, AND E. LEWIS (2010): “Should the Personal Computer Be Considered a Technological Revolution? Evidence from U.S. Metropolitan Areas,” *Journal of Political Economy*, 118, 988–1036.
- BRODA, C. AND D. E. WEINSTEIN (2006): “Globalization and the Gains From Variety,” *The Quarterly Journal of Economics*, 121, 541–585.
- BUSTOS, P. (2011): “Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms,” *American Economic Review*, 101, 304–340.

- CHETTY, R. AND N. HENDREN (2018): “The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates,” *The Quarterly Journal of Economics*, 133, 1163–1228.
- CHETTY, R., M. STEPNER, S. ABRAHAM, S. LIN, B. SCUDERI, N. TURNER, A. BERGERON, AND D. CUTLER (2016): “The Association Between Income and Life Expectancy in the United States, 2001-2014,” *JAMA*, 315, 1750–1766.
- DAVIS, D., E. MENGUS, AND T. MICHALSKI (2019): “Labor Market Polarization and the Great Divergence: Theory and Evidence,” Working Paper 26955, National Bureau of Economic Research.
- DAVIS, D. R. AND J. I. DINGEL (2014): “The Comparative Advantage of Cities,” NBER Working Papers 20602, National Bureau of Economic Research, Inc.
- (2019): “A Spatial Knowledge Economy,” *American Economic Review*, 109, 153–170.
- DAVIS, S. J., J. HALTIWANGER, AND S. SCHUH (1996): “Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts,” *Small Business Economics*, 8, 297–315.
- DECKER, R. A., J. HALTIWANGER, R. S. JARMIN, AND J. MIRANDA (2016): “Declining Business Dynamism: What We Know and the Way Forward,” *American Economic Review*, 106, 203–207.
- DIAMOND, R. (2016): “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000,” *American Economic Review*, 106, 479–524.
- ECKERT, F. (2019): “Growing Apart: Tradable Services and the Fragmentation of the U.S. Economy,” Ph.D. thesis, Yale University.
- ECKHOUT, J., C. HEDTRICH, AND R. PINHEIRO (2019): “Automation, Spatial Sorting, and Job Polarization,” .
- ENGBOM, N. (2017): “Firm and Worker Dynamics in an Aging Labor Market,” Working Paper 756, Federal Reserve Bank of Minneapolis.
- FORT, T. C. (2017): “Technology and Production Fragmentation: Domestic versus Foreign Sourcing,” *Review of Economic Studies*, 84, 650–687.
- FORT, T. C. AND S. D. KLIMEK (2018): “The Effects of Industry Classification Changes on US Employment Composition,” Working Papers 18-28, Center for Economic Studies, U.S. Census Bureau.
- FORT, T. C., J. R. PIERCE, AND P. K. SCHOTT (2018): “New Perspectives on the Decline of US Manufacturing Employment,” *Journal of Economic Perspectives*, 32, 47–72.
- GANONG, P. AND D. SHOAG (2017): “Why has regional income convergence in the U.S. declined?” *Journal of Urban Economics*, 102, 76–90.
- GAUBERT, C. (2014): “Essays on Firms and Cities,” Ph.D. thesis, Princeton University.

- GIANNONE, E. (2017): “Skill-Biased Technical Change and Regional Convergence,” Ph.D. thesis, University of Chicago.
- HERKENHOFF, K. F., L. E. OHANIAN, AND E. C. PRESCOTT (2018): “Tarnishing the golden and empire states: Land-use restrictions and the U.S. economic slowdown,” *Journal of Monetary Economics*, 93, 89–109.
- HOPENHAYN, H. A. (1992): “Entry, Exit, and Firm Dynamics in Long Run Equilibrium,” *Econometrica*, 60, 1127–1150.
- HSIEH, C.-T. AND E. MORETTI (2019): “Housing Constraints and Spatial Misallocation,” *American Economic Journal: Macroeconomics*, 11, 1–39.
- HSIEH, C.-T. AND E. ROSSI-HANSBERG (2019): “The Industrial Revolution in Services,” Working Paper 25968, National Bureau of Economic Research.
- JIAO, Y. AND L. TIAN (2019): “Geographic Fragmentation in a Knowledge Economy,” .
- KARAHAN, F., B. PUGSLEY, AND A. ŞAHIN (2019): “Demographic Origins of the Startup Deficit,” Working Paper 25874, National Bureau of Economic Research.
- KATZ, L. F. AND K. M. MURPHY (1992): “Changes in Relative Wages, 1963–1987: Supply and Demand Factors,” *The Quarterly Journal of Economics*, 107, 35–78.
- KRUEGER, A. B. (1993): “How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984–1989,” *The Quarterly Journal of Economics*, 108, 33–60.
- KRUSELL, P., L. E. OHANIAN, J.-V. RÍOS-RULL, AND G. L. VIOLANTE (2000): “Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis,” *Econometrica*, 68, 1029–1054.
- LUTTMER, E. G. J. (2007): “Selection, Growth, and the Size Distribution of Firms,” *The Quarterly Journal of Economics*, 122, 1103–1144.
- MELITZ, M. J. (2003): “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71, 1695–1725.
- MICHAELS, G., F. RAUCH, AND S. J. REDDING (2013): “Task Specialization in U.S. Cities from 1880-2000,” NBER Working Papers 18715, National Bureau of Economic Research, Inc.
- MONTE, F., S. J. REDDING, AND E. ROSSI-HANSBERG (2018): “Commuting, Migration, and Local Employment Elasticities,” *American Economic Review*, 108, 3855–90.
- MORETTI, E. (2012): *The New Geography of Jobs*, Houghton Mifflin Harcourt.
- (2013): “Real Wage Inequality,” *American Economic Journal: Applied Economics*, 5, 65–103.



- NOCKE, V. (2006): “A Gap for Me: Entrepreneurs and Entry,” *Journal of the European Economic Association*, 4, 929–956.
- PUGSLEY, B. AND A. SAHIN (2018): “Grown-up business cycles,” *The Review of Financial Studies*, 32, 1102–1147.
- RAPPAPORT, J. (2018): “The Faster Growth of Larger, Less Crowded Locations,” *Kansas City Fed Economic Review*, 4th Quarter 2018.
- ROBACK, J. (1982): “Wages, Rents, and the Quality of Life,” *Journal of Political Economy*, 90, 1257–1278.
- ROSSI-HANSBERG, E., P.-D. SARTE, AND F. SCHWARTZMAN (2019): “Cognitive Hubs and Spatial Redistribution,” Working Paper 26267, National Bureau of Economic Research.
- RUGGLES, S., S. FLOOD, R. GOEKEN, J. GROVER, E. MEYER, J. PACAS, AND M. SOBEK (2019): “IPUMS USA Version 9.0 [dataset],” *Minneapolis, MN: IPUMS*.
- SAIZ, A. (2010): “The Geographic Determinants of Housing Supply,” *The Quarterly Journal of Economics*, 125, 1253–1296.
- SANTAMARIA, C. (2018): “Small Teams in Big Cities: Inequality, City Size and the Organization of Production,” Ph.D. thesis, Princeton University.
- WALSH, C. (2020): “Firm Creation and Local Growth,” Ph.D. thesis, Yale University.
- YEAPLE, S. R. (2005): “A simple model of firm heterogeneity, international trade, and wages,” *Journal of International Economics*, 65, 1 – 20.

## 9 Tables

|                       | $\log\left(\frac{\text{ICT investment}}{\text{employment}}\right)$ |                        | $\frac{\text{ICT investment}}{\text{total investment}}$ |                          |
|-----------------------|--|------------------------|---|--------------------------|
| $\log(\text{pop})$    | 0.0541***<br>(0.0109)  | 0.0568***<br>(0.00919) | 0.0142***<br>(0.00119)                                  | 0.0145***<br>(0.00117)   |
| $\log(\text{emp})$    |  | -0.430***<br>(0.00841) |   | -0.0156***<br>(0.00147)  |
| $\log(\text{estabs})$ |  | 0.450***<br>(0.0124)   |   | -0.00624***<br>(0.00214) |
| NAICS 4 FEs           | Y  | Y                      | Y   | Y                        |
| Year FEs              | Y  | Y                      | Y   | Y                        |
| Age FEs               | N  | Y                      | N   | Y                        |
| $N$                   | 239000   | 239000                 | 269000  | 269000                   |
| $R^2$                 | 0.194  | 0.324                  | 0.146   | 0.15                     |

SEs clustered at the CBSA level

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table I: Technology adoption and city-size

Source: ICT data is from ACES and merged into the LBD for information on age, size and location of the firm. Population is working age population (ages 20-64) from the Intercensal Population Estimates. Table displays the relationship between log of ICT spending and ICT share of investment and city-size. The unit of observation is a firm and city-size is measured at the CBSA level. Standard errors are clustered at the CBSA level. Note that the difference in sample size between columns 1-2 and columns 3-4 is due to firms with 0 ICT spending.

|                    | $\Delta$ 1980 to 2014 |                       |                      |                      |                      |                      |
|--------------------|-----------------------|-----------------------|----------------------|----------------------|----------------------|----------------------|
|                    | sr                    | er                    | esr                  | eer                  | jcr                  | jdr                  |
| $\log(pop_{1980})$ | 0.349***<br>(0.0557)  | 0.0979***<br>(0.0387) | 0.322***<br>(0.0586) | 0.286***<br>(0.0444) | 0.780***<br>(0.0822) | 0.308***<br>(0.0567) |
| Observations       | 57,500                | 57,500                | 57,500               | 57,500               | 58,500               | 58,500               |
| R2                 | 0.141                 | 0.066                 | 0.101                | 0.070                | 0.073                | 0.061                |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table II: Decline in Dynamism and City Size, industry controls

Source: LBD and author calculations. Population is from the Intercensal Population Estimates. Table displays the relationship between the demeaned decline in dynamism, as measured by the establishment start-up and exit rate, the firm start-up and exit rate and job creation and destruction, between 1980 and 2014 and city-size. Specifically for a city  $j$  and year  $t$ , demeaned decline in the dynamism rate  $D_{jt}$  is  $\Delta_{1980,2014}D_j - \Delta_{1980,2014}D_t$ . The unit of observation is a CBSA  $\times$  3-digit NAICS. Regression includes a full set of industry fixed effects. Standard errors are clustered at the CBSA level. Observations are rounded to the nearest 100 for disclosure avoidance.

| A. Aggregates   |  |     |                            |
|-----------------|--|-----|----------------------------|
| $H$             | mass of high skill                     | .24 | data                       |
| $L$             | mass of low skill                      | 1   | normalization              |
| B. Elasticities |  |     |                            |
| $1 - \beta$     | housing share of income                | .4  | Monte et al. (2018)        |
| $\sigma$        | elasticity of substitution $\omega$    | 6.8 | Broda and Weinstein (2006) |
| $\nu$           | scale parameter of Gumbel distribution | 3   | Allen et al. (2018)        |

Table III: Aggregate parameters

| parameters | description                    | value  | moment              | target | model |
|------------|--------------------------------|--------|---------------------|--------|-------|
| $\mu$      | persistence of Brownian motion | -.0013 | pareto tail emp     | 1.26   | 1.26  |
| $\Psi$     | sd of Brownian motion          | .020   | sd emp. gr          | 42%    | 43%   |
| $\epsilon$ | elasticity of sub. L & H       | 1.59   | Autor et al. (2008) | 1.62   | 1.62  |
| $f^b$      | fixed building cost for firms  | .99    | 3.5% of costs       | 3.5%   | 3.5%  |

Table IV: Firm productivity process parameters

| parameters                     | description                   | targets                   |
|--------------------------------|-------------------------------|---------------------------|
| $\gamma_j$ ,                   | skill premium                 | $\frac{w_{j,h}}{w_{1,l}}$ |
| $\psi_j$                       | city-size premium             | $\frac{w_{j,l}}{w_{1,l}}$ |
| $A_{hj}$                       | skill intensity               | $\frac{H_j}{L_j}$         |
| $A_{lj}$                       | relative city-size            | $\frac{H_j+L_j}{H_1+L_1}$ |
| $f_j^e$                        | establishment start-up rate   | $esr_j$                   |
| $f_j^c$                        | establishment size            | $\frac{H_j+L_j}{M_j}$     |
| $b_j$                          | rent of city $j$              | $\frac{r_j}{w_{1,l}}$     |
| $\frac{1}{\eta_j-1}$           | elasticity of building supply | Saiz, 2010                |
| Normalizations                 |                               |                           |
| $\gamma_{1,l}, A_{h1}, A_{l1}$ |                               | 1                         |

Table V: City level fundamentals

| parameters      | description                              | parameter value | targets                                | data | model |
|-----------------|--|-----------------|--|------|-------|
| $\Gamma_\gamma$ | new high skill share                     | 1.67            | Aggregate growth in high-skilled wages | 39.9 | 39.23 |
| $\Gamma_\psi$   | productivity advantage of new technology | 1.42            | Aggregate growth in low-skilled wages  | 4.31 | -0.13 |
| $\Gamma_{fc}$   | additional fixed cost                    | 2.15            | Average establishment size             | 18.5 | 19.63 |
| $H$             | Aggregate quantity of high-skilled labor | .694            | growth in supply of high-skilled labor | .694 | .694  |
| $L$             | Aggregate quantity of low-skilled labor  | 1.24            | growth in supply of low-skilled labor  | 1.24 | 1.24  |

Table VI: Calibration of the new technology

|                               | Data: $\log(\frac{ict}{emp})$ |                     |                        |                     | Model: share of adopters |                     |                     |                      |
|-------------------------------|-------------------------------|---------------------|------------------------|---------------------|--------------------------|---------------------|---------------------|----------------------|
| $\log(pop_{1980})$            | 0.0565***<br>(0.00929)        |                     | 0.0436***<br>(0.00955) |                     | 1.68***<br>(0.166)       |                     | 0.961***<br>(0.095) |                      |
| $\frac{H_{1980}}{L_{1980}}$   |                               | 0.970***<br>(0.212) |                        | 0.462***<br>(0.142) |                          | 52.66***<br>(3.762) |                     | 32.62***<br>(2.685)  |
| $\frac{w_{h1980}}{w_{l1980}}$ |                               |                     | 0.279**<br>(0.115)     | 0.116*<br>(0.0668)  |                          |                     | 3.35<br>(8.453)     | -8.830***<br>(1.453) |
| Observations                  | 239000                        | 239000              | 239000                 | 239000              | 30                       | 30                  | 30                  | 30                   |
| $R^2$                         | 0.324                         | 0.320               | 0.322                  | 0.324               | 0.785                    | 0.875               | 0.006               | 0.976                |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table VII: Model vs. Data: ICT spending and adoption vs. initial city characteristics

Source: The left panel uses ICT data from ACES merged into the LBD for information on the location of the firm. Population is working age (ages 20-64) population from the Intercensal Population Estimates. Table displays the relationship between log of ICT spending from 2003 to 2013 and 1980 city characteristics including city-size, skill intensity and the skill premium. The regressions include a full set of 4 digit NAICS fixed effects and year fixed effects. The unit of observation is a firm and city-size is measured at the CBSA level. Standard errors are clustered at the CBSA level. The right panel uses model generated output and shows the relationship between the share of firms that adopt in a city and 1980 city characteristics.

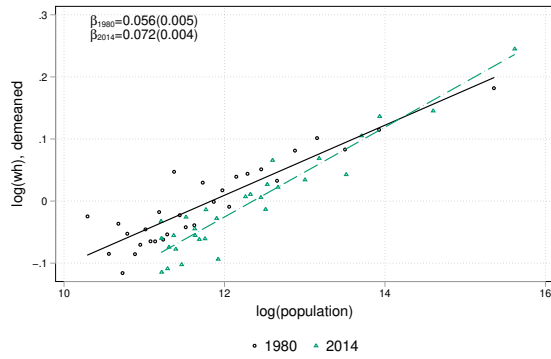


|              | $\Delta$ adoption | High-skilled |              | Low-skilled |              |
|--------------|-------------------|--------------|--------------|-------------|--------------|
|              |                   | $w_h$ gr.    | cons. equiv. | $w_l$ gr.   | cons. equiv. |
| no migration | -0.014            | 1.461        | -0.290       | 0.250       | -1.554       |
| equal HSE    | -0.305            | 2.211        | 8.627        | 0.699       | 6.104        |

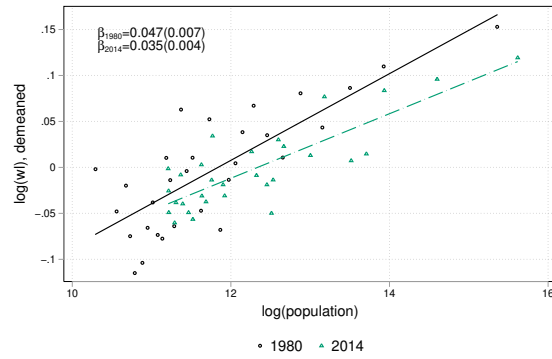
*Source:* Model calculations. Table shows the aggregate share of firms adopting, average wages and average consumption equivalent measures. Aggregates are averages across cities weighted by the number of firms, high-skilled population and low-skilled population, respectively.

Table VIII: Counterfactuals: aggregates

## 10 Figures



(a) High skilled wages



(b) Low skilled wages

Figure I: City-size wage premium by skill group

Source: Wages are from the 1980 Decennial Census and 2014 ACS. Population is working age population (ages 20-64) from the Intercensal Population Estimates. Figure displays the relationship between log average wages and city-size. The unit of observation is one of 30 city-size categories. Wages are demeaned by year. Specifically, for each skill type  $\tau$  and each year  $t$  I plot are  $\log(w_{\tau t}) - \log(\bar{w}_{\tau t})$ .

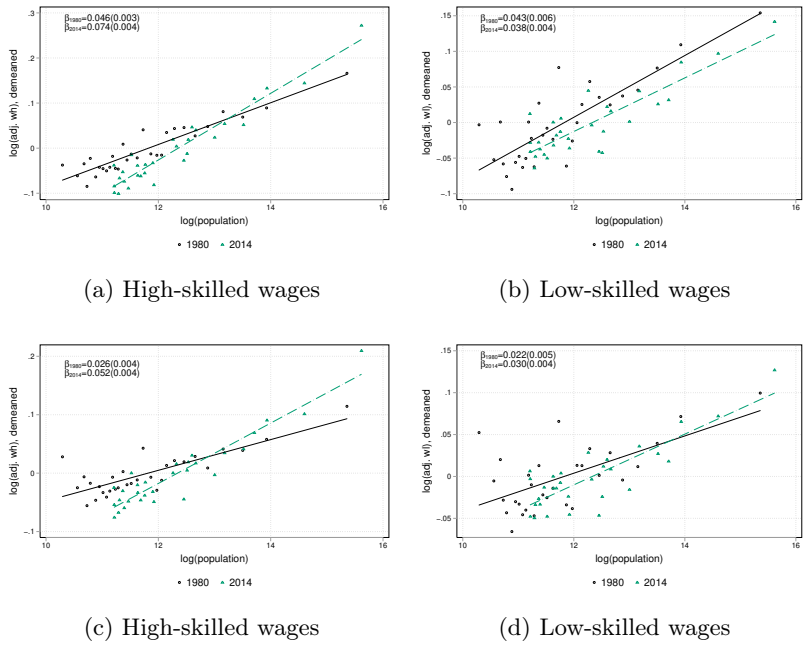


Figure II: City-size wage premium by skill group, residual wages

Source: Wages are from the 1980 Decennial Census and 2014 ACS. Population is working age population (ages 20-64) from the Intercensal Population Estimates. Figure displays the relationship between log average adjusted wages and city-size. The unit of observation is one of 30 city-size categories. Wages are demeaned by year. Specifically, for each skill type  $\tau$ , city  $j$  and each year  $t$ , I plot are  $\log(w_{\tau jt}) - \log(\bar{w}_{\tau t})$ . Panels (a) and (b) show wages adjusted for demographics including age, age squared, gender and race dummies. Panels (c) and (d) add further controls for occupation and industry fixed effects.

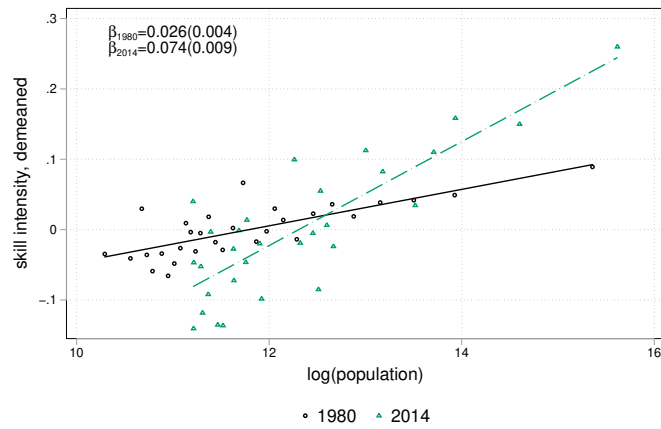
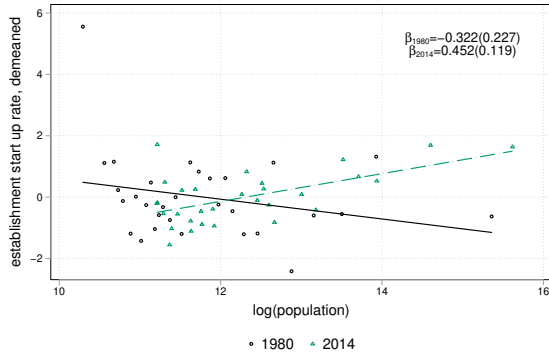
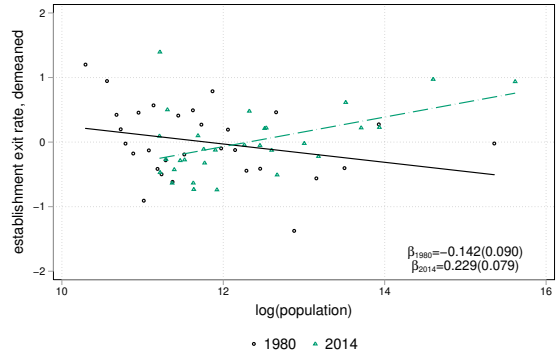


Figure III: Skill intensity and city-size

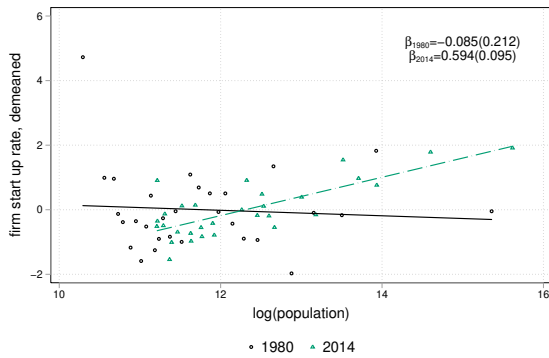
Source: Skill intensity is from the 1980 Decennial Census and 2014 ACS. Population is working age population (ages 20-64) from the Intercensal Population Estimates. Figure displays the relationship between demeaned skill intensity and city-size. Specifically for a city  $j$  and year  $t$  demeaned skill intensity is  $\frac{H}{L}_{jt} - \bar{\frac{H}{L}}_t$ . The unit of observation is one of 30 city-size categories.



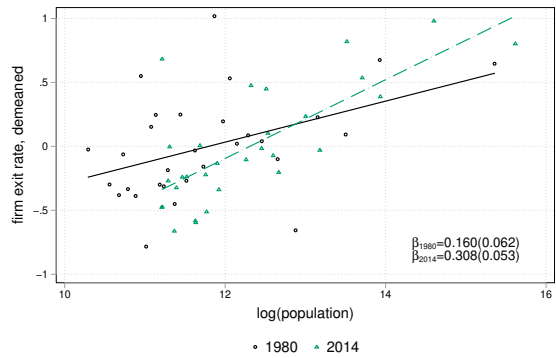
(a) Establishment start-up rate



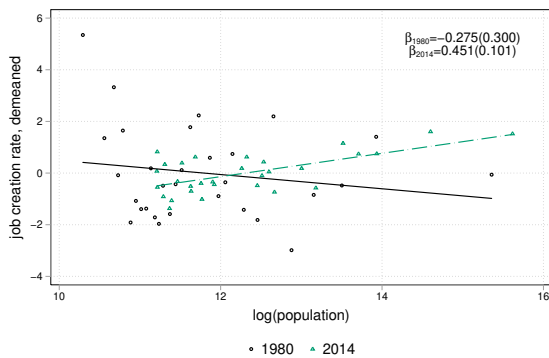
(b) Establishment exit rate



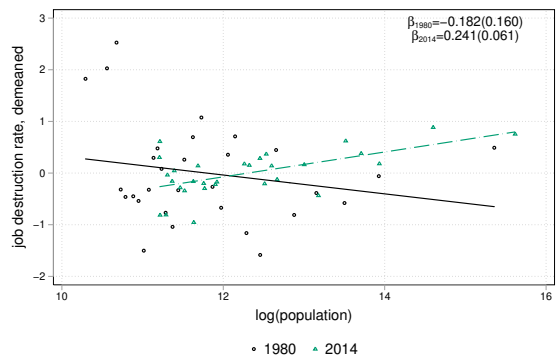
(c) Firm start-up rate



(d) Firm exit rate



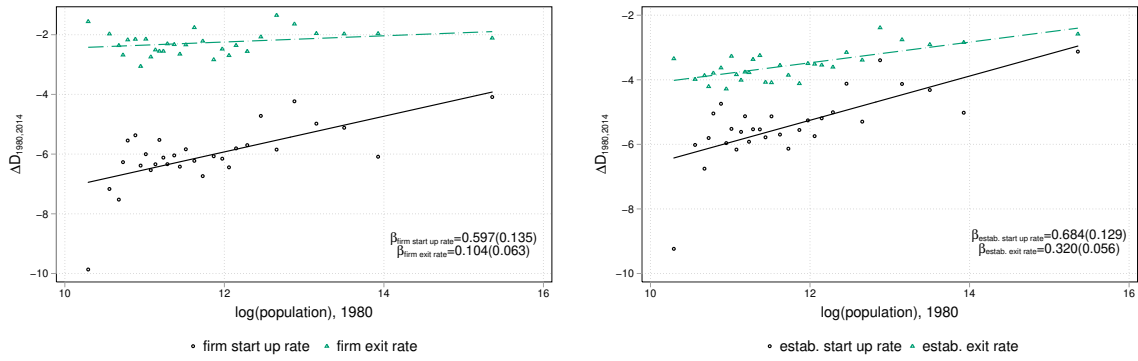
(e) Job creation rate



(f) Job destruction rate

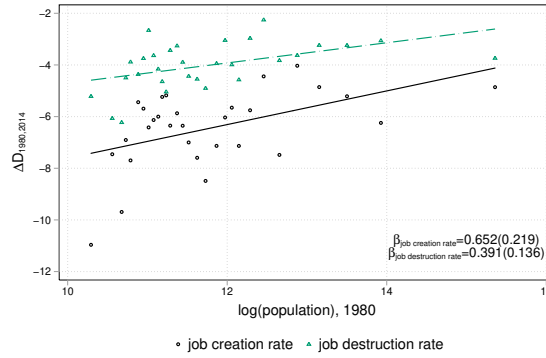
## Figure IV: Dynamism and City-size

Source: Business Dynamic Statistics and author calculations. Population is working age population (ages 20-64) from the Intercensal Population Estimates. Figure displays the relationship between demeaned dynamism, as measured by the establishment start-up and exit rate, the firm start-up and exit rate and job creation and destruction, and city-size. Specifically for a city  $j$  and year  $t$  demeaned dynamism rate  $D_{jt}$  is  $D_{jt} - \bar{D}_t$ . The unit of observation is one of 30 city-size categories.



(a) Firm start-up and exit rate

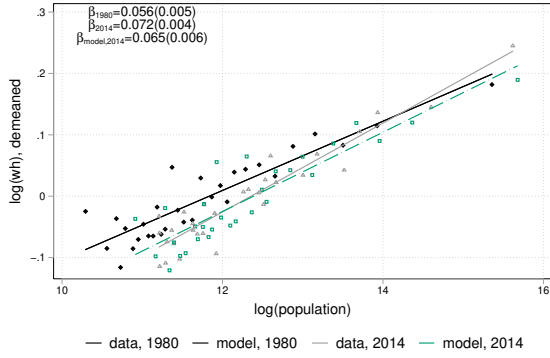
(b) Establishment start-up and exit rate



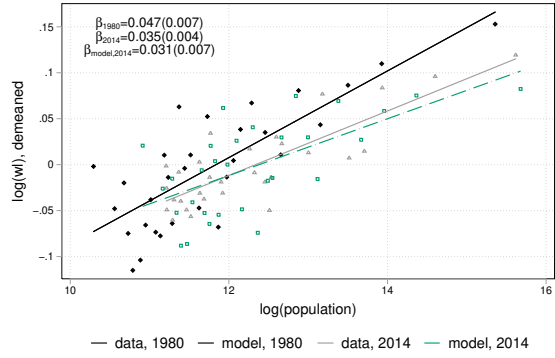
(c) Job creation and job destruction rate

Figure V: Decline in dynamism versus initial city-size

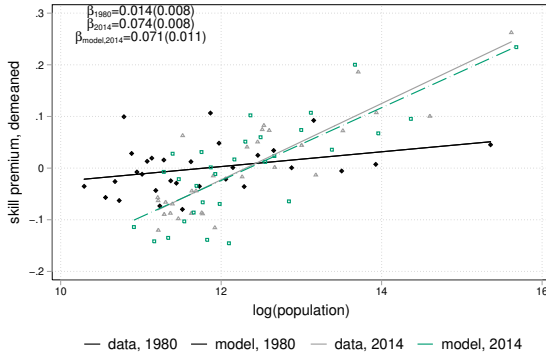
Source: Business Dynamic Statistics and author calculations. Population is from the Intercensal Population Estimates. Figure displays the relationship between the demeaned decline in dynamism, as measured by the establishment start-up and exit rate, the firm start-up and exit rate and job creation and destruction, between 1980 and 2014 and city-size. Specifically for a city  $j$  and year  $t$  demeaned dynamism rate  $D_{jt}$  is  $\Delta_{1980,2014} D_j - \Delta_{1980,2014} D_t$ . The unit of observation is one of 30 city-size categories.



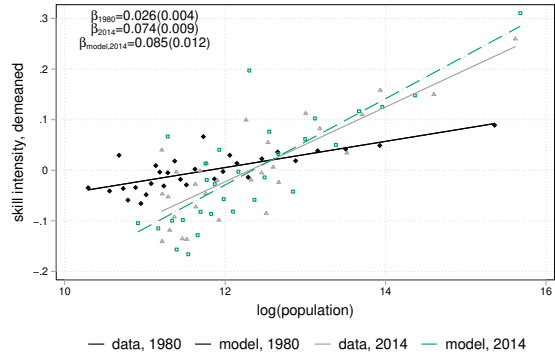
(a) City-size wage premium: high-skilled



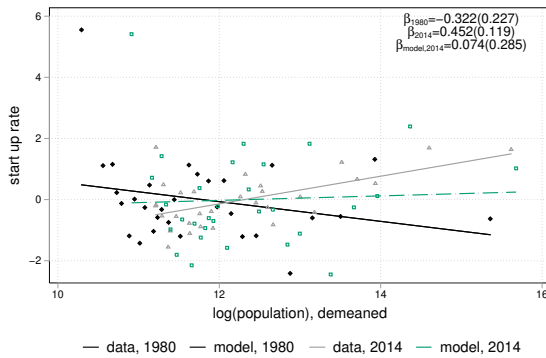
(b) City-size wage premium: low-skilled



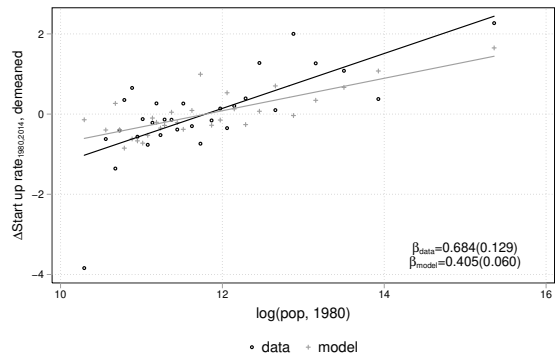
(c) Skill premium



(d) Skill intensity



(e) Establishment start-up rate



(f) Change in the establishment start-up rate

## Figure VI: Descriptive facts: data vs. model

Source: Wages and skill intensity are from the 1980 Decennial Census and 2014 ACS. Dynamism is from the Business Dynamic Statistic. Population is working age population (ages 20-64) from the Intercensal Population Estimates. Additional data from author calculated model output. Figure displays the relationship between wages, skill intensity, and dynamism and city-size in the model and the data. By construction, the model and data align perfectly in 1980 and are both shown in black. Gray gives the 2014 data and green the 2014 steady state in the model. The unit of observation is one of 30 city-size categories. All variables are demeaned by year.

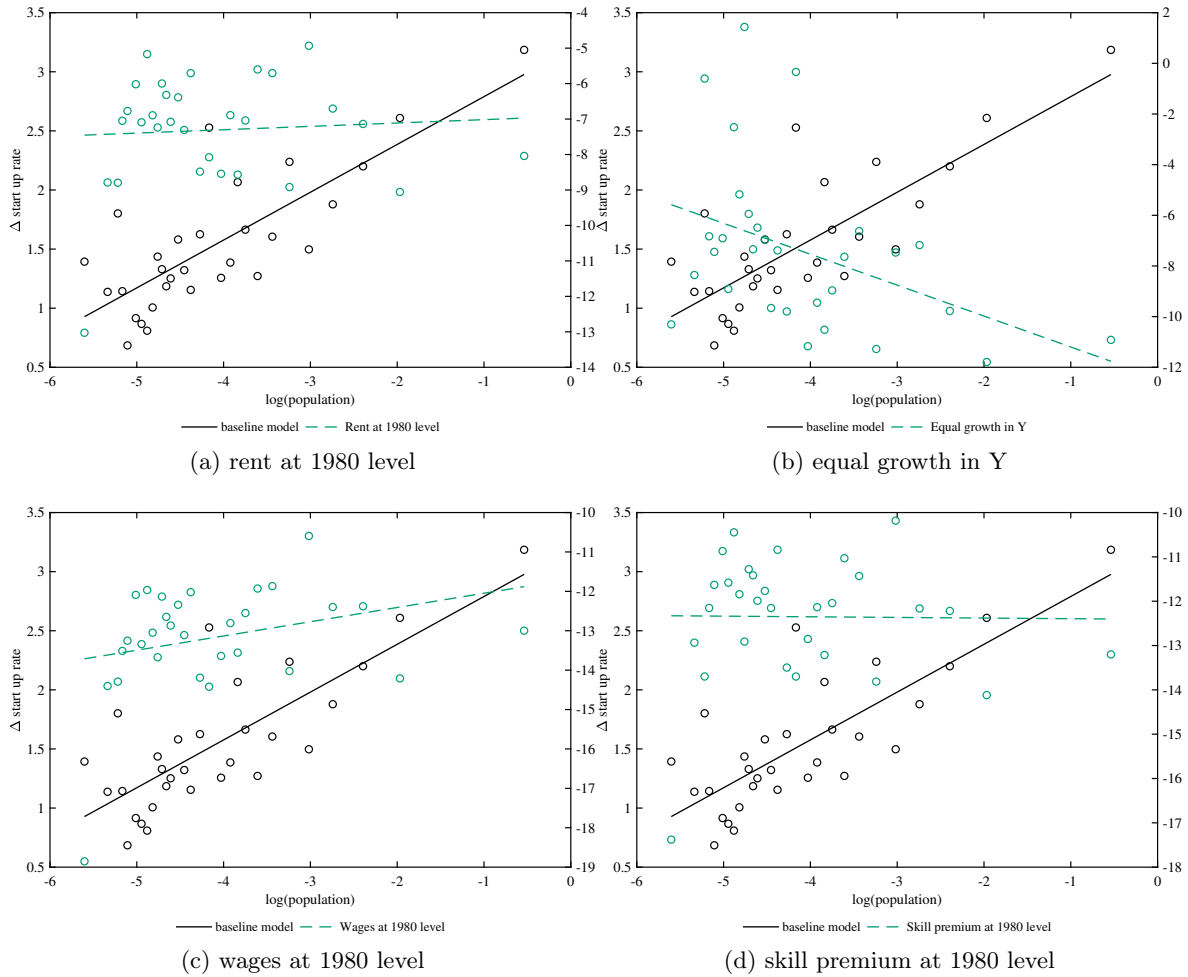


Figure VII: Drivers of the cross sectional change in dynamism

Source: Model calculations of counterfactuals cross-sectional patterns of dynamism and city-size if different variables were held constant.



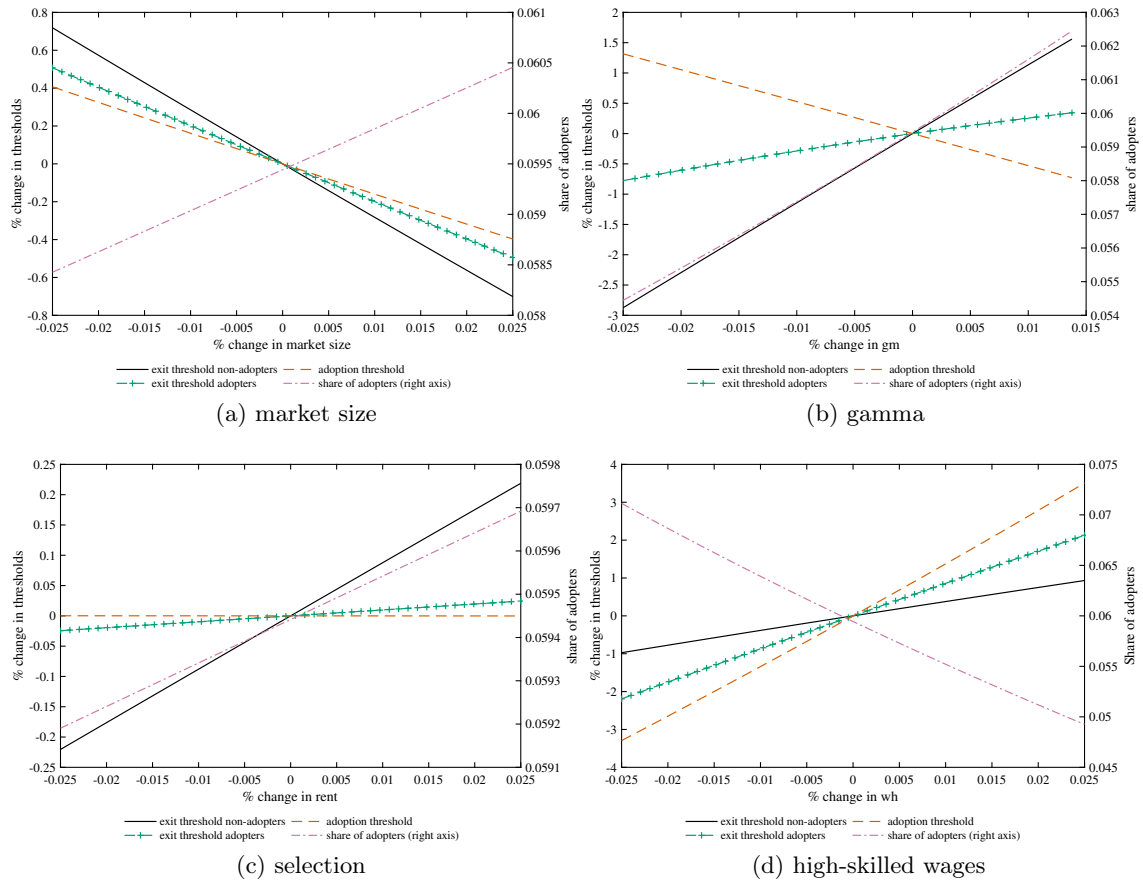
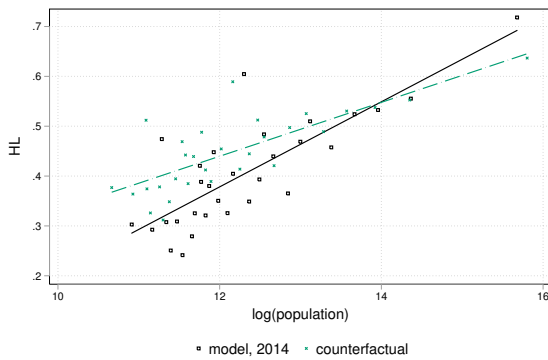
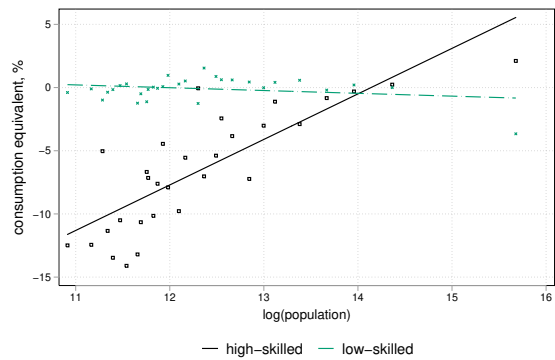


Figure VIII: Drivers of adoption

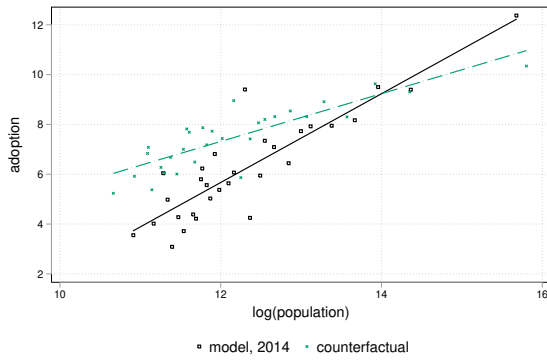
Source: Model calculations of the exit threshold for non-adopters, the exit threshold for adopters, the adoption threshold and the share of adopters versus model objects such as wages, rents and the initial weight on high-skilled labor  $\gamma_j$ . These relationships are shown for city-size category 20.



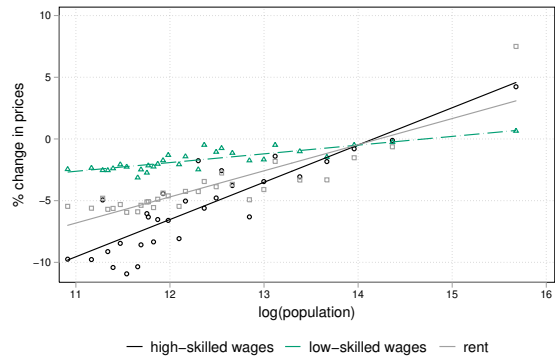
(a) Skill intensity



(b) consumption equivalence



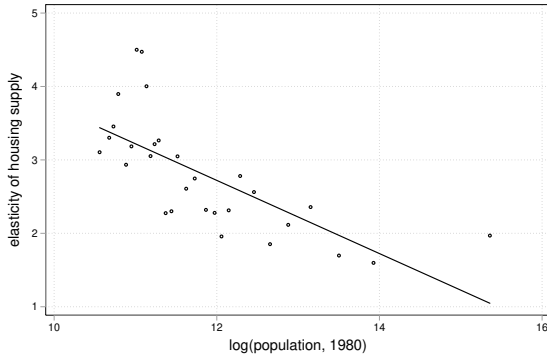
(c) adoption



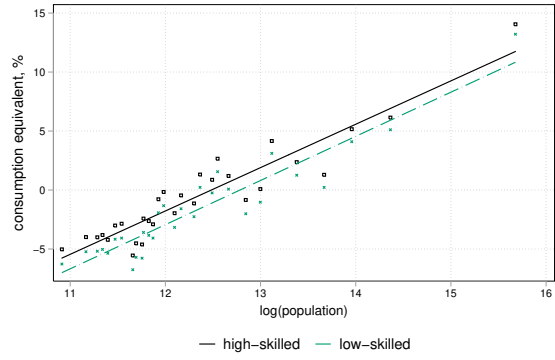
(d) high-skilled wages

## Figure IX: No Migration

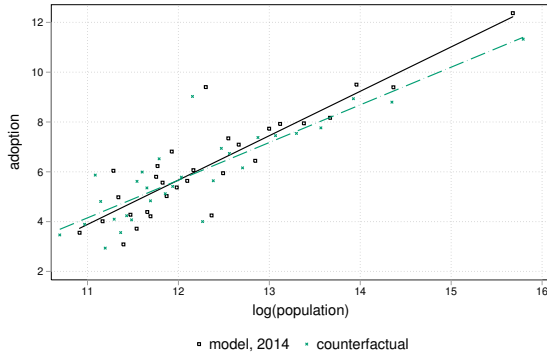
Source: Model generated output. Panel (a) shows what skill intensity ( $\frac{H}{L}$ ) looks like in the no migration counterfactual. Panel (b) shows the consumption equivalent for high- and low-skilled labor versus 2014 city-size. The consumption equivalent gives the percent of consumption they would be willing to give up in order to avoid the counterfactual world. Panel (c) shows the relationship between adoption and city-size. Panel (d) shows the relationship between the percent change in wages and rents and city-size between the no-migration counterfactual and the 2014 steady state.



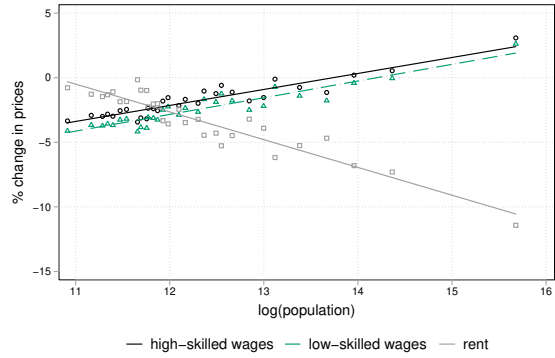
(a) Housing Supply Elasticity, [Saiz \(2010\)](#)



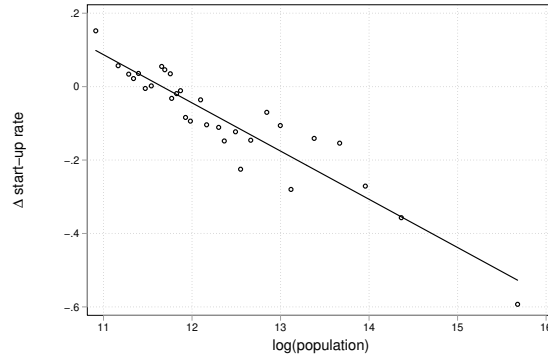
(b) consumption equivalence



(c) adoption



(d) change in prices



(e) dynamism

## Figure X: Equalize Elasticity of Housing Supply

Source: Data on housing supply elasticity is from [Saiz \(2010\)](#). Population is working age (ages 20-64) population from the Intercensal Population Estimates. Model generated output. Panel (a) shows the relationship between housing supply elasticity and 1980 city-size. Panel (b) shows the consumption equivalent for high- and low-skilled labor versus 2014 city-size. The consumption equivalent gives the percent of consumption they would be willing to give up in order to avoid the counterfactual world. Panel (c) shows the relationship between adoption and city-size and panel (d) shows the relationship of the change in wages and rent and city-size between the counterfactual and the 2014 steady state. Panel (e) shows the change in dynamism between the counterfactual and the 2014 steady state.

## Appendix

### A Data Appendix

This appendix provides robustness checks for the main empirical findings in the paper. First, I present robustness for the main demographic facts in Section A.1. Then in Section A.2 I provide robustness for the facts on business dynamism. In Section A.3, I provide more details on the ACES ICT data.

#### A.1 Descriptive facts robustness checks: Demographic facts

Figure A.I shows the city-size wage premium by year for all cities rather than the city size bins used in the main text. Panels a and b show the raw city size wage premium for high- and low-skilled workers, respectively. While Panels c and d use residual wages from a standard Mincer regression controlling for age, age squared, sex, race and foreign born status. Panels e and f add controls for industry and occupation. In each case, using the full set of cities instead of the city size bins does not change the result. With each specification the city-size wage premium for high-skilled workers increases between 1980 and 2014. For low-skilled workers it decreases, with the exception of the specification including industry and occupation controls for which it remains constant. This reflects the fact that low-skilled workers are reallocating towards industries and occupations that pay less in big cities more than in small cities. This would be consistent with a model in which there is less technology adoption in the small cities which implies less reallocation across industries and occupations. Even though the city-size wage premium is flat for low-skilled workers when controlling for industry and occupation, the divergence of the city-size wage premium between high- and low-skilled workers holds.

In Table A.I, I show the city size wage premium for more education categories: less than high school, high school, some college, college and more than college. Panel (a) shows the results using average wage by group while Panel (B) shows the results using residual wages from a standard Mincer regression. Panel (C) shows the results further controlling for a full set of industry and occupation fixed effects. The largest increase in the city-size wage premium is for the more than college group, while for less than high school, the city-size wage premium falls. The city size wage premium for college is flat when using average wages, but rising when considering the wages adjusted for demographics and industry and occupation. The high school city-size wage premium is roughly stable, while for workers with some college, there is a slight increase in the city size wage premium between 1980 and 2014.

Figure A.II shows the relationship between skill intensity and city-size for 1980 and 2014, for all cities rather than the city size bins used in the main text. Figure A.II also includes three definitions of skill intensity. First, in Panel (a) is the baseline measure of skill intensity, those with 4 or more years of college divided by those with some college or less. Panel (b) uses an alternative definition: those with some college or more divided by those with high school or less, and finally, Panel (c) uses

those with exactly 4 years of college relative to those with only a high school degree. In all three cases, the same pattern holds: the relationship between skill intensity and city-size has increased between 1980 and 2014.

## A.2 Descriptive facts robustness checks: Firm facts

In Figure A.III, I present the correlation between dynamism and city-size over time controlling for industry composition at the 3 digit NAICS level. To do this, I compute dynamism measures within a city-industry cell and then estimate the following specification

$$D_{ict} = \alpha_{it} + \sum_t \beta_t \log(pop_{ct}) \times \lambda_t + \epsilon_{ict}.$$

Figure A.III shows the correlation coefficients  $\beta_t$  over time. Panel (b) of Figure A.III shows the same correlation coefficients  $\beta_t$  over time, but with further controls for the lagged 10 year population growth in the city and the prime age worker share. These robustness checks are motivated by the work of Karahan et al. (2019) and Engbom (2017) who identify population growth and prime age workers as important drivers of business dynamics. The same conclusions still hold - the cross sectional correlation between dynamism and city size has increased between 1980 and 2014.

Because city-industry cells can have very small firm counts for small cities, I perform an alternative analysis by grouping cities into 7 size categories<sup>21</sup>. I then calculate the dynamism measures within a city-size category-industry bin, using 4-digit NAICS codes. Grouping cities into size categories allows me to control for a more detailed industry composition without worrying about small firm counts. In Figure A.IV, I plot the city-size category fixed effects from a regression of the dynamism measure on a full set of industry-year fixed effects and city-size-year fixed effects:

$$D_{ict} = \alpha_{it} + \gamma_{ct} + \epsilon_{ict}$$

The city-size category fixed effects represent the average difference, across industries, in the dynamism rates between big cities and the smallest city-size category.

In Figure A.V, I show that the results for the relationship between the change in dynamism between 1980 and 2014 and city size are robust to using a finer set of industry controls. I again group cities into size categories and calculate dynamism rates within a city-size - NAICS 4 bin. I then estimate the following specification

$$\Delta_{2014,1980} D_{ic} = \alpha_i + \gamma_c + \epsilon_{ic}$$

Figure A.V plots the city fixed effects from this regression. The city fixed effect represents the

---

<sup>21</sup>Specifically, city-size categories are based on population percentiles. City-size category 1 is cities below the 50th percentile, category 2 is the 50th to 75th percentile, 3 is the 75th to 90th percentile, 4 is the 90th to 95th percentile, 5 is 95th percentile to 98th percentile, 6 is the 98th to 99th percentile and finally category 7 is cities above the 99th percentile.

average difference across industries between the decline in dynamism for a given city-size category and the smallest city-size category. The fact that the line is increasing in city-size means that the bigger city-size categories experienced a smaller decline in dynamism than the smallest cities.

### A.3 Annualized Capital Expenditure Survey Data Appendix

In this section, I provide some information about matching the Annualized Capital Expenditure Survey (ACES) and the ICT supplement to the Longitudinal Business Database (LBD). Of all the observations in the ICT supplement of the ACES, 95.5% of them match to a firm in the LBD. Since the ACES over-samples large firms, it is more relevant that 92.5% of the total sampling weights have a match in the LBD. In Table A.II, I show some summary statistics on the firms that match versus the firms that do not match. While both matchers and non-matchers have similar ICT shares of investment, non-matchers are smaller and thus have lower levels of ICT investment and total investment.

Note that the ACES survey is conducted at the level of EIN (Employer Identification Number) while the LBD aggregates EINs that belong to one firm under a unique FIRMID. Thus, it is possible that multiple observations in the ACES match to the same FIRMID in the LBD. In this case, I aggregate the observations in the ACES, summing across the different investment categories to get total firm investment. To match to the LBD, I use the ALPHA number provided in the ACES when available. When ALPHA is not available, I use the BRID (Business Register ID). I verify the matches by checking that the broad industry classification (NAICS Sector) is the same in both datasets.

## B Derivations for firm problems

### B.1 Value function: no adoption

To find the solution to the HJB equation I use the following theorem:

**Theorem:** The general solution is the sum of a particular solution to the general equation and the general solution to the reduced equation.

So, one needs to find a particular solution to the general equation and a general solution to the reduced equation. The reduced equation is

$$0 = -\rho V(s) + \tilde{\mu} V'(s) + \frac{1}{2} \tilde{\Psi}^2 V''(s).$$

To find a general solution to the reduced equation, I use the following theorem:

**Theorem:** the generalized solution to the reduced equation is a linear combination of any two linearly independent particular solutions.

To find particular solutions, consider the class of equations  $e^{\xi s}$ . Plugging this into the reduced equation

$$0 = -\rho e^{\xi s} + \tilde{\mu} \xi e^{\xi s} + \frac{1}{2} \tilde{\Psi}^2 \xi^2 e^{\xi s}$$

dividing by  $e^{\xi s}$  gives

$$\begin{aligned} 0 &= -\rho + \tilde{\mu} \xi + \frac{1}{2} \tilde{\Psi}^2 \xi^2 \\ \implies \xi &= -\tilde{\mu} \pm \frac{\sqrt{\tilde{\mu}^2 + 2\tilde{\Psi}^2 \rho}}{\tilde{\Psi}^2}. \end{aligned}$$

Let  $\xi^+$  and  $\xi^-$  be the positive and negative roots, respectively. Then the general solution to the reduced equation is

$$C^+ e^{\xi^+ s} + C^- e^{\xi^- s}.$$

The general solution is the sum of a particular solution to the general equation and a general solution to the reduced equation. I just found the general solution to the reduced equation so now I find a particular solution to the general equation. Define this as  $V_p(m)$ . I solve for  $V_p(m)$  using the method of undetermined coefficients. I look for a solution to

$$\rho V(s) = [Z e^s - f^c - r f^b] + \tilde{\mu} V'(s) + \frac{1}{2} \tilde{\Psi}^2 V''(s)$$

guessing that the answer is  $V(s) = A e^s + B$  and plugging the guess into the HJB equation

$$\rho (A e^s + B) = [Z e^s - f^c - r f^b] + \tilde{\mu} A e^s + \frac{1}{2} \tilde{\Psi}^2 A e^s.$$

Rearranging gives

$$\begin{aligned} A e^s + B &= \frac{1}{\rho} [-f^c - r f^b] + \frac{1}{\rho} \left[ Z + \tilde{\mu} A + \frac{1}{2} \tilde{\Psi}^2 A \right] e^s \\ \implies B &= \frac{1}{\rho} [-f^c - r f^b] \end{aligned}$$

and

$$\begin{aligned} \implies A &= \frac{1}{\rho} \left[ Z + \tilde{\mu} A + \frac{1}{2} \tilde{\Psi}^2 A \right] \\ \implies A &= \frac{Z}{\rho - \tilde{\mu} - \frac{1}{2} \tilde{\Psi}^2}. \end{aligned}$$

So the particular solution to the HJB is

$$V_p(s) = \frac{Z}{\rho - \tilde{\mu} - \frac{1}{2} \tilde{\Psi}^2} e^s + \frac{1}{\rho} [-f^c - r f^b].$$

Thus a general solution to the value function will be

$$V(s) = V_p(s) + C^+ e^{\xi^+ s} + C^- e^{\xi^- s}$$

where

$$V_p(s) = \frac{Z}{\rho - \tilde{\mu} - \frac{1}{2}\tilde{\Psi}^2} e^s + \frac{1}{\rho} [-f^c - r f^b].$$

Using the no bubble condition, it must be the case that  $C^+ = 0$  otherwise  $V(s)$  will diverge from  $V_p(s)$  as  $s \rightarrow \infty$ . Then I solve for  $C^-$  using the value matching condition and the optimal threshold using the smooth pasting condition.

Value matching gives

$$\begin{aligned} V(s_x) &= 0 \\ V(s_x) &= \frac{Z}{\rho - \tilde{\mu} - \frac{1}{2}\tilde{\Psi}^2} e^{s_x} + \frac{1}{\rho} [- (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f^c - r f^b] + C^- e^{\xi^- s_x} = 0 \\ \implies C^- &= \frac{-Z}{\rho - \tilde{\mu} - \frac{1}{2}\tilde{\Psi}^2} e^{s_x - \xi^- s_x} - \frac{1}{\rho} [- (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f^c - r f^b] e^{-\xi^- s_x}. \end{aligned}$$

So now the value function is

$$\begin{aligned} V(s) &= \frac{Z}{\rho - \tilde{\mu} - \frac{1}{2}\tilde{\Psi}^2} e^s + \frac{1}{\rho} [- (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f^c - r f^b] \dots \\ &+ \left[ \frac{-Z}{\rho - \tilde{\mu} - \frac{1}{2}\tilde{\Psi}^2} e^{s_x - \xi^- s_x} - \frac{1}{\rho} [- (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f^c - r f^b] e^{-\xi^- s_x} \right] e^{\xi^- s}. \end{aligned}$$

Then using the smooth pasting condition, I solve for  $s_x$

$$\begin{aligned} V'(s_x) &= 0 \\ V'(s_x) &= \frac{Z}{\rho - \tilde{\mu} - \frac{1}{2}\tilde{\Psi}^2} e^{s_x} \dots \\ &+ \xi^- \left[ \frac{-Z}{\rho - \tilde{\mu} - \frac{1}{2}\tilde{\Psi}^2} e^{s_x - \xi^- s_x} - \frac{1}{\rho} [- (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f^c - r f^b] e^{-\xi^- s_x} \right] e^{\xi^- s_x} = 0. \end{aligned}$$

Solving for  $s_x$  gives

$$s_x = \ln \left[ \frac{1}{Z} \left( \frac{-\xi^-}{1 - \xi^-} \right) \left( \frac{\rho - \tilde{\mu} - \frac{1}{2}\tilde{\psi}^2}{\rho} \right) \left( (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f^c + r f^b \right) \right].$$

## B.2 Firm size distribution: no adoption

The corresponding Kolmogorov Forward Equation is

$$\frac{\partial g(s)}{\partial t} = -\tilde{\mu} \frac{dg(s)}{ds} + \frac{1}{2} \tilde{\psi}^2 \frac{d^2 g(s)}{ds^2}$$



In steady state  $\frac{\partial g(s)}{\partial t} = 0 \implies -\delta \frac{dg(s)}{ds} = \frac{d^2 g(s)}{ds^2}$  where  $\delta = \frac{-\tilde{\mu}}{\tilde{\Psi}^2/2}$ . The solution is going to be

$$g(s) = \begin{cases} g^-(s) & s \in (s_x, s_e) \\ g^+(s) & s > s_e \end{cases}$$

and

$$\begin{aligned} g^-(s) &= C_1 + C_2 e^{-\delta s} \\ g^+(s) &= C_3 + C_4 e^{-\delta s} \end{aligned}$$

There will be 4 boundary conditions:

1. the mass of firms is finite:  $M = \int_{s_x}^{s_e} g^-(s) ds + \int_{s_e}^{\infty} g^+(s) ds < \infty$

(a)  $\int_{s_e}^{\infty} g^+(s) ds = [C_3 s + C_4 \frac{1}{\delta} e^{-\delta s}]_{s_e}^{\infty} \implies C_3 = 0$

2.  $f(s)$  is continuous at entry:  $\lim_{s \downarrow s_e} g^+(s) = \lim_{s \uparrow s_e} g^-(s)$

(a) To show this use the discrete approximation to the Brownian motion

$$\Delta N = (1-p)hg^+(s_e+h) = phg^-(s_e-h)$$

$$\Delta N \approx (1-p)h(g^+(s_e) + g^{+'}(s_e)h + g^{+''}(s_e)h^2)$$

$$\approx ph(g^-(s_e) + g^{-'}(s_e)(-h) + g^{-''}(s_e)h^2)$$

Divide both sides by  $\sqrt{\Delta}$  and take the limit as  $\Delta \rightarrow 0$  which gives

$$g^+(s_e) = g^-(s_e)$$

(b)  $C_1 + C_2 e^{-\delta s_e} = C_4 e^{-\delta s_e} \implies C_1 = (C_4 - C_2) e^{-\delta s_e}$

3. there's no mass at the exit threshold:  $g(s_x) = 0$

(a)  $(C_4 - C_2) e^{-\delta s_e} + C_2 e^{-\delta s_x} = 0 \implies C_4 = C_2 \frac{(e^{-\delta s_e} - e^{-\delta s_x})}{e^{-\delta s_e}}$

(b) So far we have

$$g^-(s) = -C_2 e^{-\delta s_x} + C_2 e^{-\delta s}$$

$$g^+(s) = C_2 \frac{(e^{-\delta s_e} - e^{-\delta s_x})}{e^{-\delta s_e}} e^{-\delta s}$$

4. the mass of exitors is

$$\Delta E = (1-p)hg^-(s_x+h) \approx (1-p)h[g^-(s_x) + hg^{-'}(s_x) + O(h^2)]$$

dividing by  $\Delta$  and taking the limit as  $\Delta \rightarrow 0$  implies  $E = \frac{\tilde{\psi}^2}{2} g^{-'}(s_x)$

(a)  $f'(s_x) = -\delta C_2 e^{-\delta s_x} = \frac{2E}{\tilde{\psi}^2} \implies C_2 = \frac{E}{\tilde{\mu} e^{-\delta s_x}}$

### B.3 Value function: non-adopters

As before the solution  $v(s)$  will be the sum of a particular solution to the general equation and two linearly independent solutions to the reduced equation

$$V(s) = V_p(s) + C^+ e^{\xi^+ s} + C^- e^{\xi^- s}$$

where  $\xi = -\tilde{\mu} \pm \frac{\sqrt{\tilde{\mu}^2 + 2\tilde{\Psi}^2 \rho}}{\tilde{\Psi}^2}$  and  $V_p(s)$  is the same as above and solved for using the method of undetermined coefficients:

$$V_p(s) = \frac{Z}{\rho - \tilde{\mu} - \frac{1}{2}\tilde{\Psi}^2} e^s + \frac{1}{\rho} [- (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f^c - r f^b]$$

To make the notation easier, rewrite this as

$$V_p(s) = Ae^s + B$$

I use the boundary conditions to solve for  $C^+$ ,  $C^-$  and the exit and adoption thresholds.

1. Use  $v(s_x) = 0$  to solve for  $C^+$

$$C^+ = \frac{-Ae^{s_x} - B - C^- e^{\xi^- s_x}}{e^{\xi^+ s_x}}$$

2. Use  $v(s_a) = v^{a*}$  to solve for  $C^-$

$$v^{a*} = Ae^{s_a} + B + \left[ \frac{-Ae^{s_x} - B - C^- e^{\xi^- s_x}}{e^{\xi^+ s_x}} \right] e^{\xi^+ s_a} + C^- e^{\xi^- s_a}$$

$$C^- \left( e^{\xi^- s_x + \xi^+ s_a - \xi^+ s_x} - e^{\xi^- s_a} \right) = Ae^{s_a} + B + \left( \frac{-Ae^{s_x} - B}{e^{\xi^+ s_x}} \right) e^{\xi^+ s_a} - v^{a*}$$

$$C^- = \frac{Ae^{s_a} + B + (-Ae^{s_x} - B) e^{\xi^+ s_a - \xi^+ s_x} - v^{a*}}{e^{\xi^- s_x + \xi^+ s_a - \xi^+ s_x} - e^{\xi^- s_a}}$$

3. Plug in  $C^-$  to solve for  $C^+$

$$C^+ = \frac{-Ae^{s_x} - B - C^- e^{\xi^- s_x}}{e^{\xi^+ s_x}}$$

$$C^+ = \left( \frac{-Ae^{s_x} - B}{e^{\xi^+ s_x}} \right) - \left( \frac{Ae^{s_a} + B + (-Ae^{s_x} - B) e^{\xi^+ s_a - \xi^+ s_x} - v^{a*}}{e^{\xi^- s_x + \xi^+ s_a - \xi^+ s_x} - e^{\xi^- s_a}} \right) e^{s_x(\xi^- - \xi^+)}$$

4. Use the smooth pasting conditions  $v'(s_x) = 0$  and  $v'(s_a) = v^{a'}$  to solve for  $s_x$  and  $s_a$

## B.4 Firm size distribution: with adoption

### The distribution conditional on adoption

The stochastic process for firm size is  $\frac{ds}{dt} = \tilde{\mu}dt + \tilde{\Psi}dW(t)$  and the same Kolmogorov Forward Equation as in the case with no adoption

$$\frac{\partial g^a(s)}{\partial t} = -\tilde{\mu}\frac{dg^a(s)}{ds} + \frac{1}{2}\tilde{\Psi}^2\frac{d^2g^a(s)}{ds^2}$$

In steady state  $\frac{\partial g^a(s)}{\partial t} = 0$  and  $\delta = \frac{-\tilde{\mu}}{\tilde{\Psi}^2/2}$ . The solution is going to be

$$g^a(s) = \begin{cases} g^{a-}(s) & s \in (s_x, s_a) \\ g^{a+}(s) & s > s_a \end{cases}$$

and

$$\begin{aligned} g^{a-}(s) &= C_1 + C_2e^{-\delta s} \\ g^{a+}(s) &= C_3 + C_4e^{-\delta s} \end{aligned}$$

There will be 4 boundary conditions:

1. the mass of firms is finite:  $M = \int_{s_x^a}^{s_a} g^{a-}(s)ds + \int_{s_a}^{\infty} g^{a+}(s)ds < \infty$ 
  - (a)  $\implies C_3 = 0$
2.  $g^a(s)$  is continuous at the adoption threshold:  $\lim_{s \downarrow s_a} g^{a+}(s) = \lim_{s \uparrow s_a} g^{a-}(s)$ 
  - (a)  $C_1 = (C_4 - C_2)e^{-\delta s_a}$
3. there's no mass at the exit threshold:  $g^a(s_x^a) = 0$ 
  - (a)  $C_4 = C_2 \left( \frac{e^{-\delta s_a} - e^{-\delta s_x}}{e^{-\delta s_a}} \right)$
  - (b)  $\implies C_1 = -C_2e^{-\delta s_x}$
4. the mass of exitors is

$$\Delta E^a = (1-p)hg^{a-}(s_x^a + h) \approx (1-p)h [g^{a-}(s_x^a) + hg^{a-l}(s_x^a) + O(h^2)]$$

dividing by  $\Delta$  and taking the limit as  $\Delta \rightarrow 0$  implies  $E^a = \frac{\tilde{\Psi}^2}{2}g^{a-l}(s_x^a)$

(a) Now to recap we have

$$\begin{aligned} g^{a-}(s) &= -C_2e^{-\delta s_x} + C_2e^{-\delta s} \\ g^{a+}(s) &= C_2 \left( \frac{e^{-\delta s_a} - e^{-\delta s_x}}{e^{-\delta s_a}} \right) e^{-\delta s} \end{aligned}$$

(b)  $\implies C_2 = \frac{E^a 2}{-\delta \tilde{\Psi}^2} e^{\delta s_x^a} = \frac{E^a}{\tilde{\mu} e^{-\delta s_x^a}}$

(c) Note that in steady state, the mass of new adopters must equal the mass of adopters that exit,  $A = E^a$

5. So,

$$g^a(s) = \begin{cases} \frac{E^a}{\tilde{\mu}} (e^{-\delta(s-s_x^a)} - 1) & s \in (s_x^a, s_a) \\ \frac{E^a}{\tilde{\mu}} (e^{\delta s_x^a} - e^{\delta s_a}) e^{-\delta s} & s > s_a \end{cases}$$

6. Integrating over the mass of firms to get  $M^A$  gives  $M^A = \frac{E^a}{\tilde{\mu}}(s_x - s_a)$

7. Putting this in terms of  $z$

$$f^a(z) = \begin{cases} \frac{E^a}{\tilde{\mu}} (z^{\delta(1-\sigma)} e^{\delta s_x^a} - 1) \left(\frac{\sigma-1}{z}\right) & z \in (z_x^a, z_a) \\ \frac{E^a}{\tilde{\mu}} (e^{\delta s_x^a} - e^{\delta s_a}) z^{\delta(1-\sigma)} \left(\frac{\sigma-1}{z}\right) & z > z_a \end{cases}$$

### The distribution conditional on no adoption

The stochastic process for firm size is  $\frac{ds}{dt} = \tilde{\mu}dt + \tilde{\Psi}dW(t)$  and the same Kolmogorov Forward Equation is

$$\frac{\partial g^n(s)}{\partial t} = -\tilde{\mu} \frac{dg^n(s)}{ds} + \frac{1}{2} \tilde{\Psi}^2 \frac{d^2 g^n(s)}{ds^2}$$

In steady state  $\frac{\partial g^n(s)}{\partial t} = 0 \implies -\delta \frac{dg^n(s)}{ds} = \frac{d^2 g^n(s)}{ds^2}$  where  $\delta = \frac{-\tilde{\mu}}{\tilde{\Psi}^2/2}$ . The solution is going to be

$$g^n(s) = \begin{cases} g^{n-}(s) & s \in (s_x, s_e) \\ g^{n+}(s) & s \in (s_e, s_a) \end{cases}$$

and

$$\begin{aligned} g^{n-}(s) &= C_1 + C_2 e^{-\delta s} \\ g^{n+}(s) &= C_3 + C_4 e^{-\delta s} \end{aligned}$$

There will be 5 boundary conditions:

1. the mass of adopters is

$$\begin{aligned} \Delta A &= phg^{n+}(s_a - h) \approx ph [g^{n+}(s_a) - hg^{n+'}(s_a) + h^2 g^{n+''}(s_a)] \\ &= ph [-hg^{n+'}(s_a) + h^2 g^{n+''}(s_a)] \end{aligned}$$

where the second equality follows because  $g(s_a) = 0$ . Dividing both sides by  $\Delta$  and taking the limit as  $\Delta \rightarrow 0$  gives

$$A = -\frac{1}{2} \tilde{\Psi}^2 g^{n+'}(s_a)$$

2. there's no mass at the adoption threshold:  $g^n(s_a) = 0$

3.  $g^n(s)$  is continuous at the entry threshold:  $\lim_{s \downarrow s_e} g^{n+}(s) = \lim_{s \uparrow s_e} g^{n-}(s)$

4. there's no mass at the exit threshold:  $g^n(s_x) = 0$

5. the mass of exitors is

$$\Delta E^n = (1-p)hg^{n-}(s_x+h) \approx (1-p)h[g^{n-}(s_x) + hg^{n-l}(s_x) + O(h^2)]$$

Divide by  $\Delta$  and taking the limit as  $\Delta \rightarrow 0$  implies  $E^n = \frac{\tilde{\psi}^2}{2}g^{n-l}(s_x)$

6. In steady state the mass of entrants must equal the mass of non-adopters who exit plus the mass of firms that become adopters,  $N = E^n + A$ .

7. Applying 3 and 4 first gives

$$g^{n-}(s_x) = 0 \implies C_1 = -C_2e^{-\delta s_x}$$

Then

$$\begin{aligned} E^n &= \frac{\tilde{\Psi}^2}{2}g^{n-l}(s_x) = -\delta \frac{\tilde{\Psi}^2}{2}C_2e^{-\delta s_x} \implies C_2 = \frac{E^n}{\mu e^{-\delta s_x}} \\ &\implies C_1 = -\frac{E^n}{\mu} \\ &\implies g^{n-}(s) = \frac{E^n}{\tilde{\mu}} \left( e^{-\delta(s-s_x)} - 1 \right) \end{aligned}$$

Apply  $g^{n+}(s_a) = 0$ .

$$C_3 = -C_4e^{-\delta s_a}$$

Next apply  $g^{n-}(s_e) = g^{n+}(s_e)$

$$\begin{aligned} \frac{E^n}{\tilde{\mu}} \left( e^{-\delta(s_e-s_x)} - 1 \right) &= C_4 \left( e^{-\delta s_e} - e^{-\delta s_a} \right) \\ &\implies C_4 = \frac{E^n}{\tilde{\mu}} \left( \frac{e^{-\delta(s_e-s_x)} - 1}{e^{-\delta s_e} - e^{-\delta s_a}} \right) \\ &\implies g^{n+}(s) = \frac{E^n}{\tilde{\mu}} \left( \frac{e^{-\delta(s_e-s_x)} - 1}{e^{-\delta s_e} - e^{-\delta s_a}} \right) \left( e^{-\delta s} - e^{-\delta s_a} \right) \end{aligned}$$

8. So the final distribution is:

$$g^n(s) = \begin{cases} \frac{E^n}{\tilde{\mu}} \left( e^{-\delta(s-s_x)} - 1 \right) & s \in (s_x, s_e) \\ \frac{E^n}{\tilde{\mu}} \left( \frac{e^{-\delta(s_e-s_x)} - 1}{e^{-\delta s_e} - e^{-\delta s_a}} \right) \left( e^{-\delta s} - e^{-\delta s_a} \right) & s \in (s_e, s_a) \end{cases}$$

9. Integrating to get  $M^n$  gives

$$M^n = \frac{E^n}{-\delta \tilde{\mu}} \left( e^{-\delta(s_e-s_x)} - 1 \right) - \frac{E^n}{\tilde{\mu}} (s_e - s_x)$$

$$\begin{aligned}
& \dots + \frac{E^n}{\tilde{\delta}\tilde{\mu}} \left( e^{-\delta(s_e - s_x)} - 1 \right) - \frac{E^n}{\tilde{\mu}} \left( \frac{e^{-\delta(s_e - s_x)} - 1}{e^{-\delta s_e} - e^{-\delta s_a}} \right) e^{-\delta s_a} (s_a - s_e) \\
& = -\frac{E^n}{\tilde{\mu}} (s_e - s_x) - \frac{E^n}{\tilde{\mu}} \left( \frac{e^{-\delta(s_e - s_x)} - 1}{e^{-\delta s_e} - e^{-\delta s_a}} \right) e^{-\delta s_a} (s_a - s_e)
\end{aligned}$$

10. Use

$$\begin{aligned}
A & = -\frac{1}{2} \tilde{\Psi}^2 g^{n+1}(s_a) = -\frac{1}{2} \tilde{\Psi}^2 \left( -\delta \frac{E^n}{\tilde{\mu}} \left( \frac{e^{-\delta(s_e - s_x)} - 1}{e^{-\delta s_e} - e^{-\delta s_a}} \right) e^{-\delta s_a} \right) \\
& = -E^n \left( \frac{e^{-\delta(s_e - s_x)} - 1}{e^{-\delta s_e} - e^{-\delta s_a}} \right) e^{-\delta s_a}
\end{aligned}$$

to solve for  $A = E^a$  as a function of  $E^n$ . (where I used  $\delta = \frac{-\tilde{\mu}2}{\tilde{\Psi}^2}$ )

11. Distribution over  $z$

$$g^n(z) = \begin{cases} \frac{E^n}{\tilde{\mu}} (z^{\delta(1-\sigma)} e^{\delta s_x} - 1) \left( \frac{\sigma-1}{z} \right) & s \in (s_x, s_e) \\ \frac{E^n}{\tilde{\mu}} \left( \frac{e^{-\delta(s_e - s_x)} - 1}{e^{-\delta s_e} - e^{-\delta s_a}} \right) (z^{\delta(1-\sigma)} - e^{-\delta s_a}) \left( \frac{\sigma-1}{z} \right) & s \in (s_e, s_a) \end{cases}$$

## C Calibration

### C.1 Uniqueness of city fundamentals

In this section, I show that there is a unique set of fundamentals that rationalize the data being an equilibrium of the model in 1980. The vector of fundamentals is given by  $\mathbb{P}^c = \{\gamma, \psi, A_h, A_l, f^e, f^c, b, \eta\}$ ; that is, the weight on high-skilled labor,  $\gamma$ , the city specific productivity term,  $\psi$ , high- and low-skilled amenities,  $A_h$  and  $A_l$ , entry costs,  $f_e$ , fixed costs,  $f_c$ , the productivity of the building sector  $b$ , and the elasticity of building supply,  $\eta$ . Table V gives the moments in the data used to recover the parameters. I solve for the parameters recursively in the following steps:

1. Using the start-up rate, back out the implied exit threshold  $z_x$ :  $sr = \frac{\tilde{\mu}}{s_x - s_e}$  and  $z_x = e^{\frac{s_x}{\sigma-1}}$ . Knowing the exit threshold means the firm distribution given by equation 1 is determined.
2. Solve for  $\gamma_j$  using the two labor market clearing conditions. The two labor market clearing conditions are

$$\begin{aligned}
H_j & = h_j^d \left( Y_j + \frac{M_j f_j^c}{h_j^d + l_j^d} + \frac{E_j f_j^e}{h_j^d + l_j^d} \right) \\
L_j & = l_j^d \left( Y_j + \frac{M_j f_j^c}{h_j^d + l_j^d} + \frac{E_j f_j^e}{h_j^d + l_j^d} \right)
\end{aligned}$$

Divide  $\frac{H_j}{L_j} = \left( \frac{w_h}{w_l} \right)^{-\epsilon} \left( \frac{\gamma}{1-\gamma} \right)^\epsilon$ . The skill intensity of city  $j$ ,  $\frac{H_j}{L_j}$ , and the relative wages are both

data. Thus, rearranging to solve for  $\gamma$  gives  $\gamma_j = \frac{\left(\frac{H_j}{L_j}\right)^{\frac{1}{\epsilon}} \frac{w_h}{w_l}}{1 + \left(\frac{H_j}{L_j}\right)^{\frac{1}{\epsilon}} \frac{w_h}{w_l}}$ .

3. Next I use the condition that  $P_j = 1$  to solve for the level of wages and  $\psi_j$ . In city 1,  $\psi_1$  is normalized to be 1. Thus, I can solve

$$1 = \left( \int_{\Omega_j} \left( c_j(z) \frac{\sigma}{\sigma - 1} \right)^{1-\sigma} d\omega \right)^{\frac{1}{1-\sigma}}$$

where  $c_j(z) = (\psi_j z)^{-1} (\gamma_j^\epsilon w_{hj}^{1-\epsilon} + (1 - \gamma_j)^\epsilon w_{lj}^{1-\epsilon})^{\frac{1}{1-\epsilon}}$  and  $\Omega_j$  was pinned down in step 1 for  $w_{l1}$ . Rearranging gives

$$1 = w_{lj}^{\frac{1}{1-\sigma}} (\gamma_j^\epsilon \frac{w_{hj}^{1-\epsilon}}{w_{lj}^{1-\epsilon}} + (1 - \gamma_j)^\epsilon)^{\frac{1-\sigma}{1-\epsilon}} \left( \int_{\Omega_j} \left( \frac{1}{z} \frac{\sigma}{\sigma - 1} \right)^{1-\sigma} d\omega \right)$$

which means there is a unique solution for  $w_{l1}$ . Using  $w_{l1}$  pins down the level of wages in the whole economy and using data on relative wages in city  $j$ , the city-size wage premium and rents, I can solve for  $w_{hj}$ ,  $w_{lj}$  and  $r_j$  for all  $j$ .

4. Knowing wages, in all other cities  $j \geq 2$ , I can use  $P = 1$  to solve for  $\psi_j$ .

$$\psi_j = \left( \int_{\Omega_j} \left( (z)^{-1} (\gamma_j^\epsilon w_{hj}^{1-\epsilon} + (1 - \gamma_j)^\epsilon w_{lj}^{1-\epsilon})^{\frac{1}{1-\epsilon}} \frac{\sigma}{\sigma - 1} \right)^{1-\sigma} d\omega \right)^{\frac{1}{1-\sigma}}$$

5. Next, I use three equations to jointly solve for aggregate output  $Y$ , the fixed cost  $f^c$  and the entry cost  $f^e$ .

$$\text{final goods: } Y_j = w_{Hj} H_j + w_{Lj} L_j + \Pi_j^v - (\theta_j^h w_{hj} + \theta_j^l w_{lj}) (M_j f_j^c + N_j f_j^e)$$

$$\text{free entry: } V_j(z_e) = (\theta_j^h w_{hj} + \theta_j^l w_{lj}) f_j^e$$

$$\text{profit maximization (smooth pasting): } V_j'(z_x) = 0$$

where the smooth pasting condition can be re-arranged to solve uniquely for  $f^c$ ,

$$\frac{\rho^{\frac{(1-\xi^-)}{-\xi^-}} \frac{Z}{\rho - \tilde{\mu} - \frac{1}{2} \tilde{\Psi}^2} e^{s_x} - r f^b}{(\theta_j^h w_{hj} + \theta_j^l w_{lj})} = f^c.$$

Recall that  $Z_j \equiv \left( (\psi_j)^{-1} (\gamma_{jh}^\epsilon w_{jh}^{1-\epsilon} + \gamma_{jl}^\epsilon w_{jl}^{1-\epsilon})^{\frac{1}{1-\epsilon}} \frac{\sigma}{\sigma - 1} \right)^{1-\sigma} Y_j P_j \frac{1}{\sigma}$  for ease of notation. Then plugging the free entry condition and the equation for  $f^c$  into the market clearing condition gives a linear function of  $Y$  and therefore, the solution for  $Y_j$  will be unique. Knowing  $Y_j$  the free entry condition and smooth pasting condition can be used to solve for unique values of

$f_j^e$  and  $f_j^c$ .

6. The data on the elasticity of housing supply from [Saiz \(2010\)](#) maps directly into the parameter on the elasticity of building supply  $\eta_j$ . Then use the building market clearing condition to solve for  $b_j$

$$M_j f_b + (1 - \beta) H_j \frac{w_{hj}}{r_j} + (1 - \beta) L_j \frac{w_{lj}}{r_j} = b_j r^{\frac{1}{\eta-1}}$$

7. Finally, I use the labor market clearing to solve for amenities for each skill type. Define  $\bar{V}_{\tau j} \equiv \left( A_{\tau j} \beta^\beta (1 - \beta)^{(1-\beta)} w_{\tau j} q_j^{\beta-1} \right)^\nu$ . Then the probability that city  $j$  provides the highest utility to worker  $i$  will follow a multinomial logit

$$\pi_{\tau j} = P(V_{i\tau j} > V_{i\tau k}, \forall k \neq j) = \frac{\bar{V}_{\tau j}}{\sum_j \bar{V}_{\tau j}}.$$

In city 1 the amenities are normalized to be 1 so  $\bar{V}_{\tau 1}$  is known. Further, the  $\pi_{\tau j}$ s are known from the data. Define a vector  $x = [-\pi_{\tau 2} \bar{V}_{\tau 1}, \dots, -\pi_{\tau J} \bar{V}_{\tau 1}]$  and a matrix

$$P = \begin{bmatrix} \pi_{\tau 2} - 1 & \pi_{\tau 2} & \dots & \pi_{\tau 2} \\ \pi_{\tau 3} & \pi_{\tau 3} - 1 & \dots & \pi_{\tau 3} \\ \dots & \dots & \dots & \dots \\ \pi_{\tau J} & \pi_{\tau J} & \dots & \pi_{\tau J} - 1 \end{bmatrix}$$

then the the vector  $V = [\bar{V}_{\tau 2}, \dots, \bar{V}_{\tau J}]$  can be solved for

$$x = PV \implies V = P^{-1}x.$$

The determinant of the matrix P is not equal to zero, and therefore, there will be a unique solution for  $V$ , which can then be used to solve in closed form for  $A_{\tau j}$ .

## C.2 Robustness of calibration of parameters of the new technology

Figure [A.VI](#) and Figure [A.VII](#) show how the the main targets in the calibration and the main model results change with the parameters of the new technology:  $\Gamma_\gamma$ , which scales up the weight on high-skilled labor;  $\Gamma_\psi$ , which scales up the absolute productivity advantage of the new technology; and,  $\Gamma_{fc}$ , which gives the additional fixed cost of the new technology.

Figure [A.VI](#) shows the response of the main targets in the model: high- and low-skilled wage growth and average establishment size. In each case, a 5% increase or decrease in the parameters of the new technology changes the aggregate moments very little.

As  $\Gamma_\gamma$  increases, the new technology becomes more expensive to adopt which decreases adoption rates, but it also increases the weight placed on high-skilled labor. As a result, as  $\Gamma_\gamma$  increases, wages grow for high-skilled workers and fall for low-skilled workers while establishment size falls since adoption is less likely. As  $\Gamma_\psi$  increases, the new technology is more productive which increases



adoption rates and therefore SBTC. Wages for high-skilled workers again rise while they fall for low-skilled workers. The average establishment size rises since there is more adoption and the size of these firms will be larger. As  $\Gamma_{fc}$  increases, adoption becomes more expensive. Since there is less adoption wages for high-skilled workers fall while they rise for low-skilled workers. Establishment size increases since fixed costs are higher and selection is tougher.

Figure A.VII shows how the main results change with the parameters of the new technology. The main results include the share of the increase in the city-size wage premium for high-skilled workers, the share of the decrease in the city-size wage premium for low-skilled workers, the share of the changing relationship of skill-intensity and city size, and the share of the changing relationship of the start-up rate and city size.

As  $\Gamma_\gamma$  increases, high-skilled wages rise conditional on adoption, but it becomes more expensive for firms to adopt. Thus, as  $\Gamma_\gamma$  increases, the model explains a much larger share of the change in the city-size wage premium for high- and low-skilled workers. Nevertheless, with a 5% increase or decrease in  $\Gamma_\gamma$ , the model always accounts for a quantitatively significant portion of the cross-sectional results. As  $\Gamma_\psi$  and  $\Gamma_{fc}$  increase or decrease 5%, the main results are largely unchanged.

### C.3 Robustness with respect to fixed land cost, $f_b$

Figure A.VIII shows how the the main targets in the calibration and the main model results change with the fixed cost of land,  $f_b$ . The top row shows the response of the main targets in the model: high- and low-skilled wage growth and average establishment size. It also shows the response of the aggregate start-up rate. In each case, a 5% increase or decrease in  $f_b$  changes the aggregate moments very little.

The second row of Figure A.VIII shows how the main results change with  $f_b$ . The main results include the share of the increase in the city-size wage premium for high-skilled workers, the share of the decrease in the city-size wage premium for low-skilled workers, the share of the changing relationship of skill-intensity and city size, and the share of the changing relationship of the start-up rate and city size. In each case, the share of the cross-sectional results that the model explains is largely unchanged (note the tight axis on each of the graphs).

## D Additional Model Results

### D.1 Testing the model in changes

As an additional test of the success of the model, in Figure A.IX, I test the model in changes. For each of the four main variables, high-skilled wages, low-skilled wages, skill intensity and business dynamism, I show the relationship between the change in the data between 1980 and 2014 and the change in the model between the first and second steady state. With the exception of high-skilled wages, the relationship between the change in the data and the change in the model is positive and statistically significant at the 10% level. For high-skilled wages, the relationship is positive, but not

significant.

## D.2 Changes in market size: model vs. data

In this section, I verify that the relationship between population growth and initial population in the model are consistent with the data. This relationship is counter-intuitive since we would expect big cities that are doing more adoption to grow faster. However, these cities are also much more limited in their housing supply elasticity. Thus, even though they adopt more, the increase in congestion limits their population growth. Low-skilled workers, in particular, leave the big cities for less constrained smaller cities. This is consistent with evidence from [Rappaport \(2018\)](#) who documents a non-monotonic relationship of population growth to initial city-size. In particular, [Rappaport \(2018\)](#) finds that big cities grow less than less constrained and less crowded medium sized cities. However he finds that medium sized and big cities grew faster than rural areas and micropolitan statistical areas, which are not in my model.

## D.3 Changes in rent: model vs data

In this section, I show that the model does a good job of reproducing the changing relationship between rent and city-size in the data. In Figure [A.XI](#), I show the relationship between rent and city-size for 1980 in the data and in the model in black. By construction, the model matches the data perfectly in 1980. In gray, I show the relationship between rents and city-size in the 2014 data, while in green I show the relationship between rent and city-size in the second steady state of the model. In the data the correlation between rent and city-size increases from .075 to .126, while in the model it increases from .075 to .110. Thus, I slightly under-match the increasing relationship between rent and city-size in the data, but qualitatively, the model is very successful at reproducing this pattern.

## D.4 Understanding the drivers of the start-up rate

To understand what is driving the changes in the city level churn rates in the second steady state equilibrium, I proceed in three steps. First, I decompose the churn rates into three components: the churn rate for adopters, the churn rate for non-adopters and the share of firms that are adopters. Each of these components are functions of three endogenous objects: the exit threshold for adopters, the exit threshold for non-adopters and the adoption threshold. I show how the churn rates and the share of adopters are affected by each threshold. In the second step, I show how each of the thresholds, the churn rates and the share of adopters depend on wages, rents, market size and the availability of the new technology. Finally, I undertake a series of partial equilibrium counterfactuals to understand the main drivers of the changing churn rate in the second equilibrium. I ask what would have happened to the churn rates and the share of adopters had rent and wages been held constant at their 1980 levels or if market size,  $Y$ , increased equally across cities.

The start-up rate in the city is given by

$$sr_j = (1 - \mu_j)sr_j^{na} + \mu_j sr_j^a,$$

i.e. it is a weighted average of the start-up rate of non-adopters and adopters where the weight is given by the share of firms that have adopted,  $\mu_j$ . Now, I characterize the three objects in this equation  $sr_j^{na}$ ,  $sr_j^a$  and  $\mu_j$ .

Integrating the size distribution of adopters gives the start-up rate for adopters

$$sr_j^a = \frac{E_j^a}{M_j^a} = \frac{-\tilde{\mu}}{s_{aj} - s_{xj}}$$

where

$$s_x^a = \ln \left[ \frac{1}{Z_j^a} \left( \frac{-\xi^-}{1 - \xi^-} \right) \left( \frac{\rho - \tilde{\mu} - \frac{1}{2}\tilde{\psi}^2}{\rho} \right) ((\theta_j^h w_{hj} + \theta_j^l w_{lj})f_j^c + (\theta_j^{ha} w_{hj} + \theta_j^{la} w_{lj})\Gamma^{fc} + r_j f^b) \right]$$

and

$$Z_j^a = \left( (\psi_j^a)^{-1} \left( (\gamma_j^a)^\epsilon w_{jh}^{1-\epsilon} + (1 - \gamma_j^a)^\epsilon w_{jl}^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}} \frac{\sigma}{\sigma - 1} \right)^{1-\sigma} Y_j \frac{1}{\sigma}.$$

Similarly, the start-up rate for non-adopters is given by

$$sr_j^{na} = \frac{E_j^n}{M_j^n} = \frac{-\tilde{\mu}}{(s_e - s_{xj}) + \left( \frac{e^{-\delta(s_e - s_{xj})} - 1}{e^{-\delta(s_e - s_{aj})} - 1} \right) (s_{aj} - s_e)}$$

where  $s_x$  and  $s_a$  solve the smooth pasting conditions:  $v'(s_x) = 0$  and  $v'(s_a) = v^a(s_a)$ .

Finally, the share of adopters is given by

$$\begin{aligned} \mu_j &= \frac{M_j^a}{M_j^a + M_j^n} \\ &= \frac{(s_{xj}^a - s_{aj})}{(s_{xj}^a - s_{aj}) + \left( \frac{e^{-\delta(s_e - s_{xj})} - 1}{e^{-\delta s_e} - e^{-\delta s_{aj}}} \right) e^{-\delta s_{aj}} \left[ (s_e - s_{xj}) + \left( \frac{e^{-\delta(s_e - s_{xj})} - 1}{e^{-\delta s_e} - e^{-\delta s_{aj}}} \right) e^{-\delta s_{aj}} (s_{aj} - s_e) \right]} \\ &= \frac{1}{1 - \underbrace{\left( \frac{e^{-\delta(s_e - s_{xj})} - 1}{e^{-\delta(s_e - s_{aj})} - 1} \right)}_{<0} \left[ \frac{sr^a}{sr^{na}} \right]}. \end{aligned}$$

Three thresholds characterize the start-up rate for adopters and non-adopters and the share of adopters: the adoption threshold,  $s_j^a$ , the exit threshold for adopters,  $s_{xj}^a$ , and the exit threshold for non-adopters,  $s_{xj}$ . Thus, the start-up rate for adopters is characterized by the difference between the exit threshold for adopters and the adoption threshold, while the start-up rate for non-adopters is

characterized by the difference between the exit threshold for non-adopters and the entry threshold as well as the difference between the entry threshold and the adoption threshold. Finally, the share of adopters,  $\mu_j$ , is characterized by the difference between the entry and exit thresholds and the difference between the entry and adoption thresholds, as well as the relative churn rates for adopters and non-adopters.

In Figure A.XII, I show how the churn rate for adopters and non-adopters and the share of adopters changes with each exit and adoption threshold, holding the other thresholds fixed. These relationships are largely intuitive. The churn rate for adopters is in panel A.XIIa of Figure A.XII. Each line gives the churn rate for adopters for a percentage change in the exit threshold for non-adopters, the exit threshold for adopters and the adoption threshold. As the the exit threshold for non-adopters changes, holding the other thresholds fixed, the churn rate for adopters is unchanged. This is unsurprising as  $sr_j^a$  only depends on the exit threshold for adopters and the adoption threshold. As the exit threshold for adopters increases, firms are more likely to hit the exit barrier and the churn rate for adopters goes up. As the adoption threshold decreases firms adopt earlier, and holding the exit threshold for adopters fixed, they are more likely to be near the exit threshold increasing the churn rate for adopters.

Similarly, panel b plots the churn rate for non-adopters. The churn rate for non-adopters is unchanged by the exit threshold for adopters and increasing in the exit threshold for non-adopters. As the adoption threshold decreases, non-adopters are more likely to adopt and the churn rate for non-adopters increases.

Panel c shows the share of firms that adopt versus a percent change in each threshold. As the exit threshold for non-adopters increases, there will be less non-adopters in market and the share of adopters increases. Similarly, as the exit threshold for adopters increases, the share of adopters goes down since less of them remain in the market. The relationship between the adoption threshold and the share of adopters is non-monotonic. For low values of the adoption threshold, the share of adopters is increasing. This is because there is a tradeoff between a low adoption threshold, which makes firms more likely to hit the threshold and adopt but also means the new adopters are close to the exit threshold and do not stay in the market for long. As the adoption threshold rises, firms are less likely to hit the threshold, but also less likely to exit once they hit it.

Figure A.XIII shows how the churn rates, the share of adopters and the relative thresholds are affected by wages, relative wages, rents and market size. Panels a and b show how the churn rates and adoption are affected by a proportional increase in both high- and low-skilled wages. As wages increase, the churn rate for non-adopters increases. This is because both variable labor and fixed labor costs are higher and the firms need a higher productivity to stay in the market. The share of adopters goes down as wages increase because the additional fixed labor cost of the new technology becomes more burdensome as wages rise. The churn rate for non-adopters is largely unchanged as the exit threshold for adopters and the adoption threshold increase proportionally.

The second row shows what happens as high-skilled wages increase. As wages for high-skilled labor goes up, the share of adopters goes down since the return to adoption is lower. The churn rate

for adopters also falls slightly since the adoption threshold relative to the exit threshold for adopters increases and, on average, adopters will be further away from their exit threshold. The churn rate for non-adopters rises substantially as fixed and variable labor costs increase and selection becomes tougher.

In the third row, I show the how the churn rates and the share of adopters move with rent. As rent increases, so do the fixed costs paid by the firm. Selection becomes tougher and the share of adopters increases as does the churn rate for non-adopters. The churn rate for adopters is largely unaffected as the burden of the rent increase is much smaller.

Finally, row 4 shows what happens to churn rates and adoption rates and market size, given by  $Y$ , increases. As market size rises, churn rates go down as profits increase and firms are more willing to stay in the market for any given productivity. In addition, the share of firms that adopt increases as firms are more willing to absorb the additional fixed cost of adoption since they can spread it over a larger volume of sales.

## E Model Validation

### E.1 Model versus data: robustness check

This section shows robustness for Table VII. It shows the same set of regressions, but instead of using log of ICT spending per employee, it uses the ICT share of total investment as the dependent variable. In each case the sign of the coefficient is the same. However, the coefficient on the skill-premium loses significance when controlling for 1980 population and 1980 skill intensity. This is consistent with two off-setting forces in the model that make the relationship with the 1980 skill premium ex-ante ambiguous. With a high initial skill-premium, we would expect that the return to adopting would be lower, pushing down the relationship between ICT spending and the initial skill premium. On the other-hand, if the city is good at using high-skilled labor, the skill premium in the city should be high pushing up the relationship between ICT and the initial skill premium. Thus, the relationship we should see in the data between ICT spending and the initial skill premium is ambiguous, consistent with the weak relationship in the data.

## F Appendix Tables

|                    | (1)                    | (2)                     | (3)                    | (4)                    | (5)                    |
|--------------------|------------------------|-------------------------|------------------------|------------------------|------------------------|
|                    | <HS                    | HS                      | SC                     | C                      | >C                     |
| $\log(pop_{1980})$ | 0.0511***<br>(0.00550) | 0.0469***<br>(0.00357)  | 0.0504***<br>(0.00354) | 0.0642***<br>(0.00362) | 0.0605***<br>(0.00439) |
| $\log(pop_{2014})$ | -0.00225<br>(0.00518)  | 0.00928***<br>(0.00337) | 0.0296***<br>(0.00334) | 0.0629***<br>(0.00341) | 0.0757***<br>(0.00413) |
| Observations       | 1883                   | 1883                    | 1883                   | 1883                   | 1883                   |
| $R^2$              | 0.258                  | 0.172                   | 0.135                  | 0.327                  | 0.498                  |

Standard errors in parentheses

\*  $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

(a) Wages

|                    | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    |
|--------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                    | <HS                    | HS                     | SC                     | C                      | >C                     |
| $\log(pop_{1980})$ | 0.0336***<br>(0.00473) | 0.0350***<br>(0.00304) | 0.0363***<br>(0.00288) | 0.0431***<br>(0.00297) | 0.0498***<br>(0.00364) |
| $\log(pop_{2014})$ | 0.00236<br>(0.00445)   | 0.0185***<br>(0.00287) | 0.0325***<br>(0.00271) | 0.0605***<br>(0.00280) | 0.0714***<br>(0.00344) |
| Observations       | 1883                   | 1883                   | 1883                   | 1883                   | 1883                   |
| $R^2$              | 0.061                  | 0.925                  | 0.168                  | 0.815                  | 0.330                  |

Standard errors in parentheses

\*  $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

(b) Residual wages

|                    | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    |
|--------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                    | <HS                    | HS                     | SC                     | C                      | >C                     |
| $\log(pop_{1980})$ | 0.0228***<br>(0.00414) | 0.0212***<br>(0.00252) | 0.0217***<br>(0.00241) | 0.0198***<br>(0.00251) | 0.0296***<br>(0.00293) |
| $\log(pop_{2014})$ | 0.00529<br>(0.00391)   | 0.0201***<br>(0.00237) | 0.0272***<br>(0.00227) | 0.0366***<br>(0.00237) | 0.0433***<br>(0.00276) |
| Observations       | 1883                   | 1883                   | 1883                   | 1883                   | 1883                   |
| $R^2$              | 0.881                  | 0.132                  | 0.934                  | 0.884                  | 0.909                  |

Standard errors in parentheses

\*  $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

(c) Residual wages, occupation and industry controls

Table A.I: City-size wage premium, more education groups, all cities

Source: Wages are from the 1980 Decennial Census and 2014 ACS. Population is working age population (ages 20-24) from the Intercensal Population Estimates. An observation refers to a city-year pair. Tables display the relationship between log average adjusted wages and city-size. Wages are demeaned by year. Specifically, for each skill type  $\tau$ , city  $j$  and each year  $t$ , I plot are  $\log(w_{\tau jt}) - \log(\bar{w}_{\tau t})$ . Panels (a) shows average wages by skill group. Panels (b) shows wages adjusted for demographics including age, age squared, gender and race dummies. Panel (c) adds further controls for occupation and industry fixed effects.

|              | average ICT<br>investment share | average ICT<br>investment | average total<br>investment |
|--------------|---------------------------------|---------------------------|-----------------------------|
| non-matchers | 0.3579<br>(0.3543)              | 13.67<br>(1035)           | 76.78<br>(5217)             |
| matchers     | 0.3412<br>(0.3465)              | 64.98<br>(9738)           | 347.6<br>(34380)            |

Table A.II: Matchers versus non-matchers

Note: standard deviation is in parentheses. Average ICT investment and Total Investment are in thousands of U.S. dollars. Source: LBD and ACES data from the U.S. Census Bureau. Table compares characteristics of firms in the ACES that match to the LBD and firms in the ACES that do not match to the LBD.

|                                 | Data: $\frac{\text{ict inv.}}{\text{tot inv.}}$ |                      |                     |                        |
|---------------------------------|---|----------------------|---------------------|------------------------|
| $\log(\text{pop}_{1980})$       | 0.0138***<br>(0.00121)                          |                      |                     | 0.0110***<br>(0.00164) |
| $\frac{H_{1980}}{L_{1980}}$     |   | 0.238***<br>(0.0456) |                     | 0.112***<br>(0.0373)   |
| $\frac{w_h_{1980}}{w_l_{1980}}$ |   |                      | 0.0436*<br>(0.0239) | 0.0027<br>(0.0158)     |
| Observations                    | 269000  | 269000               | 269000              | 269000                 |
| $R^2$                           | 0.150   | 0.145                | 0.148               | 0.150                  |

SEs clustered at the CBSA level

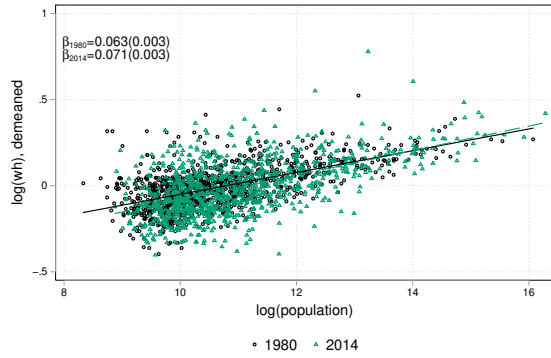
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table A.III: ICT spending vs. initial city characteristics

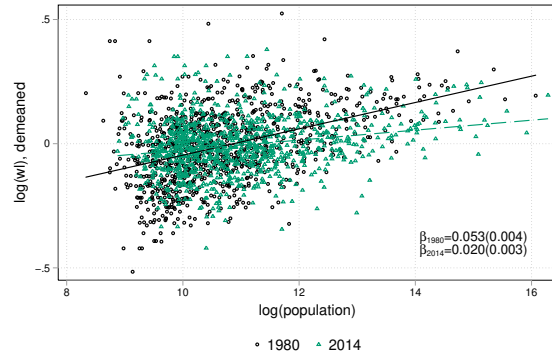
Source: The left panel uses ICT data from ACES merged into the LBD for information on the location of the firm. Population is working age (ages 20-64) population from the Intercensal Population Estimates. Table displays the relationship between log of ICT spending from 2003 to 2013 and 1980 city characteristics including city-size, skill intensity and the skill premium. The regressions include a full set of 4 digit NAICS fixed effects and year fixed effects. The unit of observation is a firm and city-size is measured at the CBSA level. Standard errors are clustered at the CBSA level. The right panel uses model generated output and shows the relationship between the share of firms that adopt in a city and 1980 city characteristics.



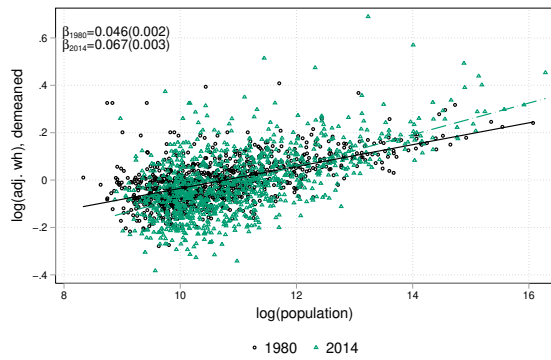
## G Appendix Figures



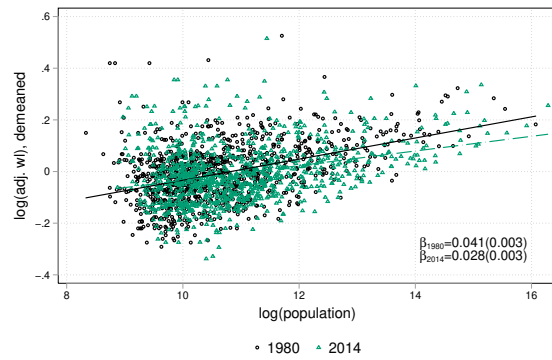
(a) High-skilled average wages



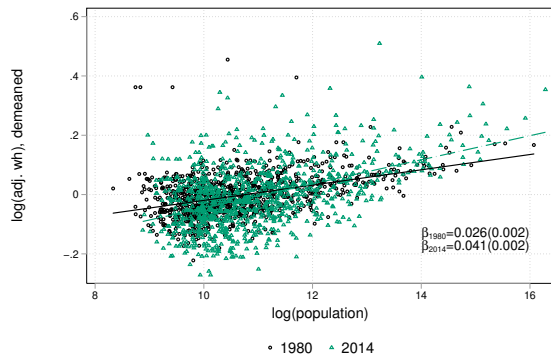
(b) Low-skilled average wages



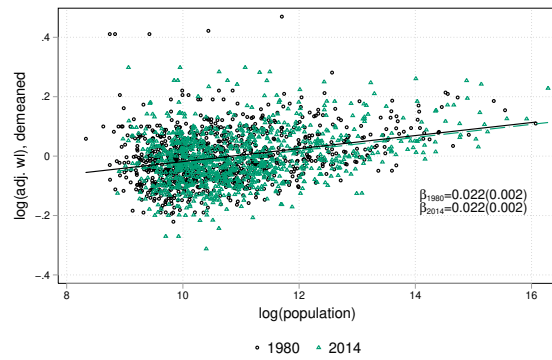
(c) High-skilled residual wages



(d) Low-skilled residual wages



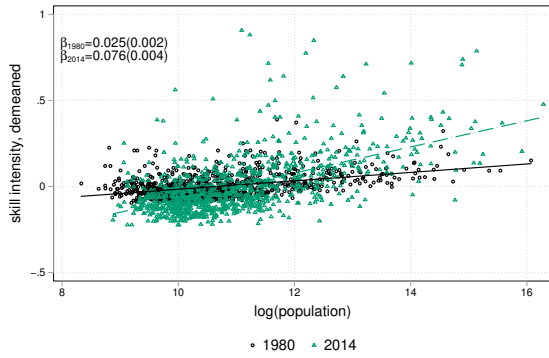
(e) High-skilled wages



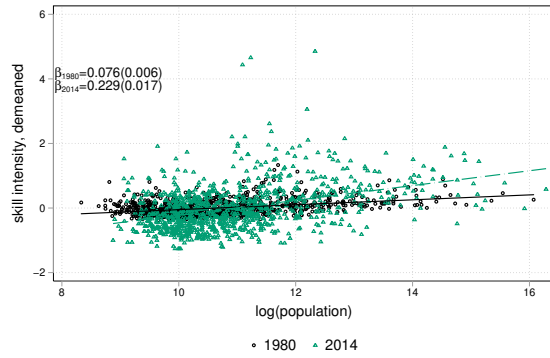
(f) Low-skilled wages

Figure A.I: City-size wage premium by skill group

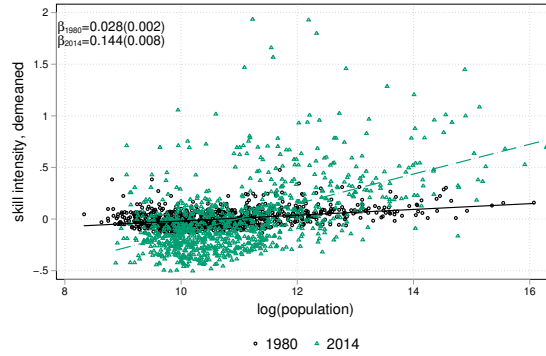
Source: Wages are from the 1980 Decennial Census and 2014 ACS. Population is working age population (ages 20-24) from the Intercensal Population Estimates. Figures display the relationship between log average adjusted wages and city-size. Wages are demeaned by year. Specifically, for each skill type  $\tau$ , city  $j$  and each year  $t$ , I plot are  $\log(w_{\tau jt}) - \log(\bar{w}_{\tau t})$ . Panels (a) and (b) show average wages by skill group. Panels (c) and (d) show wages adjusted for demographics including age, age squared, gender and race dummies. Panels (e) and (f) add further controls for occupation and industry fixed effects.



(a) More than college : Some college or less (baseline measure of skill intensity)



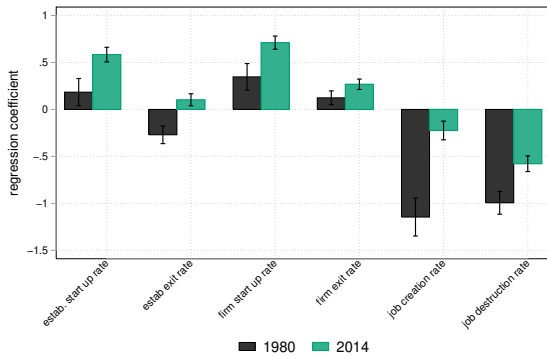
(b) More than HS : HS or less



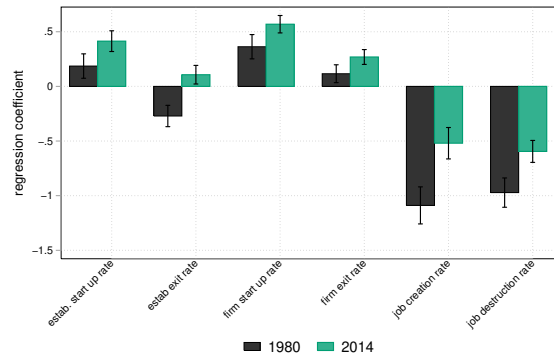
(c) College : High School

## Figure A.II: Skill intensity and city-size, all cities

Source: Skill intensity is from the 1980 Decennial Census and 2014 ACS. Population is working age (ages 20-64) from the Intercensal Population Estimates. Figure displays the relationship between demeaned skill intensity and city-size. Specifically for a city  $j$  and year  $t$  demeaned skill intensity is  $\frac{H}{L}_{jt} - \frac{H}{L}_t$ . Panel (a) defines  $H$  as those with 4+ years of college and  $L$  as those with less than 4 years of college. Panel (b) defines  $H$  as those with some college or more and  $L$  as those with a high school diploma or less. Panel (c) defines  $H$  as those with exactly 4 years of college and  $L$  as those with exactly a high school diploma.



(a) industry controls only



(b) adding demographic controls

Figure A.III: Correlation coefficients, dynamism and city-size controlling for industry

Source: LBD and author calculations. Population is working age population (ages 20-64) from the Intercensal Population Estimates. Figure displays the correlation coefficients between dynamism, as measured by the establishment start-up and exit rate, the firm start-up and exit rate and job creation and destruction, and city-size, controlling for a full set of year by NAICS 3 fixed effects. Panel (b) further controls for city level prime age worker share (ages 25-54) and ten year city-level population growth rates between years  $t$  and  $t - 10$ . The unit of observation is a CBSA  $\times$  3-digit NAICS bin. Standard errors are clustered at the CBSA level.

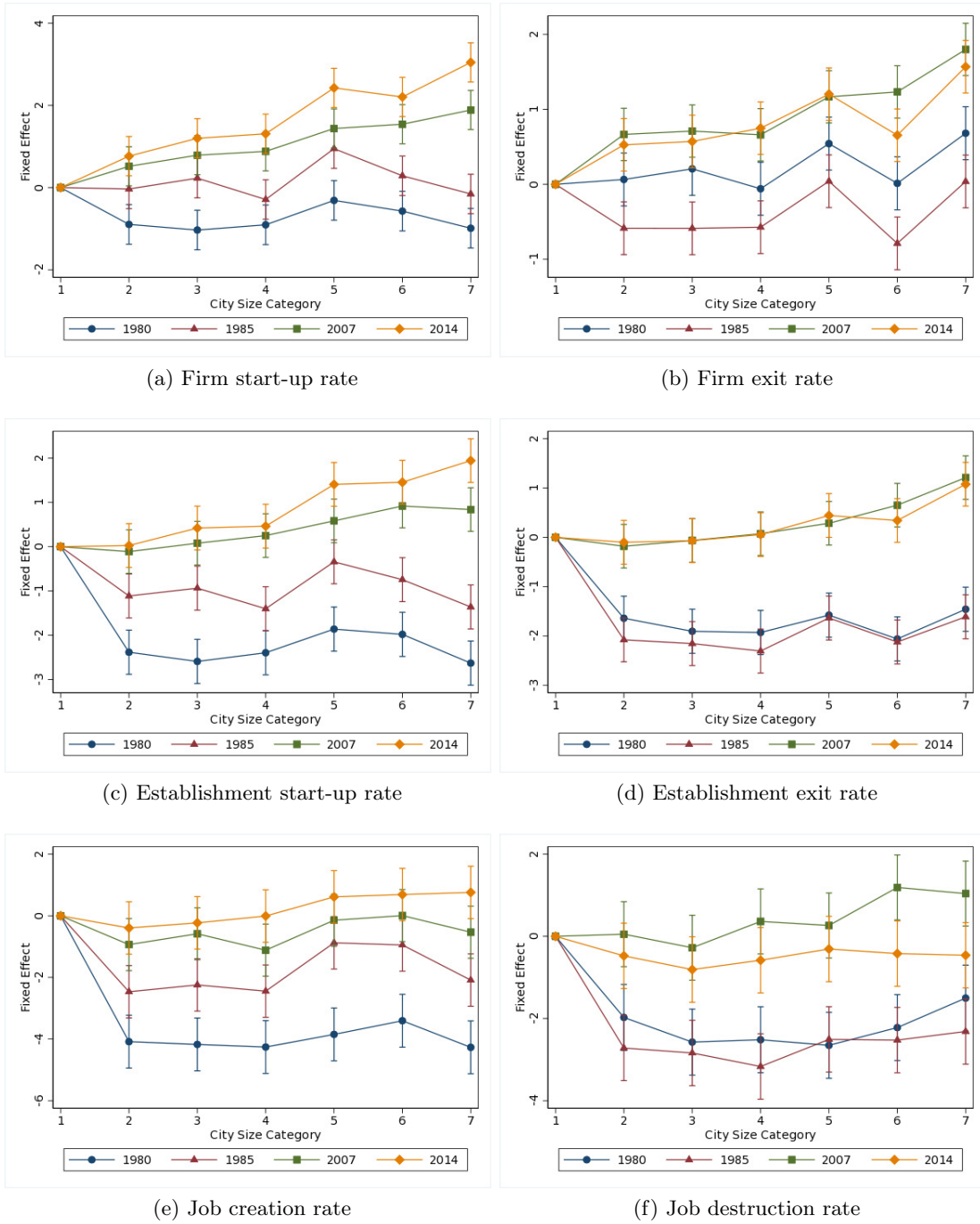
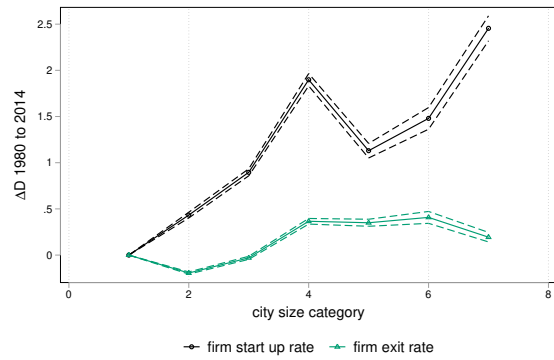
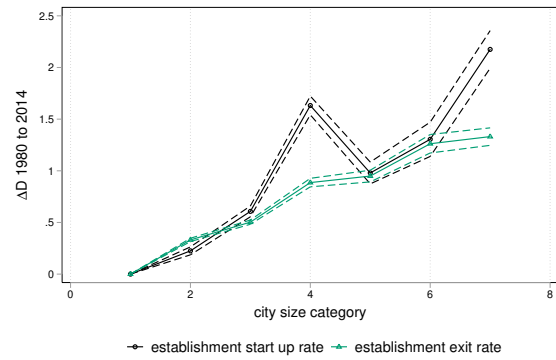


Figure A.IV: Dynamism by city-size category, NAICS 4 controls

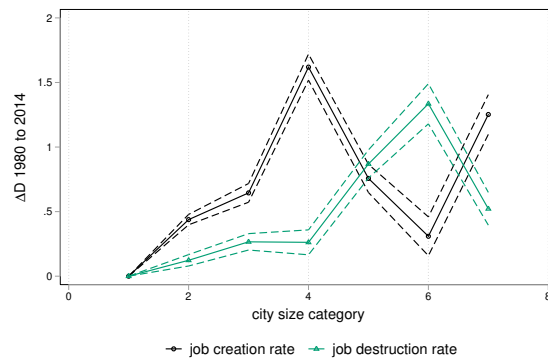
Source: LBD and author calculations. City size categories. Figure displays the city-size category fixed effects from a regression of dynamism, as measured by the establishment start-up and exit rate, the firm start-up and exit rate and job creation and destruction, on city-size, controlling for a full set of year by NAICS 4 fixed effects. Standard errors are clustered at the city-size category level.



(a) Firm start-up and exit rate



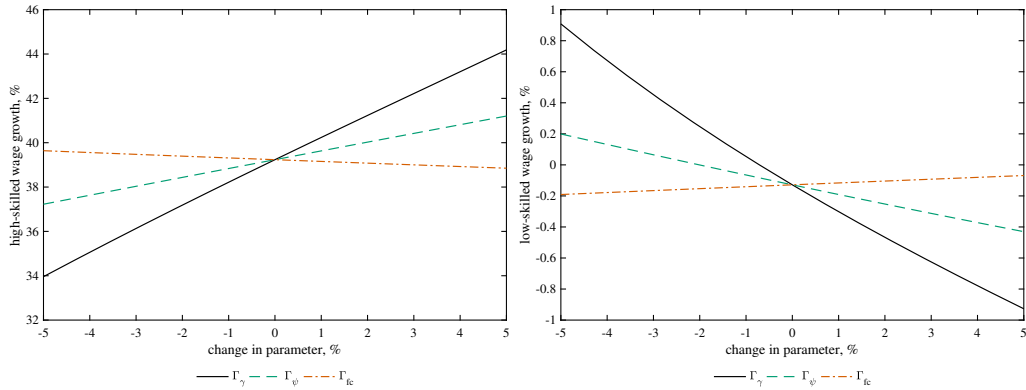
(b) Establishment start-up and exit rate



(c) Job creation and destruction rate

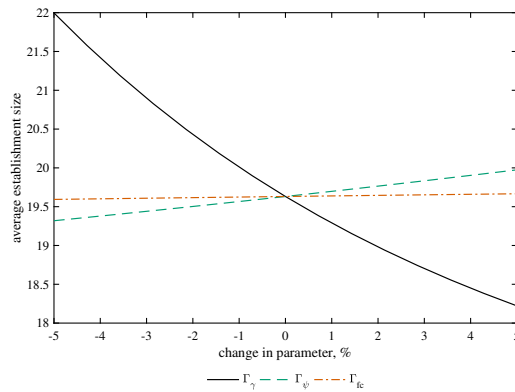
### Figure A.V: Decline in Dynamism and City Size

Source: LBD and author calculations. City size categories. Figure displays the city-size category fixed effects from a regression of the decline in dynamism between 1980 and 2014, as measured by the establishment start-up and exit rate, the firm start-up and exit rate and job creation and destruction, on city-size, controlling for a full set of NAICS 4 fixed effects. Standard errors are clustered at the city-size category level.



(a) high-skilled wage growth

(b) low-skilled wage growth



(c) average establishment size

Figure A.VI: Robustness of targets with parameters of the new technology

Source: Model generated output. Figures show how the main targets used to calibrate the new technology (high-skilled wage growth, low-skilled wage growth, and average establishment size) change as the three parameters of the new technology change.

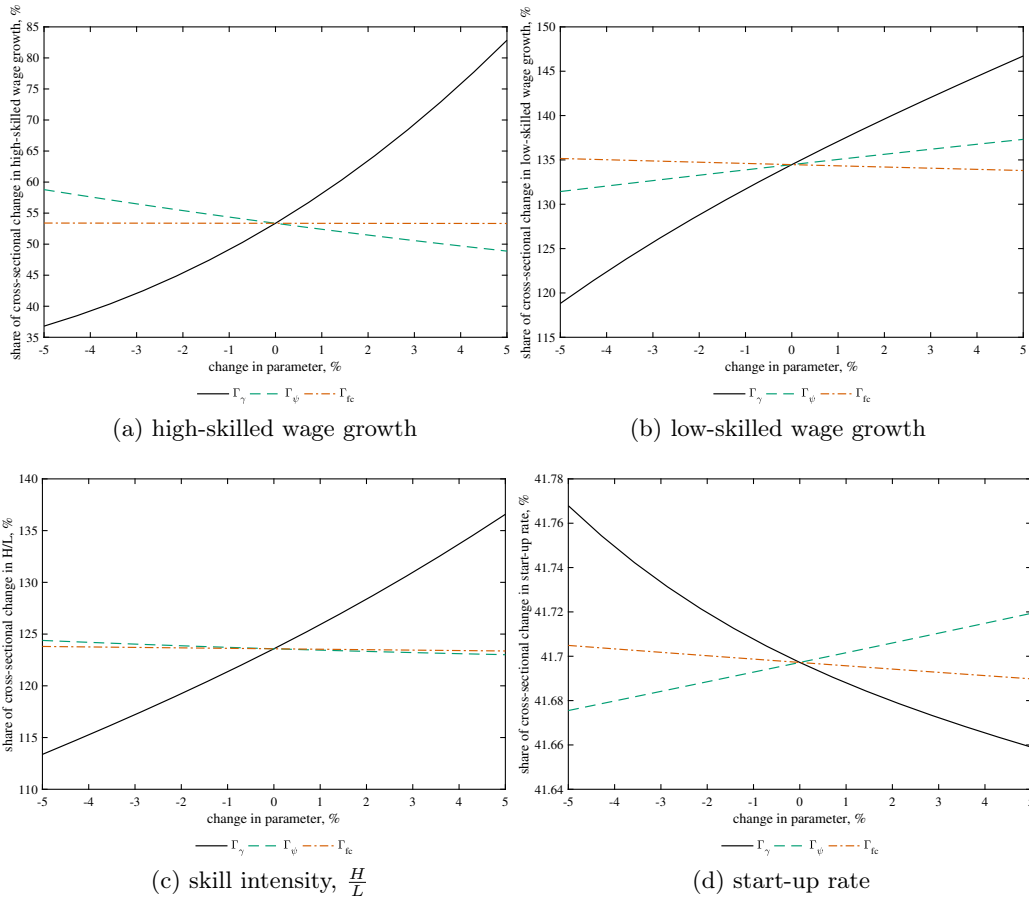


Figure A.VII: Robustness of main results to parameters of the new technology

Source: Model generated output. Figures show how the main results, the share of the cross sectional change in the four main variables (high-skilled wage growth, low-skilled wage growth, skill intensity and the start-up rate) change as the three parameters of the new technology change. For example, Panel (a) shows that even if  $\Gamma_\gamma$  was estimated to be 5% lower, the model would still explain 40% of the increase in the city-size wage premium for high-skilled workers.

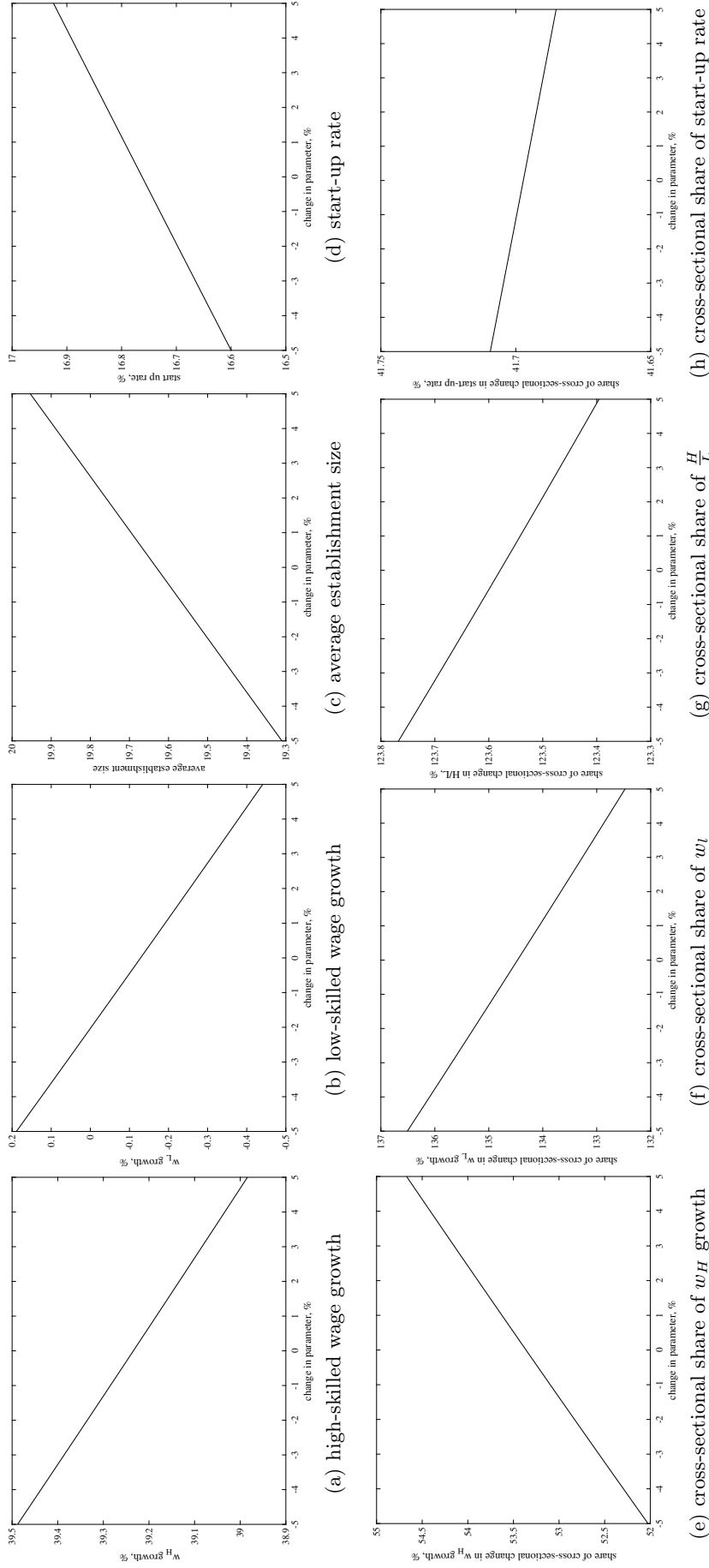
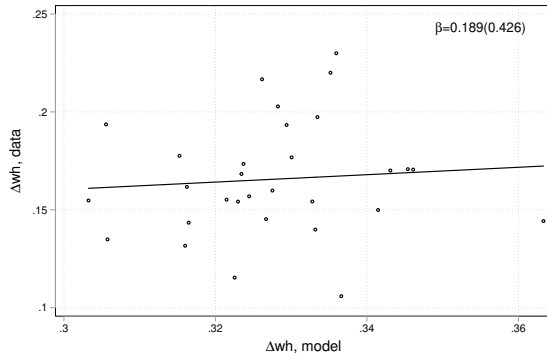


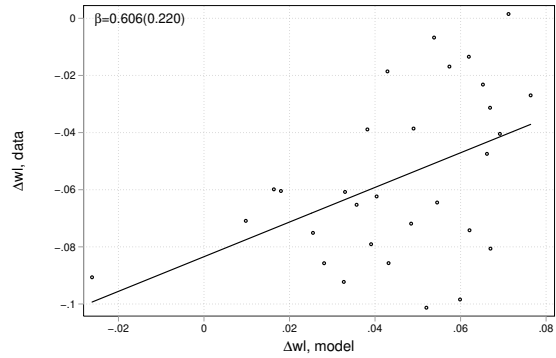
Figure A.VIII: Robustness of main results to fixed land price,  $f_b$

Source: Model generated output. Figures show how the calibration of the new technology and the main results, the share of the cross sectional change in the four main variables (high-skilled wage growth, low-skilled wage growth, skill intensity and the start-up rate) change as the fixed cost of land changes. For example, Panel (a) shows that even if the fixed cost of land was estimated to be 5% lower, the model would still predict high-skilled wage growth of  $\sim 35\%$ .

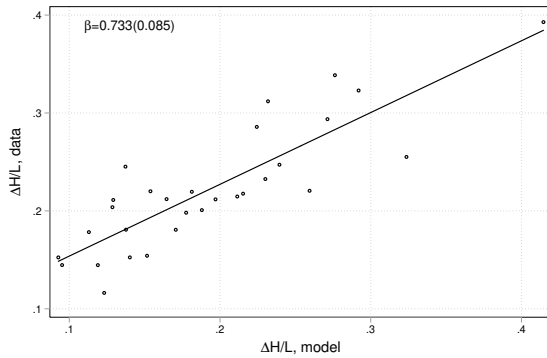




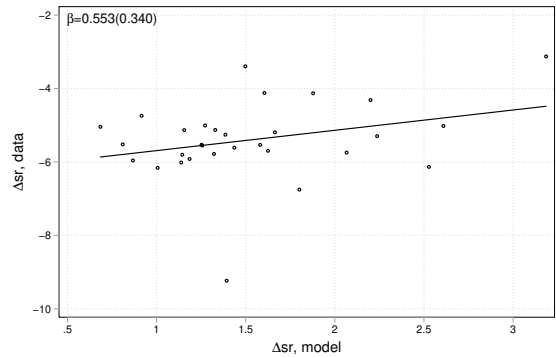
(a) high-skilled wages



(b) low-skilled wages



(c) skill intensity



(d) establishment start-up rate

Figure A.IX: Testing the model in changes

Note: Figures show the change in the data between 1980 and 2014 versus the change in the model between the first and second steady state.

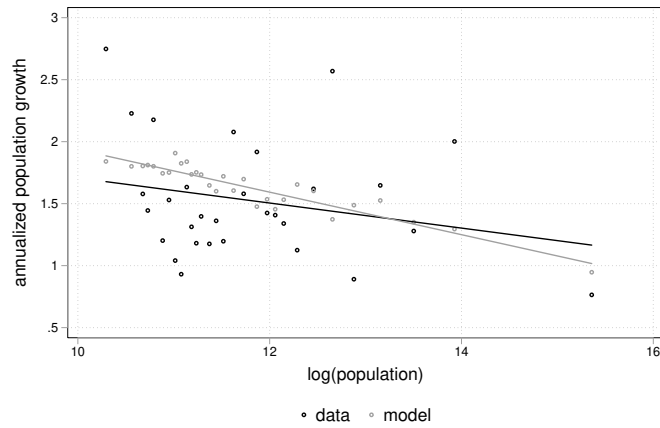


Figure A.X: Population growth vs initial population, model vs. data

Source: Population growth and population are working age population (ages 20-64) from the Intercensal Population Estimates. Additional data from author calculated model output. Figure displays the relationship between population growth and initial city-size in the model and the data. Black gives the 2014 data and green the 2014 steady state in the model. The unit of observation is one of 30 city-size categories.

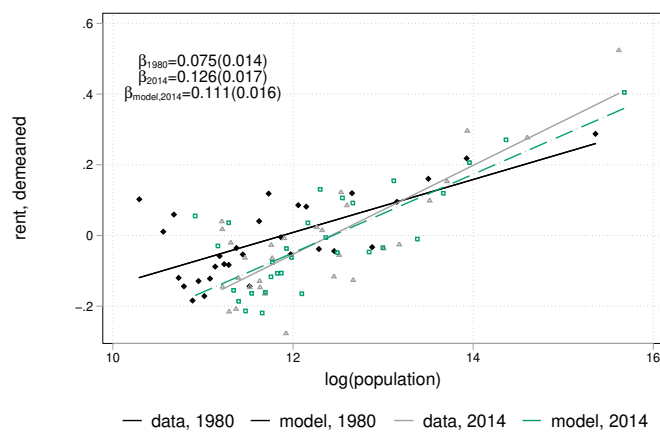
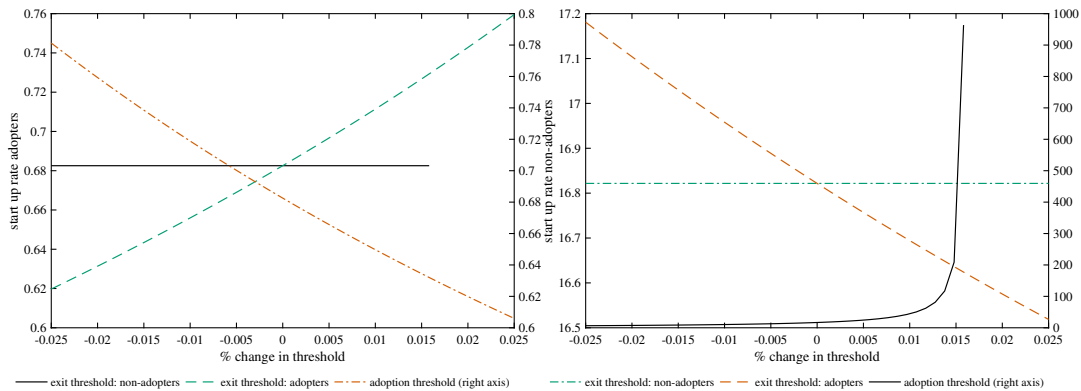


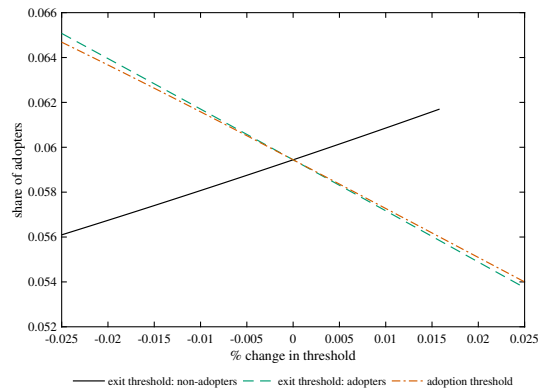
Figure A.XI: Rent vs population, model vs. data

Source: Rent is computed from the 1980 Decennial Census and 2014 ACS. Population is working age population (ages 20-64) from the Intercensal Population Estimates. Additional data from author calculated model output. Figure displays the relationship between rent and city-size in the model and the data. By construction, the model and data align perfectly in 1980 and are both shown in black. Gray gives the 2014 data and green the 2014 steady state in the model. The unit of observation is one of 30 city-size categories. All variables are demeaned by year.



(a) Churn rate for adopters

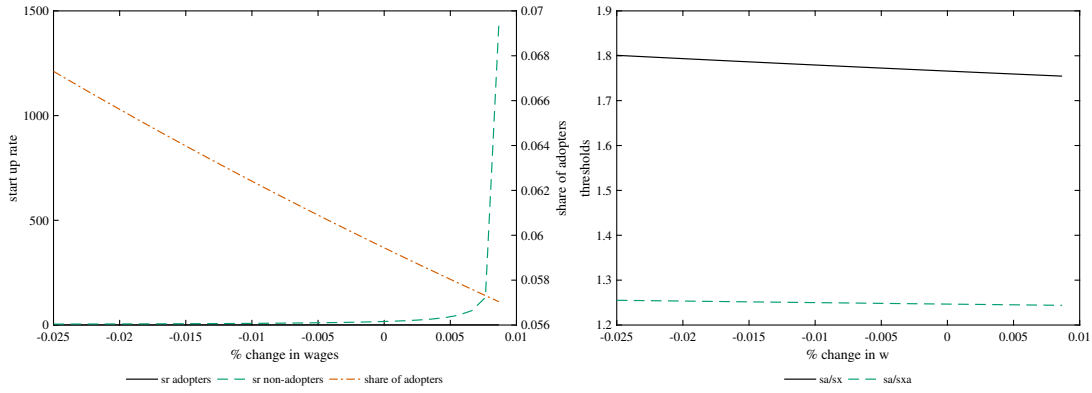
(b) Churn rate for non-adopters



(c) Share of adopters

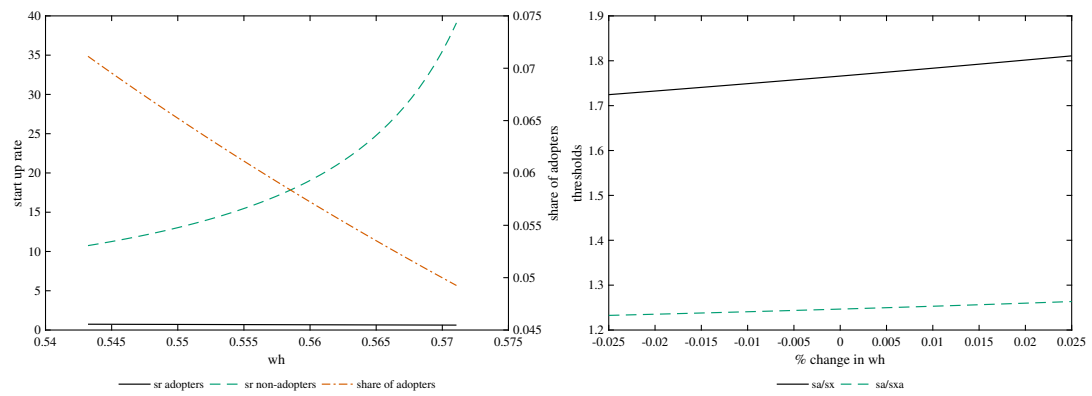
Figure A.XII: Churn rates and share of adopters versus exit and adoption thresholds

Source: Model calculations showing how measures of dynamism and the share of adopters change for a given city with respect to the three thresholds that pin down the firm size distribution.



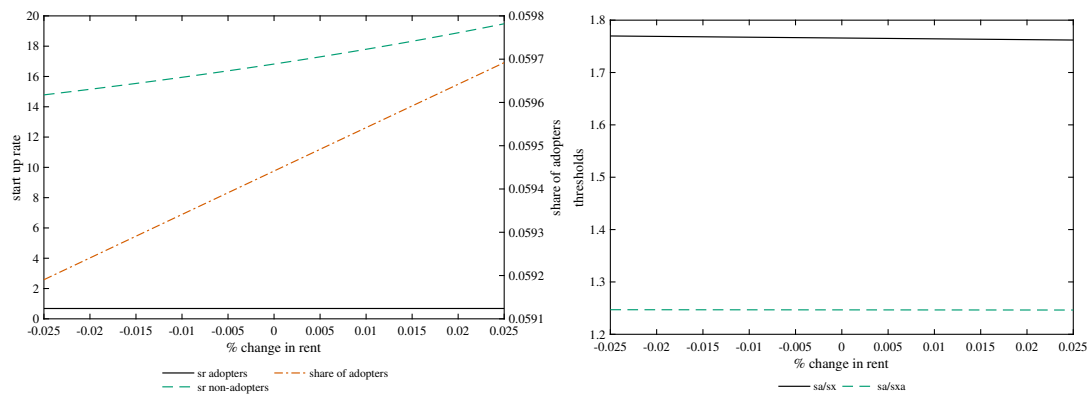
(a) Churn rates,  $\mu$

(b) thresholds



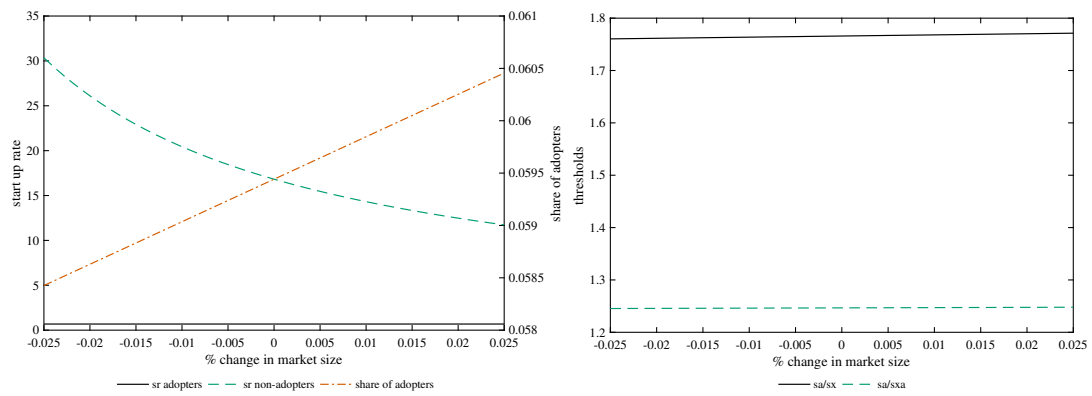
(c) Churn rates,  $\mu$

(d) thresholds



(e) Churn rates,  $\mu$

(f) thresholds



(g) Churn rates,  $\mu$

(h) thresholds

Figure A.XIII: Churn rates, share of adopters, and thresholds versus endogenous objects