ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

The Geography of Business Dynamism and Skill-Biased Technical Change

Authors	Hannah Rubinton
Working Paper Number	2020-020G
Revision Date	September 2025
Citable Link	https://doi.org/10.20955/wp.2020.020
Suggested Citation	Rubinton, H., 2025; The Geography of Business Dynamism and Skill-Biased Technical Change, Federal Reserve Bank of St. Louis Working Paper 2020-020. URL https://doi.org/10.20955/wp.2020.020

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

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

Hannah Rubinton†

September 3, 2025

Click here for the latest version

Abstract

This paper shows that the growing disparities between big and small cities in the U.S. since 1980 can be explained by firms endogenously responding to a skill-biased technology shock. With the introduction of a new skill-biased technology that is high fixed cost but low marginal cost, firms endogenously adopt more in big cities, cities that offer abundant amenities for high-skilled workers, and cities that are more productive in using high-skilled labor. In cities with more adoption, small and unproductive firms are more likely to exit the market, increasing the equilibrium rate of turnover or business dynamism—a selection effect similar to Melitz (2003). Differences in technology adoption and selection account for three key components of the growing regional disparities, known as the Great Divergence: (1) big cities saw a larger increase in the relative wages and supply of skilled workers, (2) big cities saw a smaller decline in business dynamism, and (3) firms in big cities invest more intensively in information and communication technology (ICT).

*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. I also thank Kurt Mitman and three anonymous referees for their helpful comments. Finally, 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 and 2539. The views expressed herein are solely those of the author and do not necessarily reflect those of the Federal Reserve Bank of St. Louis.

†Federal Reserve Bank of St. Louis; Email: hannah.rubinton@stls.frb.org

1 Introduction

Since 1980, economic growth in the United States has been concentrated in large urban areas, leaving small cities and rural areas behind. This trend, known as the Great Divergence (Moretti, 2012), became an increasing concern for policymakers as they learned of its implications for gaps in health outcomes, social connections, economic mobility, and political tensions.[1] Yet, little is known about the underlying causes.

This paper seeks a joint explanation of three key components of the Great Divergence. First, big cities experienced larger increases than small cities in the wages and supply of high-skilled workers relative to low-skilled workers. Second, big cities became more dynamic relative to small cities; in 1980, big and small cities had similar rates of business dynamism, but today big cities are more dynamic than small cities. Third, firms in big cities spend more intensively on information and communications technology (ICT) than firms in small cities. In the aggregate, the simultaneous increase in the wage growth and the abundance of high-skilled workers is typically explained by skill-biased technical change (SBTC) (Katz and Murphy, 1992). But, this aggregate story does not explain why SBTC was different across cities.

I develop a spatial equilibrium model with heterogeneous firms in which the amount of SBTC in each city is determined endogenously. I consider the introduction of a new technology that lowers the marginal cost for a firm, but places a higher weight on the use of high-skilled labor and has a higher fixed cost. The key insight from the model is that firms face different incentives to adopt the new technology based on the characteristics of their city. Firms adopt more in big cities, cities with amenities attractive to high-skilled workers, and cities that are ex-ante more productive in using high-skilled labor. The extent of adoption, and therefore SBTC, varies endogenously across cities, driving a divergence in TFP and explaining the changing relationships of wages and skill intensity with city size.

Differences in adoption rates across cities also affect rates of business dynamism. In cities with high adoption rates, wages and rents increase and the small, unproductive firms that do not adopt the new technology become less profitable. The productivity threshold below which they exit the market shifts up, and the probability that they exit increases, a selection effect similar to Melitz (2003). In models of firm dynamics, an increase in the probability of exit increases the equilibrium rate of turnover, or business dynamism. The firms most affected by the increase in selection are the less-productive firms using the old technology, which places more weight on low-skilled labor. As a result, the increase in selection amplifies the differences across cities in TFP growth and SBTC.

The growing divergence in wages, skill supply, and business dynamism between big and small cities has important implications for welfare inequality, economic mobility, and declining regional convergence (Diamond, 2016; Moretti, 2013; Baum-Snow et al., 2018; Giannone, 2017). Understanding the drivers of these trends is key to policymakers who wish to influence them. I make

[1]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).

progress by showing that they can be explained by differences in technology adoption across cities, which amplifies existing geographic inequalities along the dimensions for which the new technology is most suitable.

My analysis proceeds in several steps. First, I document empirical changes in the distribution of economic activity across cities since 1980. I focus on introducing two new facts to the literature on the Great Divergence. First, I document the changing relationship between business dynamism and city size. In 1980, measures of business dynamism, such as establishment entry and exit rates, were similar in big and small cities. By 2018, big cities exhibited much faster rates of dynamism than small cities. The increasing relationship between the probability an establishment exits the market and city size is most pronounced for establishments owned by young, rather than old firms. These patterns are consistent with selection becoming tougher in big cities relative to small cities, where selection is defined as the culling of unproductive firms from the market. Because new firms are, on average, smaller and less productive, they are more sensitive to changes in selection. Second, I use a novel dataset on ICT purchases matched with firm-level microdata to show that firms in big cities spend more on ICT per employee and devote a larger share of their total investment budget to ICT.[2] Finally, I reproduce two facts from the literature that are central to my analysis: big cities are increasingly more skill abundant and offer a higher skill premium than small cities.

My second contribution is to build a model that allows a joint consideration of the geographic distribution of relative wage inequality, firm dynamics, and technology adoption. I embed a rich model of firm dynamics into an otherwise standard spatial equilibrium model with high- and low-skilled workers. Within each city, there is a continuum of monopolistically competitive firms that use high- and low-skilled labor and floor space to produce a non-tradable intermediate good. Firms pay a fixed cost in units of high- and low-skilled labor and floor space in the city where they produce. Firms receive idiosyncratic productivity shocks and make dynamic entry and exit decisions. I calibrate the model to match key features of the data in 1980, including the cross-sectional patterns of wages, skill supply, and business dynamism.

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 it is with the old technology. Firms can choose to adopt the new technology, which lowers their marginal cost but 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, 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. I show that the model can match all of the changing relationship between the skill premium and city size, all of the changing relationship between skill intensity and city size, and just

[2] Closely related work by Beaudry et al. (2010) looks at city-level differences in computer purchases. I discuss the differences in our findings in Section 2.3.

over half of the changing relationship between business dynamism and city size.

In the model, three channels drive the higher technology adoption rates in the big city. The first channel is the market size effect. For a given productivity, firms in big cities receive a higher sales volume and are more willing to pay the higher fixed cost to achieve the marginal cost saving from adopting the new technology. The second channel 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 weight on skilled labor 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 is 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, further increasing the return to adoption. As a validation test of the model, I show that ICT spending is higher in cities already big in 1980, cities that were more skill-intensive in 1980, and cities with a higher skill premium in 1980, consistent with the patterns of technology adoption in the model.

The change in business dynamism across cities is driven by changes in 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 relatively better at using low-skilled labor than big firms that adopt the new technology. When selection becomes tougher, the small firms exit the market, amplifying the amount of SBTC that the city experiences and increasing dynamism rates in big cities relative to small cities. An added effect of selection is that it changes the shape of the firm productivity distribution, truncating it from below and increasing the mass of firms around the entry-productivity threshold. The change in the distribution of TFP amplifies the divergence in market-share weighted average of TFP between the big and small cities. I decompose within-city TFP growth and find that, on average, 46 percent of the growing gap between big and small cities is explained by selection rather than technology adoption. Similarly, 29 percent of the growing gap in the relative demand for skilled workers is explained by selection rather than adoption.

Finally, I use the model to analyze a series of policies aimed at increasing adoption. The analysis suggests a tension between increasing aggregate technology adoption and decreasing geographic inequality. First, I examine the effect of subsidizing the fixed cost of the new technology. Even if the government offers this subsidy everywhere, the take-up rate is endogenously higher for firms in big cities than in small cities—the skill premium increases everywhere, but even more in the big city. Second, I analyze a policy of subsidizing building construction in big cities. This policy substantially equalizes rents across cities, and workers move to the big cities. Thus, average wages and the total number of firms using the new technology increase due to the shift in population, but at the expense of small cities that lose population.

I make two contributions to the literature on the Great Divergence. The first is empirical. The previous literature has primarily focused on the geographic distribution of the skill premium and skilled workers (Baum-Snow et al., 2018; Moretti, 2013; Giannone, 2017). To the best of my knowledge, this is the first paper to document the changing relationship of business dynamism with city size and that the aggregate decline in dynamism was more severe in small cities.[3] This is also the first paper to document the positive relationship between ICT investment intensity and city size. Thus, an essential contribution of this paper is documenting these facts and showing that they are robust to controlling for industry composition and changing demographics. The second contribution is theoretical and quantitative. This paper proposes a new mechanism for the underlying cause of the Great Divergence: the uneven adoption across cities of a new skill-biased technology.[4] Contemporaneous work by Davis et al. (2019), Eeckhout et al. (2019), and Eckert et al. (2019) has a similar goal in that they show that introducing a new skill-biased technology will endogenously have different effects in different cities. My model complements these papers by adding a rich model of firm dynamics and assessing the role of selection in amplifying the Great Divergence.

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 a technology adoption decision. In Section 4, I discuss the calibration of the model and show that it can match the relevant features of the Great Divergence. In Section 5, I discuss the channels that drive differences in technology adoption across cities and provide a decomposition of within-city productivity growth into a component from adoption versus selection. Finally, in Section 6, I use the model to analyze a series of economic policies before concluding in Section 7.

2 Descriptive Facts

In this section, I document several facts on the relationship of firm behavior with city size. First, I document the changing relationship between business dynamism and city size between 1980 and 2018. In 1980, big and small cities exhibited similar rates of business dynamism, while today, big cities are more dynamic than small cities. I argue that part of the changing relationship between dynamism and city size can be explained by differential changes in selection across cities, where selection is defined as the culling of unproductive firms from the market. Specifically, in the model, selection is an increase in the productivity threshold below which firms exit the market. As selection becomes tougher, firms are more likely to exit, and the firm turnover rate increases. An increase in selection should affect young firms more than old firms because they are, on average, less productive and, therefore, closer to the exit threshold. I show that the increasing relationship between the probability of exit and city size is particularly pronounced for establishments that are part of young

[3] 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. Walsh (2020) analyzes age-conditional firm exit rates by city size but does not examine how these patterns change over time.

[4] The previous literature has instead focused on agglomeration economies that are skill-biased or a decline in communication costs (Baum-Snow et al. 2018; Rossi-Hansberg et al. 2019; Giannone 2017; Davis and Dingel 2014, 2019; Jiao and Tian 2019; Eckert 2019).

firms.

Second, I show that firms in big cities spend more on ICT per employee than firms in small cities, and a larger 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. Assuming that ICT use will be higher for firms that adopt new technologies, this pattern suggests that technology adoption has been more prevalent for firms in big cities than firms in small cities.

Finally, in Appendix A.1, I show that the relationships between the relative wages of skilled workers and the relative supply of skilled workers with city size have increased over time. These facts are well known in the literature on the Great Divergence, and I reproduce them as they are important inputs to my theoretical analysis.

2.1 Data description

Data on Cities To document facts on the changing relationship of economic activity with city size over time, I use several datasets on individuals and establishments from the U.S. Census Bureau. Throughout, I use Core-Based Statistical Areas (CBSAs) as the definition of cities.[5] 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. The final sample includes 934 CBSAs per year.

Data on Firm and Establishment Dynamics To document facts on business dynamics and technology adoption, I use several sources of firm and establishment microdata from the U.S. Census. The primary dataset is the public-use Business Dynamic Statistics (BDS), which is available at the CBSA by sector level and provides information on the number of firms and establishments, job creation, and job destruction by firm and establishment age. The BDS is made from the Longitudinal Business Database (LBD), a confidential panel dataset of all U.S. establishments with at least one employee. I use the LBD to show that the results are robust to controlling for industry composition across cities at levels of disaggregation that are finer than those available in the BDS.

The LBD includes information on establishment county, employment, industry, and payroll from 1976 to 2014. While the dataset is primarily at the establishment level, some of the analysis is at the firm level. Establishments are linked to their parent firm using a firm ID, and several firm-level variables must be created using information from the firm's establishments. Firm age is imputed from the first time its oldest establishment is observed in the LBD. Because the LBD started in 1976, meaningful age information begins in 1987 when firm age is known through 11+ years. Firm industry is imputed from the six-digit NAICS codes, which are available for each of the firm's establishments. When a firm has establishments in several industries, I assign a firm industry using the industry with the firm's highest share of employment. I assign the two-digit NAICS code first

[5] I use the 2018 definitions following the public use Business Dynamic Statistics.

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 at the establishment level in the LBD from Fort and Klimek (2018).

Throughout the analysis, I use establishment start-up and exit rates as measures of dynamism. Following Davis et al. (1996), the establishment start-up rate and exit rates at time t are defined as $esr_t = \frac{\text{age 0 establishments}_t}{(\text{establishments}_t + \text{establishments}_{t-1})/2}$ and $eer_t = \frac{\text{establishments exits}_t}{(\text{establishments}_t + \text{establishments}_{t-1})/2}$. Additional measures of dynamism are shown in Appendix A.2. I take five-year averages[6] of all dynamism rates to smooth out spurious variation in the data from city-industry bins with low firm counts and imputation error resulting from economic census years (years ending in 2 and 7).

Data on ICT Spending For ICT expenditures, I use data from the Annual Capital Expenditures Survey (ACES) and its ICT supplement between 2003 and 2013. The survey is a repeated cross-section of firms that includes data on capitalized and non-capitalized ICT expenditures and 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 development-related payroll.

Where possible, I supplement the ACES data with employment, payroll, sales, and industry information from the LBD and the Economic Census: 92% of records contain the firm identifiers necessary to match to the LBD producing a sample of approximately 256,000 firms. 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 is already available. The survey over-samples large firms but is representative of the U.S. economy when using sample weights.

For the ACES data, which is at the firm level, I assign a firm location using the location of their establishments in the LBD. When a firm has establishments in more than one location, I use the following algorithm to assign their location. First, if the firm has an establishment with a NAICS code of 55 (devoted to the management of companies), I assign the firm location as the location of that establishment. 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. In Appendix A.4, I show that the results are robust to using only single-establishment firms for which there is no ambiguity about their location. I also show results using computer purchases at the establishment level available for manufacturing firms in the Census of Manufacturers.

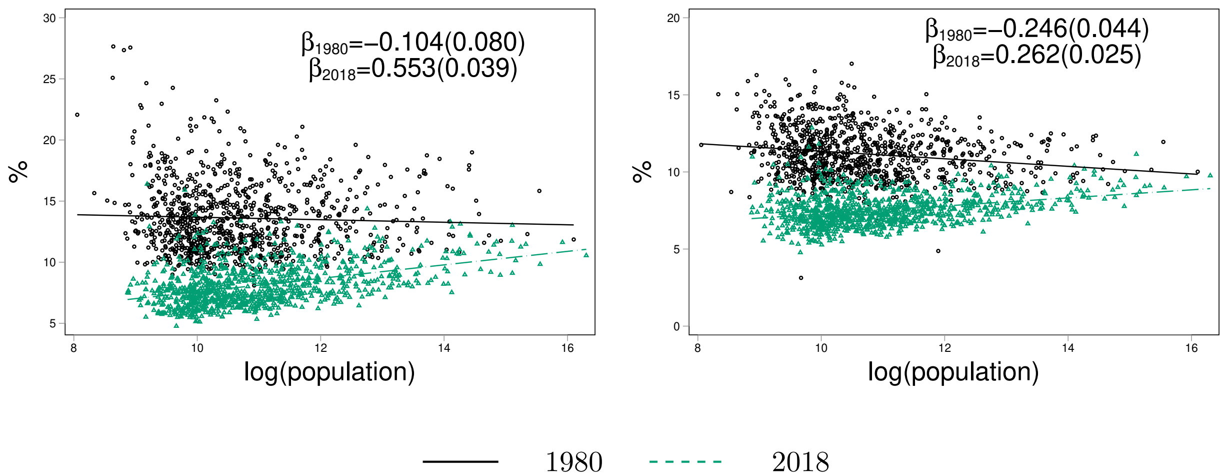
2.2 Firm Dynamics Across Cities

Fact 1: Big cities today are more dynamic than small cities. This was not true in 1980.

[6] Note that in 1980 the five year bin includes 1978-1982 while in every other year it includes year t to t-4. This is because the year t-4 is not available in the BDS data in 1980. For consistency with the Decennial Census data, I refer to this bin as 1980 throughout.

Figure I: Dynamism and city size

(a) Establishment start-up rate (b) Establishment exit rate



Note: Figure shows city-level dynamism, measured by establishment start-up and exit rates, plotted against working-age population (ages 20-64). Source: Business Dynamics Statistics and author calculations. Population data come from the Intercensal Population Estimates.

Figure I shows the relationship between the establishment start-up rate and city size in panel (a) and the establishment exit rate and city size in panel (b) for 1980 and 2018, using data from the public-use Business Dynamics Statistics. The well-documented aggregate decline in business dynamism (Decker et al., 2016; Karahan et al., 2019; Pugsley and Sahin, 2018) is apparent—the green line is substantially below the black line. However, what has not been previously documented is the change in the slope. In 1980, there was either no relationship or a negative relationship between dynamism and city size. By 2018, big cities had become more dynamic than small cities. In 2018, the semi-elasticity of the establishment start-up rate with respect to city size was 0.553 percentage points, meaning that the establishment start-up rate is almost half a percentage point higher in cities twice as large. Analogous results for the firm start-up and exit rates and job creation and destruction rates are presented in Appendix A.2.

These results are robust to controlling for city-level industry composition at the sector level. To show this, I compute dynamism measures within a city-industry cell and then estimate the following regression in each year t:

$$D_{ijt} = \alpha_{it} + \beta_t log(pop_{jt}) + \epsilon_{ict} \tag{1}$$

where D_{ijt} is the measure of dynamism for an industry i and a CBSA j in year t and α_{it} is a full set of industry fixed effects. The regression is run separately for each year, giving an estimate of the cross-sectional relationship between dynamism and city size, β_t.

The results are shown in Table I. The same pattern holds—the relationship between the establishment start-up rate and city size increased substantially between 1980 and 2018. In Appendix A.2, I show that the changing relationship between dynamism and city size is robust to controlling

Table I: Dynamism and city size, sector controls

	Establishment Start-up Rate		Establishment Exit Rate	
$log(pop_{1980})$	-0.305***		-0.234***	
	(0.0366)		(0.0229)	
$log(pop_{2018})$		0.423***		0.169***
		(0.0208)		(0.0151)
Observations	4534	4534	4534	4534
R^2	0.437	0.607	0.718	0.663

Note: Results are from a regression of dynamism–measured by establishment start-up and exit rates–on working-age population, controlling for sector fixed effects. Source: Business Dynamics Statistics and author calculations. Working-age population is from the Intercensal Population Estimates. The unit of observation is the CBSA-sector. CBSA-sector pairs with missing dynamism values in either year are excluded.

for differences across cities in population growth and the prime-age share of the population, which have been shown to be important drivers of dynamism (Karahan et al., 2019 and Engbom, 2017). I also use the LBD to show that the results are robust to using finer industry classifications.

The relationship between the establishment exit rate and city-size masks considerable heterogeneity in how the exit rate has changed across the firm-age distribution. In Table II, I show the results from estimating Equation (1) using establishment exit rates for young firms (ages 1 to 5) as the dependent variable.[7] Like exit rates overall, the relationship between exit for young firms and city-size has grown steeper over time. In 1980, establishments that were part of new firms were less likely to exit in big cities, while by 2018, there was no statistically significant difference across cities. In Appendix Table A.II, I show that this is not true for older firms.

Table II: Exit rates for young firms and city size

	Establishment Exit Rate: Ages 1-5	
$log(pop_{1980})$	-0.448***	
	(0.0497)	
$log(pop_{2018})$		0.00605
		(0.0351)
Observations	3106	3106
R^2	0.599	0.416

Note: Results are from a regression of the exit rate for establishments belonging to firms aged 1 to 5 on city working-age population, controlling for sector fixed effects. Source: Business Dynamics Statistics and author calculations. Population (ages 20-64) is from the Intercensal Population Estimates. The unit of observation is the CBSA-sector. CBSA-sector pairs with missing dynamism values in either year are excluded.

[7] The BDS releases tabulations of establishment exit by firm age rather than establishment age. The two differ in the case where an old multi-establishment firm opens a new establishment.

These patterns are consistent with selection increasing in big cities relative to small cities, as predicted by my model. When selection becomes tougher, the productivity threshold below which establishments exit the market increases (Combes et al., 2012; Melitz, 2003; Syverson, 2004). This predominantly impacts unproductive establishments that are close to the exit threshold. In a typical model of firm dynamics (e.g. Hopenhayn, 1992; Luttmer, 2007), new entrants have lower average productivity than older firms. Thus, if selection increases, we should primarily see an increase in exit for young establishments that are closer to the productivity threshold.

Using aggregate and age-dependent exit rates to identify selection is consistent with the work of Arkolakis (2016), Luttmer (2007), and Sampson (2015), who examine the implications of selection for aggregate productivity. However, the approach differs from the work of Combes et al. (2012), who examine differences in selection across cities. There are several differences between this work and the work of Combes et al. (2012), who find no differences in selection across cities in French firm-level data. The first differences are related to the sample, time period, and setting. I use the complete firm counts from the BDS, while their sample is limited to firms in the manufacturing sector and the consulting, advertising, and business services sector. They pool data between 1994 and 2002 and do not discuss whether the selection patterns across cities changed over time. The different results could also stem from differences in the urban environment between France and the United States. However, perhaps more important are the methodological differences. I argue that selection can be identified by looking at patterns of firm dynamics. Instead, they estimate the amount of shift, dilation, and selection that best transforms the distribution of firm productivity in small cities into the distribution of firm productivity in big cities. An important assumption in their approach is that tougher selection scales up the mass of the rest of the firm size distribution proportionally. This assumption holds in Melitz (2003)-style models in which the firm productivity is fixed after entry, but it is violated in typical models of firm dynamics (as in Hopenhayn, 1992 and Luttmer, 2007).

2.3 Information and Communications Technology Use Across Cities

Fact 2: Today, firms in big cities invest more in ICT than firms in small cities.

Table III shows the results from a regression of the log of ICT expenses per employee and ICT share of investment on city size, controlling for industry fixed effects. The elasticity of ICT spending per employee with respect to city size is .052 percent, meaning that firms in cities that are twice as large spend almost 5 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.

One possibility is that firms in big cities invest more intensively in all investment categories and that ICT is not unique. I investigate this in columns 3 and 4, which show the same regressions using the ICT share of total investment as the dependent variable. Firms in cities twice as large spend, on average, almost 1.2 percentage points more of their total investment budget on ICT-related

Table III: Technology adoption and city size

	$log(\frac{\text{ICT investment}}{\text{employment}})$		$\frac{\text{ICT investment}}{\text{total investment}}$	
log(pop)	0.0519***	0.0554***	0.0116***	0.0119***
	(0.00946)	(0.00737)	(0.00126)	(0.00126)
log(emp)		-0.429***		-0.0120***
		(0.00916)		(0.00167)
log(estabs)		0.479***		-0.00676***
		(0.0115)		(0.00220)
NAICS 4 FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Age FE	N	Y	N	Y
Observations	224000	224000	256000	256000
R-squared	0.234	0.351	0.172	0.175

SEs clustered at the CBSA level

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

Note: The table reports estimates from regressions of ICT spending on city size. ICT data come from ACES and are merged with the LBD for firm age, size, and location. Working-age population (ages 20-64) is from the Intercensal Population Estimates. Note that the difference in sample size between columns 1-2 and 3-4 is due to firms with zero ICT spending.

expenses such as computer equipment and software development. Controlling for size and age has little effect on the coefficient. The fact that the ICT share of total investment is increasing in city size suggests that, while investments in structures and equipment may also increase with city size, the relationship is stronger for ICT.

Because the data are only available starting in 2003, I interpret these results as informative about the second steady state in my model in which the new technology becomes available. In 1980, computers were not yet widely adopted; only about 5% of manufacturing firms purchased computers (Fort et al., 2018). In Appendix A.4, I show that these results are robust to only using single establishment firms. I also show that the results hold in an alternative dataset—the sample of manufacturing firms from the Census of Manufacturers, which contains information on computer purchases at the establishment level rather than the firm level. I find that even controlling for firm fixed effects, firms invest more intensively in computer purchases at their establishments in big cities than at those in small cities.

Relative to Beaudry et al. (2010), I make two contributions. First, I show that ICT spending per employee is increasing with city size. Beaudry et al. (2010) document that firm spending on personal computers is increasing in a city's relative supply of high-skilled workers and decreasing in the relative wages of high-skilled workers. Because both the relative wages of high-skilled workers and the relative supply of high-skilled workers are increasing in city size, the relationship between spending on personal computers and city size is indeterminate from their work. My second contribution relative to Beaudry et al. (2010) is to show that the ICT share of total investment is increasing in city size. Thus, while other types of investment may also increase in city size, the relationship between ICT spending and city size is stronger than for other investment categories.

3 A Spatial Equilibrium Model with Firm Dynamics

I build a model to formalize the hypothesis that the Great Divergence is driven by differences across cities in the adoption of a new skill-biased technology and then amplified by differential increases in selection. The model embeds a standard model of firm dynamics (Hopenhayn, 1992; Luttmer, 2007) and technology adoption (Bustos, 2011; Yeaple, 2005) into a benchmark spatial equilibrium framework (Roback, 1982).

Two types of workers, high- and low-skilled, 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, a final goods producer 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.

Each city has a continuum of monopolistically competitive firms that produce using high-skilled labor, low-skilled labor, and commercial floor space. Firms can choose between two bundles of technology; one with high fixed costs and low marginal costs that uses high-skilled labor more intensively and one with low fixed costs and high marginal costs that uses low-skilled labor more intensively. There is no sunk cost of adoption, so firms make a static adoption choice, trading off the higher fixed cost with the lower marginal cost. Fixed costs are paid in units of labor and land. They sell their output to a local final goods producer, who produces a non-differentiated, perfectly tradable final good.

Time is continuous, and firms are subject to an idiosyncratic 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 and land. Firms cannot move once they enter.

3.1 Final good producer

In each city j a final goods producer uses a continuum of local intermediate varieties, indexed by ω, aggregated according to a standard CES demand function

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

where Ω_j is determined endogenously and gives the set of firms operating in market j. Solving the profit maximization problem of the final goods producer, an intermediate firm faces the demand function

$$q_j(\omega) = Y_j P_j^{\sigma} p_j(\omega)^{-\sigma} \tag{2}$$

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

Intermediate firms have an idiosyncratic productivity term, $z(\omega)$, indexed by their variety, ω, which varies stochastically over time according to a geometric Brownian motion. Their profit maximization problem can be separated into a static component and a dynamic component. In the static problem, firms make a technology choice, trading off the fixed and marginal costs of using the old and new technology. Conditional on their technology, firms choose their optimal bundle of high-skilled labor, low-skilled labor, and commercial floor space to minimize their unit cost function, and they choose their 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 determine whether or not to enter.

Technology In each city j, a firm's technology is given by the bundle of parameters $\{\psi_j^k, \gamma_j^k, f_j^k\}$ where the ψ_j^k is a factor-neutral productivity term, γ_j^k is a skill-biased productivity term, and f_j^k is the fixed cost. There are two bundles of technology available indexed by k. Firms choose to be a non-adopter or an adopter, $k \in \{n, a\}$, trading off the benefit of the higher ψ technology against a higher fixed cost and a higher weight on high-skilled labor. The technology differs across cities to match key moments of the city-size distribution observed in 1980 (hence the subscript j on each technology term).

I assume that the adopter's technology scales up the parameters of the non-adopter's technology

$$\psi_j^a = \Gamma^\psi \psi_j^n; \;\; \gamma_j^a = \Gamma^\gamma \gamma_j^n; \;\; f_j^a = \Gamma^f f_j^n. \tag{3}$$

The parameters defining the difference between the old and new technology, $\{\Gamma^\psi, \Gamma^\gamma, \Gamma^f\}$, do not differ across cities so that the relative productivity differences across cities remain the same. $\{\Gamma^\psi, \Gamma^\gamma, \Gamma^f\}$ are all greater than 1. In the initial steady state, I assume that Γ^f is prohibitively large so that no firm adopts the new technology.

Static problem Given a bundle of technology, $\{\psi_j^k, \gamma_j^k, f_j^k\}$, intermediate firms in city j produce by combining labor and commercial floor space, b_{jt}, according to a Cobb-Douglas production function. Nested within the Cobb-Douglas function, they choose their labor inputs using a CES bundle of high- and low-skilled labor

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

where $z_t(\omega)$ is a firm-specific stochastically varying productivity, β_L is the cost share of production paid towards labor, and ϵ is the elasticity of substitution between low- and high-skilled labor.

First, firms choose the bundle of high- and low-skilled labor and commercial floor space that minimizes their unit cost function. Solving their cost minimization problem gives the unit cost function,

$$c_j^k(z) = C_j^k \frac{1}{\psi_j^k z}$$

for a firm in city j with productivity z where C_j^k is a cost index equal to

$$C_j^k = \frac{(\gamma_j^{k\epsilon} w_{hj}^{1-\epsilon} + (1-\gamma_j^k)^\epsilon w_{lj}^{1-\epsilon})^{\frac{\beta_L}{1-\epsilon}} r_j^{(1-\beta_L)}}{\beta_L^{\beta_L}(1-\beta_L)^{(1-\beta_L)}}.$$

The cost index depends on the technology k because their optimal bundle of inputs changes with the technology choice.

Next, the intermediate firm chooses output and price to maximize profits

$$\pi_j^k(z) = \max_{p_j(\omega), q_j(\omega)} p_j^k(\omega) q_j^k(\omega) - q_j^k(\omega) c_j^k(z) - C_j^k f_j^k \tag{4}$$

given the demand function from the final goods producer (equation 2). Their fixed costs, $C_j^k f_j^k$, are paid towards factors of production in the same proportion as variable costs.[8] Because fixed costs are paid in factors of production, they endogenously increase when the city becomes more congested and wages and rents increase. If they are paid in units of the final good, they become cheaper relative to their variable costs as factor prices increase. Appendix C.2 shows that the main results are robust to having the fixed costs paid in units of the final good.

Firms choose their price as a mark-up over their marginal cost, $p_j^k(z) = c_j^k(z)\frac{\sigma}{\sigma-1}$. Output is given by

$$q_j^k(\omega) = Y_j P_j^\sigma \left(\frac{\sigma}{\sigma-1} C_j^k\right)^{-\sigma} \left(\psi_j^k z\right)^\sigma \tag{5}$$

and variable profits are $\left(\psi_j^k z\right)^{\sigma-1} \left(C_j^k \frac{\sigma}{\sigma-1}\right)^{1-\sigma} Y_j P_j^\sigma \frac{1}{\sigma}$.

Technology adoption I assume that the firm's adoption choice is costless and reversible so that firms make a static choice to use the technology that maximizes their profits

$$k = \arg\max_{k \in \{n,a\}} \{\pi_j^n(z), \pi_j^a(z)\},$$

where profits are defined in Equation (4). As a result, solving the firm's technology adoption problem amounts to finding the threshold above which they use the new technology and below which they use the old technology. The marginal adopter will be indifferent between the old and

[8]Using the Hicksian demand functions for high-skilled labor $h_j^k(z)$, low-skilled labor $l_j^k(z)$, and land $b_j^k(z)$, the fixed cost can be written $F_j^k = \left(h_j^k(z) w_{hj} + l_j^k(z) w_{lj} + b_j^k(z) r_j\right) \psi_j^k z f_j^k = C_j^k f_j^k$.

new technology. The adoption threshold, z_{aj}, is found by setting their profits equal and is given by

$$log(z_{aj}) = \frac{1}{\sigma-1}\left[log \ \underbrace{\left(C_j^a f_j^a - C_j^n f_j^n\right)}_{\text{difference in fixed cost}} - log\left(\underbrace{\left(\frac{\psi_j^n}{C_j^n}\right)^{\sigma-1} - \left(\frac{\psi_j^a}{C_j^a}\right)^{\sigma-1}}_{\text{difference in productivity}}\right) - log\left(\underbrace{Y_j}_{\text{output}}\right) + log\left(\left(\frac{\sigma-1}{\sigma}\right)^{1-\sigma}\sigma\right)\right] \tag{6}$$

Three terms determine the productivity threshold above which firms adopt the new technology. First, the threshold is decreasing in market size, Y_j. Second, it is decreasing in the productivity difference between the old technology, $\left(\frac{\psi_j^n}{C_j^n}\right)$, and the new technology, $\left(\frac{\psi_j^a}{C_j^a}\right)$.[9] Finally, the adoption threshold is increasing in the difference between the fixed cost of using the new technology, $C_j^a f_j^a$, and the old technology, $C_j^n f_j^n$. Note that as the adoption threshold decreases (increases), more (less) firms become adopters.

Dynamic problem exit The productivity of intermediate firms varies over time according to a geometric Brownian motion, and firms discount future profits at an interest rate ρ. The dynamic problem of the firm is to choose the optimal exit threshold z_x, below which they will exit the market

$$V(z) = \max_{z_{xj} \leq z} \mathbb{E}_z \int_0^{T(z_{xj})} e^{-\rho t} \left(\pi_j(z(t))\right) dt$$

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

where $T(z_{xj})$ is the time at exit, $V(z)$ is the value of a firm with productivity z, and the drift of the productivity, μ, is assumed to be less than 0. Defining $s = \ln z^{\sigma-1}$, I show in Appendix B.1 that the exit threshold is

$$\begin{aligned} s_x = ln\left(\frac{R_2-1}{R_2}\right) + ln\left(\left(C_j^a f_j^a - C_j^n f_j^n\right) e^{-R_2 s_a} + C_j^n f_j^n\right)... \\ - ln\left(Y_j P_j^\sigma \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\left(e^{s_a - R_2 s_a}\left(\left(\frac{\psi_j^n}{C_j^n}\right)^{\sigma-1} - \left(\frac{\psi_j^a}{C_j^a}\right)^{\sigma-1}\right) + \left(\frac{C_j^n}{\psi_j^n}\right)\right)\right) \end{aligned} \tag{7}$$

where $R_2 = \frac{-\tilde{\mu}+\sqrt{\tilde{\mu}^2+2\tilde{\Psi}^2\rho}}{\tilde{\Psi}^2}$.

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 and floor space and start with initial productivity of z_e. Firms choose the city in which to enter that will maximize their expected value.

[9] Specifically, these terms are the Hicks-neutral productivity, ψ_j^k augmented by the cost index of each technology C_j^k. The productivity terms are raised to the $\sigma-1$, which gives the extent to which productivity differences translate into variable profits in the CES framework.

Once they start in a city, they cannot move. As a result, a free entry condition holds in each market

$$V_j(z_e) = C_j^n f_j^e, \tag{8}$$

such that entrants are indifferent between potential entry locations. As with fixed costs, entry costs are paid towards factors of production in the same proportion as variable costs.

Firms enter with an initial productivity z_e. In the calibrated model, all entering firms use the old technology. Once they enter, their productivity starts to drift down, but a firm may get lucky with a positive productivity shock. If they cross the adoption threshold z_a, they start to use the new technology. Eventually, their productivity will fall, and they will switch to the old technology before exiting.

3.3 Firm size distribution

The mass of firms with size s is given by $g(s)$.[10] This evolves according to the Kolmogorov Forward Equation (KFE)

$$\frac{\partial g_j(s)}{\partial t} = -\tilde{\mu}\frac{dg_j(s)}{ds} + \frac{1}{2}\tilde{\psi}^2\frac{d^2 g_j(s)}{ds^2}. \tag{9}$$

The left-hand side shows how the mass of firms with size s evolves over time. This depends on the drift of these firms, $\tilde{\mu}$, as well as the variance of the underlying stochastic process for size, $\tilde{\psi}^2$.[11] In steady state, the distribution must be unchanging, $\frac{\partial g_j(s)}{dt} = 0$, 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 in closed form and is described by the following equation:

$$g_j(s|s_e) = \begin{cases} \frac{E_j}{\tilde{\mu}}\left(e^{-\delta(s-s_{xj})} - 1\right) & s \in (s_{xj}, s_e) \\ \frac{E_j}{\tilde{\mu}}\left(e^{\delta s_{xj}} - e^{\delta s_e}\right)e^{-\delta s} & s > s_e. \end{cases} \tag{10}$$

Integrating over the distribution of firms in city j gives the mass of adopters and non-adopters

$$M^a = \int_{s_{aj}}^{\infty} g_j(s) = \frac{E_j}{\tilde{\mu}}\left(e^{\delta s_{xj}} - e^{\delta s_e}\right)\frac{e^{-\delta s_a}}{\delta} \tag{11}$$

$$M^n = \int_{s_{xj}}^{s_{aj}} g_j(s) = \frac{E_j}{\tilde{\mu}}\left(\frac{-e^{-\delta(s_e - s_x)} + 1}{\delta} - (s_e - s_x)\right) - \frac{E_j}{\tilde{\mu}}\left(e^{\delta s_{xj}} - e^{\delta s_e}\right)\frac{e^{-\delta s_a} - e^{-\delta s_e}}{\delta}. \tag{12}$$

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

$$\text{StartUpRate}_j = \frac{E_j}{M_j^a + M_j^n} = \frac{-\tilde{\mu}}{s_e - s_{xj}}. \tag{13}$$

[10] As above, firm size s is a function of productivity, $s(z) = \ln z^{\sigma-1}$.

[11] For more on KFEs in macroeconomics, see Achdou et al. (2021).

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, i.e., there is an increase in selection, the probability that a firm exits increases. Since the entry rate and exit rate must be equal in steady state, an increase in 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 τ in city j is given by

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

subject to their budget constraint

$$w_{\tau j} = c_{\tau j} + r_j b_{\tau j},$$

where r_j is the price of floor space and ζ_{ij} is an idiosyncratic location preference drawn from a Fréchet distribution with shape parameter ν, $\Pr(\zeta_{ij} < x) = \exp(-x^{-\nu})$. $A_{\tau j}$ is a type-specific amenity enjoyed when living in city j. Define $\bar{V}_{\tau j}$ as the indirect utility of an individual of type τ who lives in city j net of their idiosyncratic preference shock. Then the probability that a worker of skill τ chooses to live in city j follows a multinomial logit

$$\pi_{\tau j} = P(V_{i\tau j} > V_{i\tau k}, \forall k \neq j) = \frac{\bar{V}_{\tau j}^{\nu}}{\sum_k \bar{V}_{\tau k}^{\nu}}. \tag{14}$$

3.5 Land

A landlord uses the final good to build floorspace according to the production function $B = \left(\phi^{\frac{1}{\eta_j}} \frac{1+\eta_j}{\eta_j} Y\right)^{\frac{\eta_j}{1+\eta_j}}$. They rent it to firms and workers at the rental rate r. The equilibrium amount of floor space is an increasing function of rent,

$$B_j = \phi_j r_j^{\eta_j},$$

where η_j is the elasticity of building supply and ϕ_j is the productivity of the building sector.

3.6 Equilibrium

Profits are made by the landlords and by the owners of the firms who 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.

A steady-state equilibrium is a set of prices, labor allocations, output, the mass of firms and entrants, the building supply, and exit and adoption thresholds, $\{w_{lj}, w_{hj}, r_j, L_j, H_j, Y_j, M_j, E_j, B_j, z_x, z_a\}$, for each city, such that:

1. there is an invariant distribution of firms, $g_j(s)$, that satisfies the Kolmogorov Forward Equation (Equation 9),
2. there is a constant mass of firms and entrants, M_{jt} and E_{jt},
3. exit thresholds, z_{xj}, satisfy Equation (7) and the free entry condition holds (Equation 8) in each city,
4. the adoption thresholds, z_{aj}, satisfy Equation (6)
5. firms and landlords maximize profits (Equation 4)
6. workers choose their city (Equation 14), consumption, and housing to maximize utility,
7. the final goods market clears in each city,

$$\beta w_{Hj} H_j + \beta w_{Lj} L_j + \left(\Pi_j^v - M_j^n C_j^n f_j^n - M_j^a C_j^a f_j^a - E_j C_j^n f_j^e\right) + \Pi_j^B + c_j^B = Y_j$$

where Π_j^B and c_j^B give the profits and costs of the landlord, respectively; the land market clears in each city,

$$M_j \int_z b_j(z) q_j(z) g_j(z) dz + b_j^n M_j^n f_j^n + b_j^n E_j f_j^e + b_j^a M_j^a f_j^a + (1-\beta) H_j \frac{w_{hj}}{r_j} + (1-\beta) L_j \frac{w_{lj}}{r_j} = B_j^s,$$

where the left-hand side gives demand for buildings from firms and workers. Finally, labor markets for high- and low-skilled workers clear in each city,

$$\pi_{lj} L = M_j \int_z l_j(z) q(z) g(z) dz + l_j^n M_j^n f_j^n + l_j^n E_j f_j^e + l_j^a M_j^a f_j^a$$

$$\pi_{Hj} H = M_j \int_z h_j(z) q(z) g(z) dz + h_j^n M_j^n f_j^n + h_j^n E_j f_j^e + h_j^a M_j^a f_j^a.$$

3.7 Discussion

In this section, I discuss the relationship between the motivating facts documented in Section 2 and the model.

Technology adoption Section 2.3 documents that firms located in big cities are using ICT more intensively, consistent with the mechanism in the model in which firms in big cities are more likely to adopt the new technology. Equation (6) shows how the adoption threshold changes with city-level equilibrium variables and model parameters.

There are three key terms that determine the adoption threshold. First, the adoption threshold falls with the market size, Y_j. Firms' sales increase with the market size, making them more willing to cover the fixed costs associated with adopting the technology. Second, the adoption threshold is falling with the productivity difference between the old technology, $\left(\frac{\psi_j^n}{C_j^n}\right)$, and the new technology, $\left(\frac{\psi_j^a}{C_j^a}\right)$. Specifically, $\frac{\psi_j^k}{C_j^k}$ is the Hicks-neutral productivity divided by the cost index of using the technology. Firms will not adopt if the Hicks-neutral productivity advantage of the new technology is small or if the bundle of inputs used in the new technology is too expensive. These terms are raised to the $\sigma - 1$, which determines the extent to which the productivity advantages impact the firm's variable profits in the CES framework. Finally, the adoption threshold increases with the difference between the fixed cost of the new technology and old technology, $C_j^a f_j^a - C_j^n f_j^n$.

Importantly, all of these terms will be determined in general equilibrium and depend on the calibration of the city-level fundamentals. Section 5.1 shows the city-level fundamentals that are quantitatively most important for driving the differences in adoption. First is γ_j, which governs the weight on high-skilled versus low-skilled labor, impacting the cost indices, C_j^k. Second is the relative amenities for high-skilled workers versus low-skilled workers. This will impact the relative supply and wages of high-skilled workers. If the skill premium is low, the cost index for the new technology relative to the old technology, C_j^a/C_j^n, will be low. Third is the level of amenities for both low and high-skilled workers, which governs the market size Y_j.

When the initial steady state of the model is calibrated to match the features of the data in 1980, all of these terms will be correlated with population, meaning that the adoption threshold will be lower in big cities. Assuming that adopters of the new technology use ICT more intensively, there will be a tight link in the model between city size and the share of firms that adopt the new technology, matching the correlation between ICT intensity and city size in the data.

Selection In Section 2.2, I show evidence that establishment entry and exit rates are higher in big cities than in small cities—though this was not the case in 1980. The increase in exit rates in big cities is particularly pronounced for young firms, and I argue that this is consistent with selection becoming tougher in big cities relative to small cities.

In the model, Equation (13) shows that there is a direct mapping between the city-level exit threshold, s_{xj}, and the city-level start-up and exit rates. All else equal, an increase in the city-level exit threshold, or an increase in selection, will lead to an increase in firm entry and exit rates, consistent with the patterns in the data in Section 2. Thus, through the lens of the model, differences in selection across cities can be identified by examining the patterns of business dynamism. This approach is consistent with the work of Arkolakis (2016), Luttmer (2007), and Sampson (2015), who use similar models of firm dynamics to look at the effect of selection on aggregate productivity but do not look at the impact of selection across markets.

Selection and Adoption In the model, there is a close relationship between selection and adoption. This connection can be seen analytically in the model in three ways.

First, the link between selection and adoption can be seen by examining the exit threshold in Equation (7). The adoption threshold s_a shows up directly in the exit threshold, as do many of the same terms that govern the adoption threshold (Equation 6), such as market size and the productivity advantage of the new technology. When determining the exit threshold, firms consider the option value of remaining in the market, which accounts for potential profits they might earn after a series of positive productivity shocks. The benefit of adoption is reflected in the firm's discounted expected profits, resulting in a lower exit threshold.

Second, using the equations for the mass of adopters and non-adopters (Equations 11 and 12), the share of firms adopting can be written,

$$\mu^a = \frac{M^a}{M^a + M^n} = \frac{(e^{\delta s_e} - e^{\delta s_x})e^{-\delta s_a}}{(s_e - s_x)\delta}. \tag{15}$$

The share of firms adopting decreases with the adoption threshold; the higher the adoption threshold, the fewer firms adopt. However, it also increases with the exit threshold.[12] A higher exit threshold makes it more likely for non-adopters to exit the market, thereby increasing the share of firms that are adopters. Thus, as selection becomes tougher, it amplifies the share of firms in the market that are adopting the new technology, increasing the extent of skill-biased technical change. Section 5.2 shows that this is a quantitatively important channel for driving the divergence in the extent of skill-biased technical change between big and small cities.

Third, it is worth emphasizing that both the adoption and exit thresholds are determined in general equilibrium, and they will impact each other through general equilibrium effects on prices. For instance, the wages of high-skilled workers will increase with adoption. As wages increase, it will impact the profits and fixed costs of non-adopting firms, thereby influencing the exit threshold.

In Section 5.2, I provide a quantitative analysis of the extent to which selection amplifies the degree of skill-biased technical change within a city.

4 Quantitative Analysis

In this section, I take the model to the data to test whether the uneven diffusion of a skill-biased technology can match the key features of the Great Divergence. I first calibrate the parameters of the initial 1980 steady state, assuming that the fixed cost of the new technology is so high as to be prohibitive and no firm adopts. Then, in Section 4.2, I calibrate the parameters of the new technology. I emphasize that I do not target any cross-sectional moments across cities in this exercise, and, with the exception of the aggregate quantities of high- and low-skilled workers, I hold all other parameters constant at their 1980 values. I then test whether introducing the new technology into the economy as it was in 1980, holding all else constant, can quantitatively match the features of the Great Divergence. I show the results of this exercise in Section 4.3.

[12] The derivative of the adoption share with respect to the exit threshold is $\frac{\partial \mu^a}{\partial s_x} = \left[e^{\delta s_e} - e^{\delta s_x}\left(\delta(s_e - s_x) + 1\right)\right]\left[\frac{e^{-\delta s_a}}{\delta(s_e - s_x)^2}\right]$, which is greater than zero meaning the adoption share increases with the exit threshold. The second term in brackets is greater than zero since $\delta > 0$ and $s_e - s_x > 0$. The first term is greater than zero if $e^{\delta s_e} > e^{\delta s_x}\left(\delta(s_e - s_x) + 1\right)$ (taking logs of both sides and rearranging gives: $\delta(s_e - s_x) > log\left(1 + \delta(s_e - s_x)\right)$), which is always true for $s_e - s_x > 0$.

Table IV: Calibration of Initial Steady State

Panel A: Aggregate Parameters

	A. Aggregates		
H	Mass of high skill	0.28	data
L	Mass of low skill	1.0	Normalization
	B. Elasticities		
$1-\beta$	Housing share of income	0.4	Monte et al. (2018)
σ	Elasticity of substitution σ	6.5	Broda and Weinstein (2006)
$1-\beta_L$	Land cost share of production	0.26	Albouy (2016)
ϵ	Elasticity of sub. L & H	1.62	Autor et al. (2008)
ν	Shape parameter of Fréchet distribution	1.5	Fajgelbaum et al. (2019)
	C. Firm productivity process		
$\tilde{\mu}$	Persistence of Brownian motion for s	-0.12	Luttmer (2007)
$\tilde{\psi}$	SD of Brownian motion for s	0.43	Luttmer (2007)

Panel B: City-Level Fundamentals

Parameter	Description of parameter	Description of target	Target
γ_j^n	Initial skill intensity	Skill premium	$\frac{w_{j,h}}{w_{1,l}}$
ψ_j^n	Hicks-neutral productivity	City-size premium	$\frac{w_{j,l}}{w_{j,l}}$
A_{hj}	Amenities, high-skilled workers	Skill intensity	$\frac{H_j}{L_j}$
A_{lj}	Amenities, low-skilled workers	Relative city size	$\frac{H_j+L_j}{H_1+L_1}$
f_j^e	Entry cost	Establishment start-up rate	$\frac{E_j}{M_j}$
f_j^n	Fixed cost	Establishment size	$\frac{H_j+L_j}{M_j}$
ϕ_j	Productivity of building sector	Rent of city j	$\frac{r_j}{w_{1,l}}$
η_j	Elasticity of building sector	Elasticity of building supply	Saiz (2010)
	Normalizations		
$\gamma_{1,l}, A_{h1}, A_{l1}$			1

4.1 Calibration of the initial steady state

To calibrate the model, I first aggregate cities into 30 bins based on 1980 city size, each containing approximately 3.3 percent of the total population.[13] Binning by city size has the advantage that I can calibrate and solve my model with 30 bins, which would be computationally infeasible with the complete set of cities. Binning does not change any of the empirical facts.

To calibrate the initial steady state of the model, I first calibrate a set of aggregate parameters that are constant across cities. Some of these parameters are common values taken from the literature and others are calibrated to match moments in the data. Next, I recover the city-level fundamentals that 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 consistent with the data in 1980 being a steady-state equilibrium of the model.

Panel A of Table IV gives the aggregate parameters and the corresponding moments or sources I use to calibrate them. The mass of low-skilled workers is normalized to 1, and the mass of high-skilled workers, 0.28, is calculated using the aggregate quantity of workers with four or more years of college in 1980 relative to the mass of low-skilled workers. Therefore, the total mass of workers in the economy is 1.28. The share of income spent on rent, $1-\beta$, is set to 0.4 following Monte et al. (2018), and $1-\beta_L$ gives the land cost share of production and is set at .26 as estimated by Albouy (2016). The elasticity of substitution between varieties is also a standard parameter from the literature. I set it to 6.5 in line with the estimates from Broda and Weinstein (2006).[14] I take the shape parameter of the Fréchet distribution, which governs the migration elasticity, from Fajgelbaum et al. (2019) and set it to 1.5. In Appendix C.2, I show that the results are robust to using alternative values for σ and ν. The elasticity of substitution between high and low-skilled labor is 1.62 following Autor et al. (2008). Finally, I set the drift and standard deviation of the Brownian motion for s to $-.12$ and $.43$ following Luttmer (2007).

Next, I calibrate the parameters that vary across cities. Let

$$\mathbb{P}^c_{1980} = \{\gamma^n_c, \psi^n_c, A^h_c, A^l_c, f^e_c, f^n_c, \phi_c, \eta_c\} \tag{16}$$

be the vector of city-level fundamentals that is, the weight on high-skilled labor, γ^n, the Hicks-neutral productivity term, ψ^n, high- and low-skilled amenities, A_h and A_l, entry costs, f^e, fixed costs, f^n, the productivity of the building sector, ϕ, and the elasticity of building supply, η. Panel B of Table IV shows the moments from the data used to identify each of the fundamentals. In 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.

[13] I do not divide cities between bins, so some bins will have more or less than 3.3 percent of the population.

[14] There is some debate about this number in the literature. The elasticity of substitution will control the love for variety and, therefore, the agglomeration force in the model. Monte et al. (2018) use a value of 4 while Behrens et al. (2014) estimate it to be closer to 13.5. I use a value of 6.5, a common value in the literature estimated by Broda and Weinstein (2006).

Table V: Calibration of the new technology

Parameter	Description	Value	Target	Data	Model
Γ^{γ}	Skill-bias of new technology	1.90	Aggregate growth in high-skilled wages	28.70	28.71
Γ^{ψ}	Productivity advantage of new technology	1.24	Aggregate growth in low-skilled wages	-3.03	-3.01
Γ^{f}	Additional fixed cost	5.67	Aggregate growth in average establishment size	10.10	10.09
H	Aggregate quantity of high-skilled labor	0.79	Growth in supply of high-skilled labor	0.79	0.79
L	Aggregate quantity of low-skilled labor	1.10	Growth in supply of low-skilled labor	1.10	1.10

4.2 Calibration of the new technology

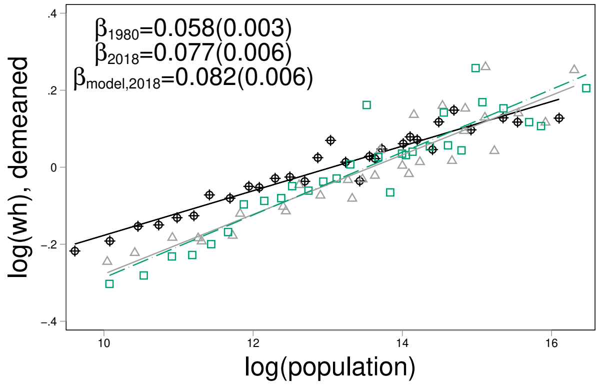
Next, starting from the economy with no adoption, I introduce a new skill-biased technology. There are three parameters of the new technology that need to be calibrated: Γ^{γ}, which scales the weight on high-skilled labor; Γ^{ψ}, which scales the absolute productivity advantage of the new technology; and Γ^{f}, which scales up the fixed cost of production, f_j^n. I calibrate the parameters to match moments on the aggregate growth in high- and low-skilled wages and average establishment size between 1980 and 2018. In addition, Katz and Murphy (1992) show that to understand the extent of SBTC, it is necessary 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 2018.

I choose the three parameters of the new technology to minimize the difference between the model and data in the aggregate growth of high-skilled wages, low-skilled wages, and establishment size, $\min_{\{\Gamma_{\gamma},\Gamma_{\psi},\Gamma_{fc}\}} X'X$, where X is a vector of the difference between moments and targets. I emphasize that no cross-sectional information is used to calibrate the new technology. I only target aggregate changes in wages and establishment size. The five parameters and the corresponding targets are listed in Table V. The model can match each of the targets perfectly.

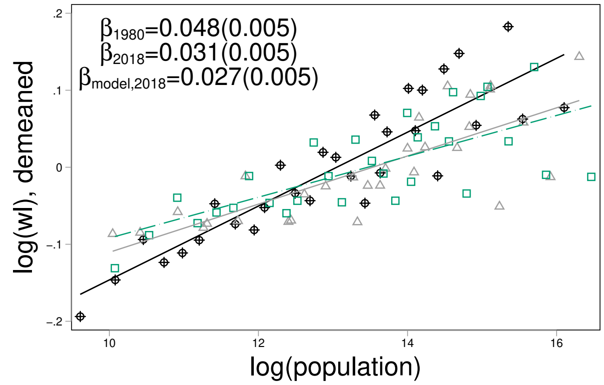
The baseline calibration assumes that all aggregate wage growth is attributed to the adoption of the new technology. Section 4.4 shows how the results change under two alternative assumptions for calibrating the new technology. Appendix C.2 presents additional robustness checks showing how the results change when varying the parameters of the new technology within a 5% range of their calibrated values, that the main results are robust to denominating the fixed costs in terms of the final good (rather than labor and land), and to using alternative values for σ and ν.

Figure II: Descriptive facts: data vs. model

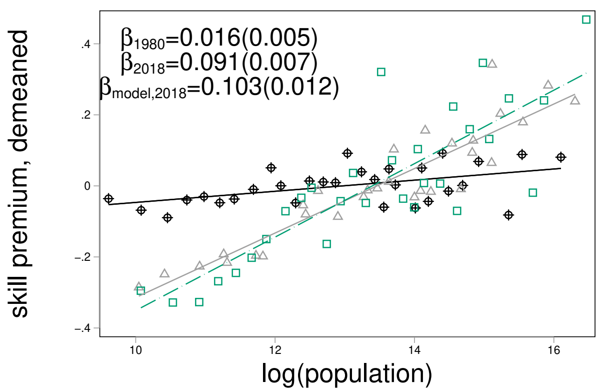
(a) City-size wage premium, high skill



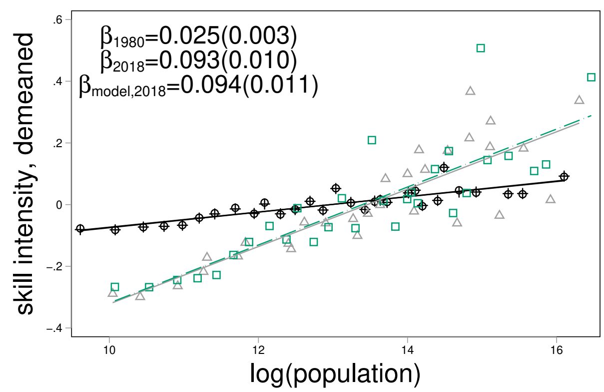
(b) City-size wage premium, low skill



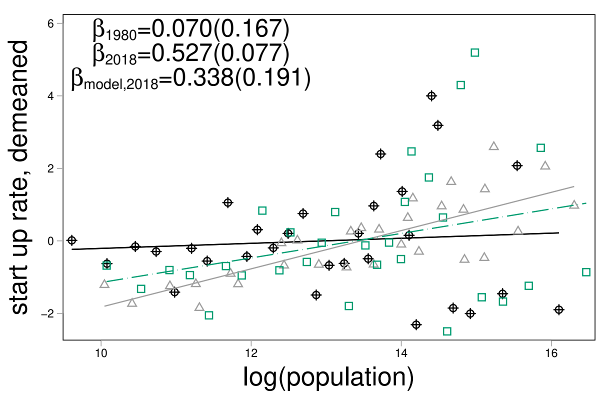
(c) Skill premium, W_h/W_l



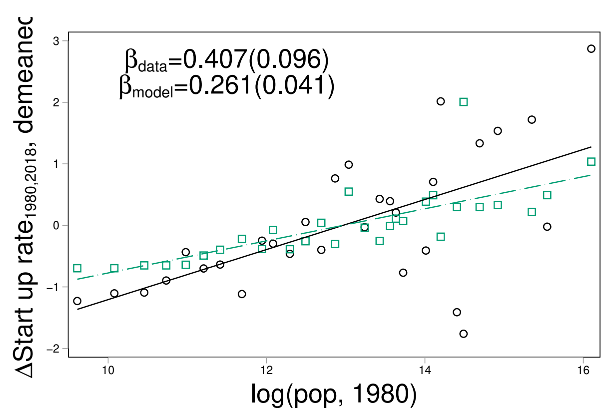
(d) Relative supply of skilled workers, H/L



(e) Establishment start-up rate



(f) Change in the establishment start-up rate



○ 1980, data + 1980, model △ 2018, data □ 2018, model

Notes: Source: Wages and skill intensity are from the 1980 Decennial Census and 2018 ACS. Dynamism is from the Business Dynamic Statistics. Population is working-age population (ages 20-64) from the Intercensal Population Estimates. Additional data are model output. Figure displays the relationship between wages, skill intensity, 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. Grey gives the 2018 data and green the 2018 steady state in the model. The unit of observation is one of 30 city-size categories. All variables are demeaned by year.

4.3 Model predictions versus the data

Figure II shows that introducing the skill-biased technology can reproduce the data patterns described in Section 2. In the initial steady state, the model is calibrated to perfectly match the empirical relationships between wages, the skill premium, skill intensity, and dynamism with city size in 1980. I then show these relationships in the 2018 data and the second steady state of the model. In the second steady state, the only change is to the aggregate quantities of high- and low-skilled labor and the availability of the new technology.

The model does a good job of predicting the cross-sectional changes for both high- and low-skilled wages shown in Panels (a) and (b). The city-size wage premium for high-skilled workers increased from 0.058 percent to 0.077 percent in the data and from 0.058 to 0.082 percent in the model. Similarly, the city-size wage premium for low-skilled workers decreases from 0.048 to 0.031 percent in the data and from 0.048 to 0.027 percent in the model. The model successfully matches the divergence by skill group; the city-size wage premium rises for high-skilled workers while it falls for low-skilled workers. The changing relationship between the skill premium and city size is shown in Panel (c). In 1980, the semi-elasticity between the skill premium and city size was .016 percentage points. By 2018, the semi-elasticity had increased to .091 percentage points, while in the model, the semi-elasticity increased from .016 to .103 percentage points. Thus, the model explains all of the increasing cross-sectional relationship between the skill-premium and city size.

The model also matches the changing relationship of skill intensity with city size (Panel d). In 1980, the semi-elasticity between skill intensity and city size was 0.025 percentage points. By 2018, it had increased to 0.093 percentage points. The model matches this fact well. The relationship between skill intensity and city size grows from 0.025 in the initial steady state to 0.094 in the final steady state. Thus, the model explains all of the increasing cross-sectional relationship between skill-intensity and city size.

The changing relationship between the establishment start-up rate and city size is shown in Panel (e). In the data, the semi-elasticity between the establishment start-up rate and city size increased from 0.070 to 0.527, while in the model it increased from 0.070 to 0.338. Thus, the model predicts 59 percent of the changing slope. Finally, Panel (f) shows that the model matches the fact that the decline in dynamism was larger in cities that were small in 1980.

Table VI: Robustness to Calibration of the New Technology

Share of cross-sectional change explained by model in:			
	Skill-premium	Skill-intensity	Start-up rate
Baseline	116.8	101.4	58.6
Target 50% of wage growth	138.9	116.5	36.4
Allow change in old technology	63.3	81.0	55.1

Notes: Table shows the share of the cross-sectional changes explained by the model. The first row gives the results for the baseline calibration. Rows 2-3 show the robustness of the main results to alternative calibrations.

4.4 Robustness to calibration of the new technology

In this section, I examine how the results would change under alternative assumptions for calibrating the new technology. In the baseline model, the three parameters of the new technology are calibrated to match aggregate growth in high-skilled wages, low-skilled wages, and average establishment size between 1980 and 2018. This approach makes a strong assumption, attributing all wage changes over the period to the introduction of the new technology.

I provide two robustness exercises that relax this assumption. First, I calibrate the parameters of the new technology to match only 50% of the aggregate wage changes.[15] Second, I allow for changes in the old (non-adopter) technology. Specifically, in addition to allowing firms to adopt the new technology, I also scale up the Hicks-neutral parameter of the old technology, ψ_j^n, and the weight on high-skilled labor in the old technology, γ_j^n. These parameters are scaled so that 50% of the aggregate wage growth for high- and low-skilled workers is attributed to changes in the old technology rather than to adoption. I then calibrate the parameters of the new technology to match the remaining growth in aggregate wages and establishment size.

Table VI shows how the baseline results change under the alternative calibrations. Specifically, I report the share of the cross-sectional changes in the skill premium, skill intensity, and the start-up rate explained by the introduction of the new technology. For comparison, the baseline results are reprinted in Row (1) (these are the same results that are shown graphically in Figure II).

When targeting only 50% of wage growth, the quantitative results remain similar to those in the baseline model (see Table C.I, Row 2). In this calibration, the extent to which the new technology increases the relative demand for skilled workers, Γ^{γ}, is comparable to the baseline. However, the productivity advantage of the new technology, Γ^{ψ}, is significantly smaller. In the baseline model, the new technology is 24% more productive than the old technology, while in the alternative calibration, it is 7% more productive. As a result, wage growth for skilled workers is lower, but the technology is still adopted more in big cities than in small cities, which drives cross-sectional results.

When allowing for technical change in the old technology, the results remain qualitatively similar, but the new technology accounts for a smaller share of the cross-sectional changes in both the skill premium and skill intensity (see Table VI, Row 3). This is due to the increasing skill bias of old technology. As a result, to match the aggregate growth in high-skilled wages, the new technology increases the skill-bias much less compared to the baseline model. Nonetheless, adoption rates are still over twice as high in the largest city compared to the smallest city, which drives the qualitatively similar cross-sectional changes observed in both the skill premium and skill intensity.

[15] While there is no definitive number for how much of the wage changes should be attributed to the diffusion of new skill-biased technologies versus other factors, DiNardo et al. (1996) show that about 50% of the change in between-group wage inequality can be attributed to the combination of supply and demand factors and unexplained changes in technology (see their Table IV).

5 The Drivers of the Great Divergence

A key result in the quantitative exercise is that technology adoption is higher in big cities than in small cities. This happens despite firms in all cities having equal access to the new technology. Growing differences across cities in selection then amplify the initial differences in technology adoption. This section aims to determine what factors drive the growing gaps between cities. Section 5.1 discusses which sources of city-level heterogeneity are most important in causing the differences in adoption rates. Section 5.2 presents an accounting exercise that decomposes city-level productivity and skill intensity growth into that arising from technology adoption versus that arising from selection. Section 5.3 shows that the patterns of adoption across cities align with patterns of ICT spending in the data.

5.1 The drivers of technology adoption

In this section, I discuss the forces that drive the differences in technology adoption across cities. The model defines cities by the eight city-level parameters given in Equation (16). To determine how each of these city-level parameters affects adoption in general equilibrium, I ask what would happen to the cross-sectional pattern of adoption if these sources of city-level heterogeneity were removed. In a series of counterfactuals, I set the parameters to the average across cities and re-solve the model in general equilibrium. In addition, for amenities, I shut down differences in relative amenities for high- and low-skilled workers. Figure III shows the effect on adoption in the counterfactuals versus the baseline model for the three main drivers of heterogeneity. Panel (d) shows that adoption rates are equalized across cities when all three of the main drivers are equalized. The minimal impact of equalizing the additional model parameters is shown in Appendix D.3.

Three city-level parameters generate the most significant differences in adoption rates across cities. The first is the weight on high-skilled labor in the city's initial technology, given by γ_j. Panel (a) of Figure III shows the effect of equalizing γ_j across cities. When the γ is equalized, adoption rates are equalized: the semi-elasticity between adoption and the initial population declines from 0.655 to 0.357. The initial weight on high-skilled labor drives significant differences in adoption rates across cities because the cost of adopting is smaller for firms that start with a higher gamma. As γ_j increases, the difference in the cost index between the old and new technology is smaller.

The second force driving the heterogeneity in adoption rates across cities is the role of relative wages. This can be seen in the response of adoption to shutting down differences in relative amenities for high- and low-skilled workers in panel (b). Specifically, I set the amenity for high-skilled workers to the level of the amenity for low-skilled workers, $A_{Hj} = A_{Lj}$, so that the relative amenity, A_H/A_L, is one for all cities. In the baseline calibration, the relative amenity for high-skilled workers is increasing in city size. When they are equalized, high-skilled workers are less likely to choose to live in big cities. To attract them to the big cities, wages for high-skilled workers increase, decreasing the return to adoption. The semi-elasticity between adoption and initial city size decreases from 0.655 to 0.443. This exercise demonstrates that amenities are essential in driving the differences in technology adoption across cities. If amenities responded endogenously to the presence of high-skilled workers, as in Diamond (2016), wages for high-skilled workers would fall. This would increase the returns to adoption and amplify the growing technological differences across cities.

The third force is the market-size effect or an increase in the firm's scale through Y_j. The market-size effect can best be seen in the counterfactual that shuts down the heterogeneity in amenities for both low- and high-skilled workers, A_{Hj} and A_{Lj}, in panel (c). In the initial equilibrium, both high- and low-skilled amenities are increasing in city size. Equalizing amenities equalizes population across cities, and adoption equalizes as population equalizes. The semi-elasticity between population and adoption falls from 0.655 to 0.544.

Finally, panel (d) shows what happens if you shut down all three sources of heterogeneity in the model by setting γ_j, A_{Hj}, and A_{Lj} to their average value. Without these three channels, there is no longer a significant positive correlation between adoption and the initial city size.

5.2 The growing productivity differences between cities: selection versus adoption

There are two sources of the growing productivity differences between cities: those arising from selection and those arising from adoption. In this section, I perform an accounting exercise to decompose the change in productivity coming from each.

Productivity gains from adoption accrue as firms choose to use the technology with the higher Hicks-neutral productivity term, ψ_j^a, versus that used by non-adopters, ψ_j^n. Productivity gains from selection occur when the distribution's lower bound, z_{xjt}, shifts up. In a static Melitz (2003) model, the only impact of selection is to truncate the left side of the productivity distribution and scale the remainder of the firm distribution proportionally. In contrast, in a model with firm dynamics, selection changes the shape of the entire productivity distribution, given by Equation (10).

The TFP of a city j is given by the market-share weighted average of the Hicks-neutral productivity term

$$Z_{j,t} = \int_z \psi_j^k(z) z m_{j,t}(z) g_{j,t} dz$$

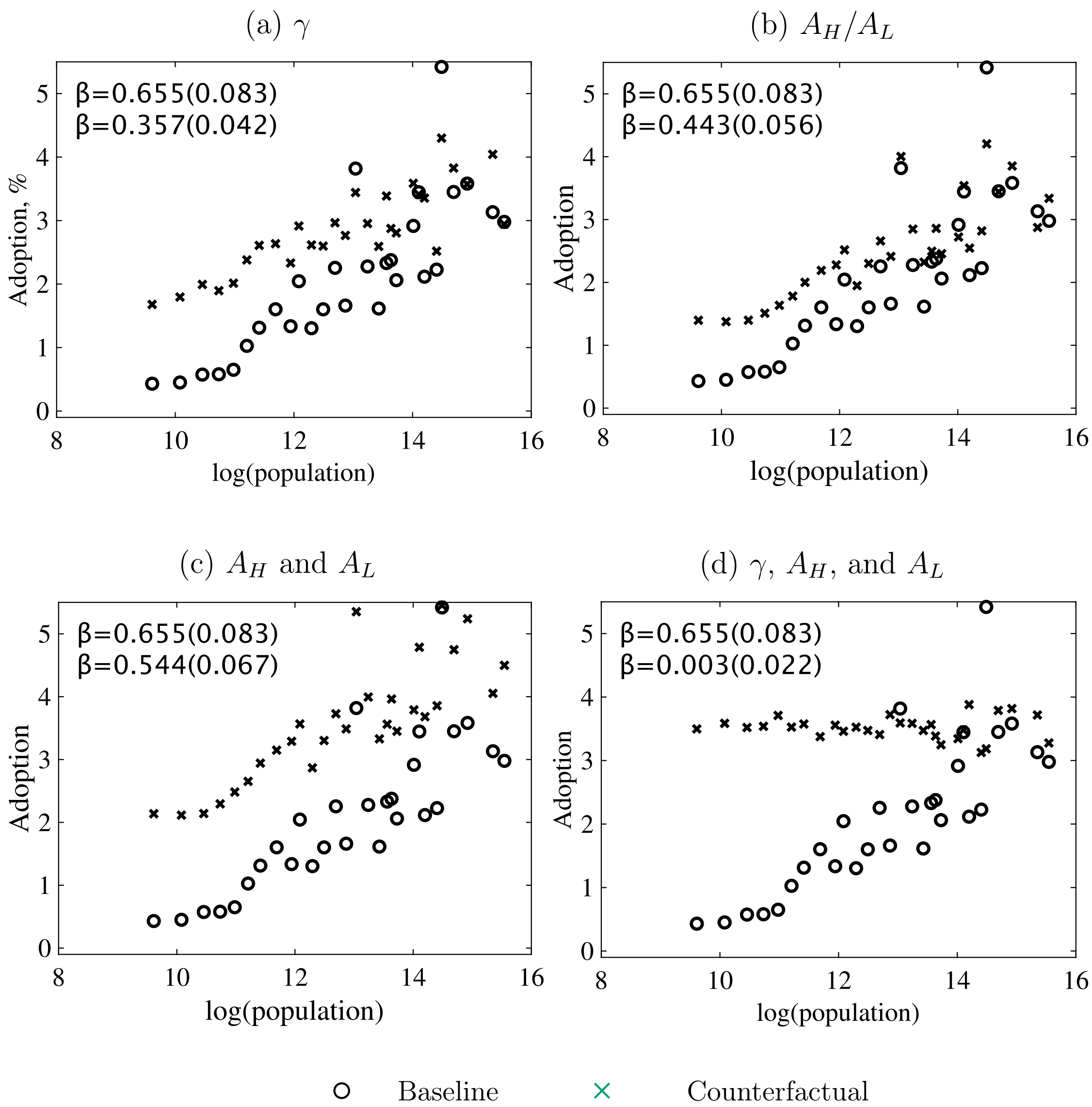
where $m_{j,t}(z) = p_j(z)q_j(z)/Y_j$ is the market share of a firm with productivity z. Then, the growth in TFP between the initial and final steady state,

$$\frac{Z_{j,2018} - Z_{j,1980}}{Z_{j,1980}} = \frac{1}{Z_{j,1980}} \left[\underbrace{\int_{z_{x,2018}}^{\infty} \psi_j^k(z) z m_{j,2018}(z) g_{j2018}(z) dz}_{=Z_{2018}} - \underbrace{\int_{z_{x,1980}}^{\infty} \psi_j^n z m_{j,1980}(z) g_{j1980}(z) dz}_{=Z_{1980}} \right],$$

can be decomposed into a component coming from selection and a component coming from adoption by adding and subtracting the term

$$\int_{z_a}^{\infty} \psi_j^n z m_{j,2018}(z) g_{j,2018}(z) dz + \left(\psi_j^a - \psi_j^n\right) \int_{z_a}^{\infty} z m_{j,1980}(z) g_{j,1980}(z)$$

Figure III: Counterfactual adoption rates



Note: Figure displays counterfactual adoption rates when parameters defining a city are equalized.

and rearranging,

$$\frac{Z_{j,2018}-Z_{j,1980}}{Z_{j,1980}} = \underbrace{\frac{\left(\psi_j^a-\psi_j^n\right)}{Z_{j,1980}}\int_{z_a}^{\infty} z m_{j,1980}(z) g_{j,1980}(z)}_{\text{adoption}}$$
$$+\underbrace{\frac{1}{Z_{j,1980}}\left[\int_{z_{x,2018}}^{\infty}\psi_j^n z m_{j,2018}(z) g_{j2018}(z)dz-\int_{z_{x,1980}}^{\infty}\psi_j^n z m_{j,1980}(z) g_{j1980}(z)dz\right]}_{\text{selection}}$$
$$\underbrace{+\frac{\left(\psi_j^a-\psi_j^n\right)}{Z_{j,1980}}\int_{z_a}^{\infty} z\left(m_{j,2018}(z)g_{j,2018}(z)-m_{j,1980}(z)g_{j,1980}(z)\right)dz}_{\text{selection}}. \tag{17}$$

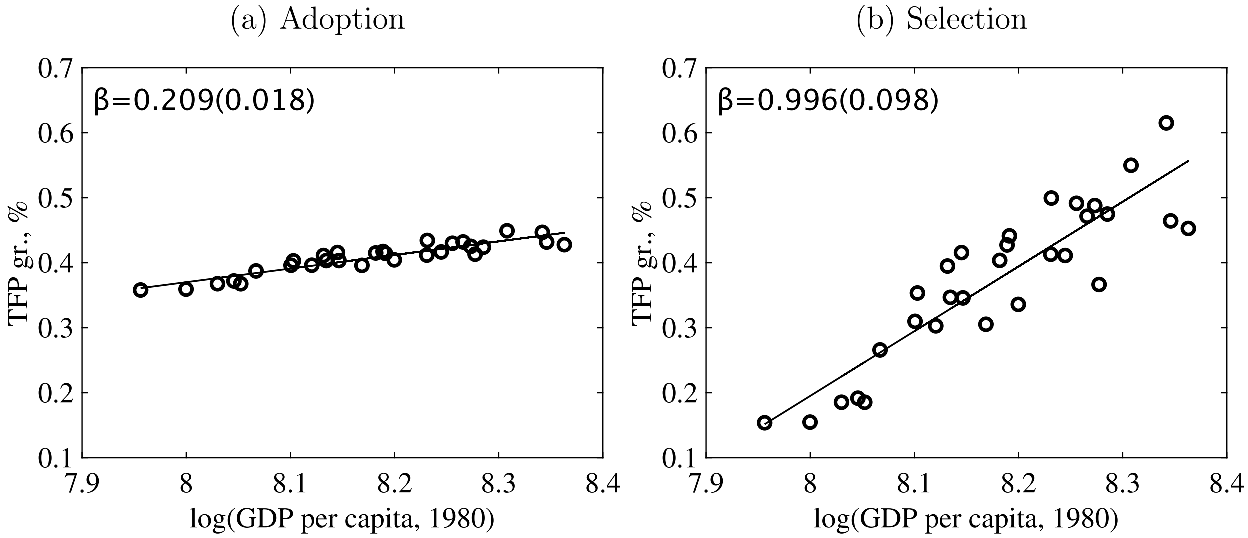
The first term is the growth in TFP attributed to adopting the new technology. It gives the change in the average Hicks-neutral productivity term if one were to keep the productivity distribution and the market shares as they were in 1980. The second two terms give the growth in TFP attributed to selection. The term in brackets is the growth in TFP that would have occurred if no firm had adopted, but there were changes in the distribution of firm TFP and market shares through selection. The final term is the growth from the additional mass of firms above the adoption threshold due to the increase in selection.

Figure IV shows annualized TFP growth arising from adoption in Panel (a) and selection in Panel (b) plotted against initial GDP per capita. The slope gives the β–convergence measure, with a β above 0 indicating that there is divergence, meaning that cities which started off with high TFP have grown faster since 1980 (Baumol, 1986; Giannone, 2017).

Technology adoption and selection are of similar importance for explaining within-city TFP growth, both contributing an average of around 0.4% of TFP growth per year. Further, both selection and adoption contribute to the divergence of TFP with β–convergence measures greater than 0. However, selection has been much more important for driving divergence than adoption. The β–convergence measure for selection is 0.996, meaning that a city with 1% higher GDP per capita in 1980 experienced, on average, TFP growth due to selection that was 0.0099 percentage points higher per year. On the other hand, TFP growth due to adoption was only 0.0021 percentage points higher per year in cities with 1% higher GDP per capita. This is consistent with the empirical findings in Section 2.2, which show that selection has become tougher in big cities relative to small cities. As a result, the share of TFP growth that can be attributed to selection is increasing in city size. In city 30, 56 percent of TFP growth can be attributed to selection versus 29 percent in city 1. On average, 46 percent of TFP growth can be attributed to selection rather than adoption.

Next, I examine the role of selection versus adoption in generating changes in the relative demand for high-skilled labor. An analogous version of Equation (17) can be written using the weight on skilled workers of adopters, γ_j^a, versus non-adopters, γ_j^n, instead of productivity, ψ_j^τ. This gives the growth in skill intensity accounted for by adoption versus the growth in skill intensity accounted

Figure IV: TFP growth from adoption versus selection



Note: Figure displays the results of a decomposition of annualized TFP growth into growth attributed to adoption and growth attributed to selection.

for by selection.

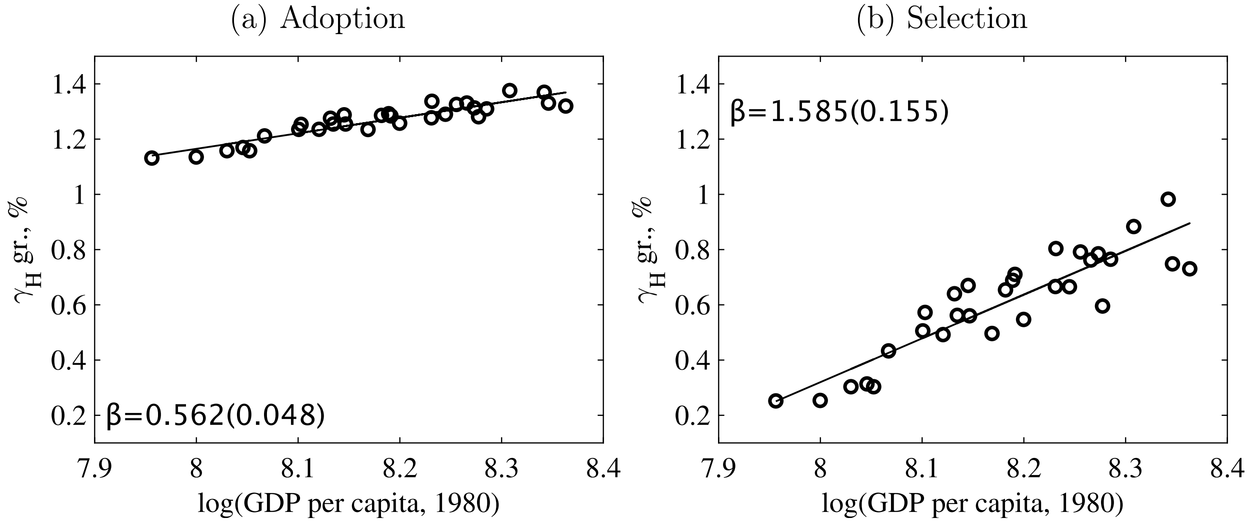
The results of this exercise are shown in Figure V. Panel (a) shows the annualized growth in skill intensity accounted for by adoption, and Panel (b) shows the annualized growth accounted for by selection plotted against initial log GDP per capita. Across all cities, there is a substantial increase in the average skill intensity term. The direct effect of adoption is much more important overall. However, as with TFP growth, the impact of selection is more important for the divergence across cities. The β-convergence measure is 1.59 for selection versus 0.56 for adoption. In city 30, 37 percent of the growth in skill intensity is accounted for by selection rather than adoption, versus 16 percent in city 1. On average across cities, 29 percent of growth in skill intensity can be attributed to selection.

5.3 Model validation using data on ICT spending

As discussed in Section 5.1, there are three main channels that generate the heterogeneity in adoption rates across cities: market size, initial weight on high-skilled workers, and relative wages driven by amenities. 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 firm investment decisions in new technologies and thus provides suggestive evidence for the model's mechanisms.

The elasticity between ICT spending per employee today and 1980 population is 5.58 percent, meaning that ICT spending per employee is about 3.87 percent higher today in markets that were

Figure V: Growth in skill-intensity, γ_j, from adoption versus selection



Note: Figure displays the results of a decomposition of annualized growth in skill intensity into growth attributed to adoption and growth attributed to selection.

twice as large in 1980. Similarly, in the model, the share of adopters is 0.454 percentage points higher in cities that were initially twice as large. This suggests that market size is important in a firm's 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 (γ_j) and the amenities for high-skilled labor. Firms that are in locations that are already good at using high-skilled labor will be more likely to adopt the new technology. In such locations, we expect to see a high initial skill supply 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 (γ_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

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

	Data: $log(\frac{ict}{emp})$			Model: share of adopters		
$log(pop_{1980})$	0.0558***			0.655***		
	(0.00708)			(0.0799)		
$\frac{H_{1980}}{L_{1980}}$		2.205***			26.38***	
		(0.427)			(1.085)	
$\frac{w_{h1980}}{w_{l1980}}$			0.313***			11.91***
			(0.105)			(1.178)
Observations	224000	224000	224000	30	30	30
R^2	0.351	0.348	0.347	0.706	0.955	0.236

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

Source: The left panel uses ICT data from ACES merged into the LBD for information on the firm's location. Population is working-age population from the Intercensal Population Estimates. The table displays the relationship between the 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, firm age, and year fixed effects. The regressions also control for the log of firm employment and number of establishments. 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.

skill intensity. Cities with a 1 percentage point higher skill intensity in 1980 have 2.205 percent more ICT spending per employee today. Similarly, in the model, cities with a 1 percentage point higher skill intensity have an adoption rate that is, on average, 0.26 percentage points higher. This provides suggestive evidence for the initial technology and labor supply mechanisms. 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 also show a positive relationship between ICT spending and initial skill premium, providing suggestive evidence for the importance of initial technology on subsequent adoption.

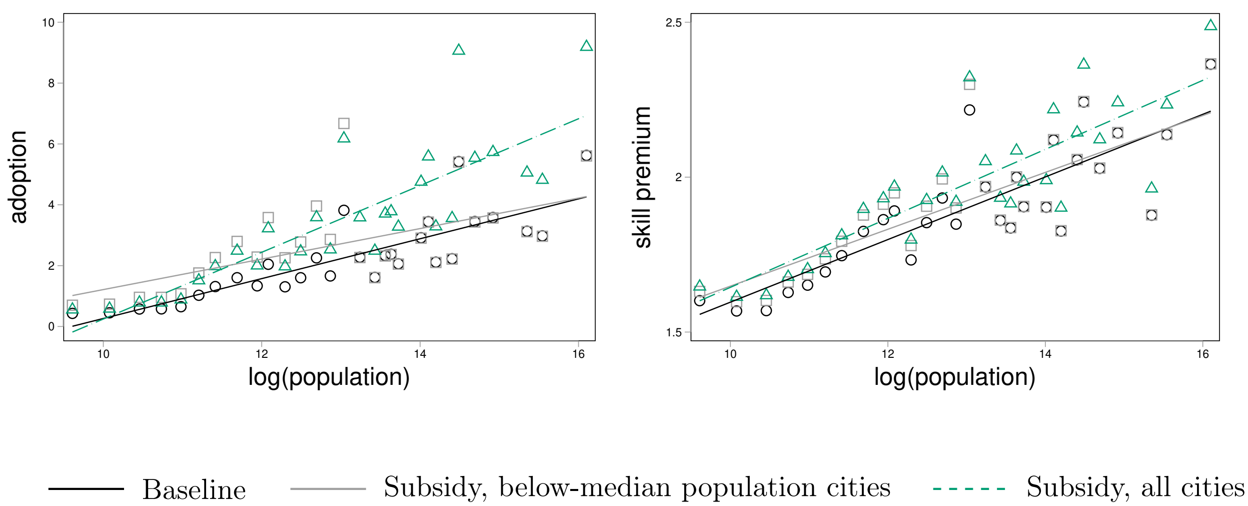
6 Policy Analysis: Subsidizing the fixed costs of adoption

This section examines several policies aimed at increasing technology adoption while reversing growing geographic inequality. I first examine a government subsidy for the fixed cost of using the new technology. I consider two versions of this policy: one where the subsidy is available everywhere and one where it is available only for firms in cities with below-median population. Second, I examine a subsidy for constructing new buildings in big, constrained cities. This is motivated by recent work (Ganong and Shoag, 2017, Herkenhoff et al., 2018, and Hsieh and Moretti, 2019) showing that housing supply constraints are a significant drag on aggregate growth. In both cases, I show that the government faces a tradeoff between increasing aggregate adoption rates and exacerbating the patterns of the Great Divergence.

Section 4.4 shows how the main results change under alternative assumptions for how to calibrate

Figure VI: Subsidizing the fixed cost of the new technology

(a) Adoption (b) Skill premium



Source: Model-generated output. Panel (a) shows the relationship between adoption and city size and Panel (b) shows the relationship between the skill-premium and city size in the counterfactuals with lower adoption costs.

the parameters of the new technology. Appendix C.3 shows that the impact of the policies is qualitatively similar under these alternative calibrations.

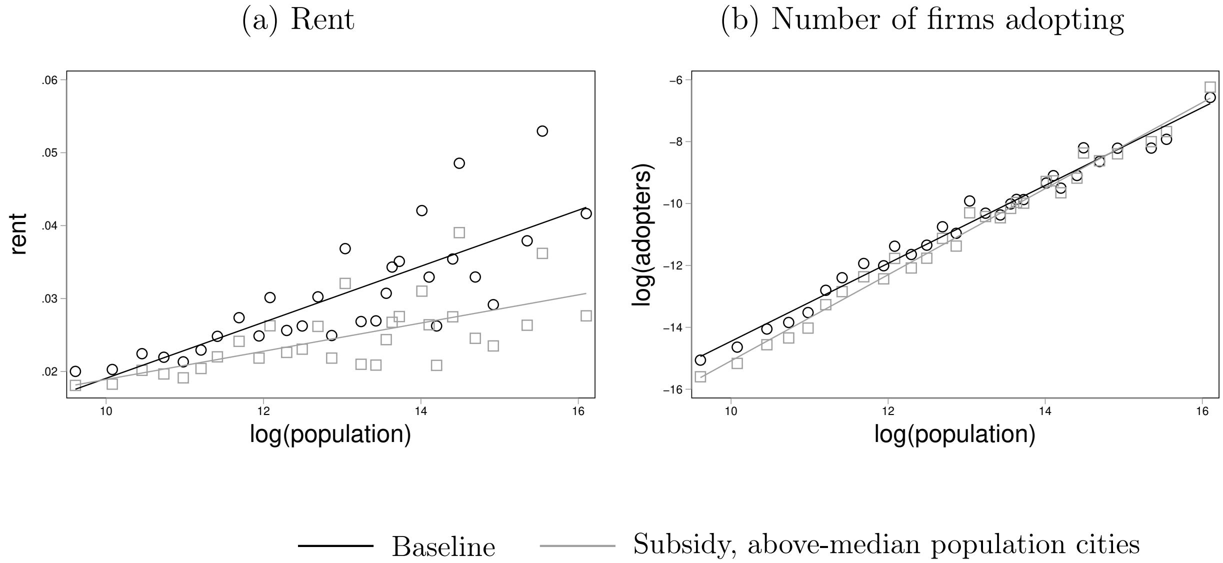
6.1 Subsidizing adoption costs

A common goal among policymakers is to encourage technology adoption, but these policies can amplify the growing differences between big and small cities. In this section, I use the model to analyze two different policy scenarios aimed at increasing technology adoption. First, I ask what would happen if the national government subsidized the fixed cost of using the new technology everywhere. Second, I ask what would happen if the federal government subsidized the fixed cost of using the new technology only in cities with below-median populations. In both cases, I assume that the subsidy is set such that the fixed cost of adoption is scaled down by 50% relative to the baseline. The subsidy is funded by the national government, which imposes a lump-sum tax on all residents regardless of where they live and their skill level. The equilibrium tax rate adjusts endogenously based on the take-up rate of the policy—the more firms that adopt, the more expensive it will be to provide the subsidy.

Figure VI shows the effect on adoption and the skill premium in the two counterfactuals.[16] The effect on adoption rates is shown in Panel (a), with the approach of subsidizing adoption only in small cities in gray and the approach of subsidizing adoption everywhere in green. Subsidizing the fixed cost of the new technology substantially increases adoption rates. This is unsurprising since

[16]A plot for skill intensity would look similar to the one for the skill premium.

Figure VII: Subsidizing housing



Source: Model-generated output. Panel (a) shows the relationship between rents and city size in the baseline model and the counterfactual with subsidized buildings in cities with above-median population. Panel (b) shows the effect of the subsidy on the number of firms who adopt the new technology.

the difference between the fixed cost of the new and old technology, $C_j^a f_j^a - C_j^n f_j^n$, is a term in the equation defining the adoption threshold, Equation (6). When the fixed cost of the new technology is much higher than the fixed cost of the old technology, the adoption threshold is higher, meaning only the most productive firms adopt. When the government subsidizes the difference, the adoption threshold shifts down, and more firms become adopters.

If one had the goal of equalizing adoption rates across cities, then lowering the cost of adoption only in small cities would be very effective. Under this policy, adoption rates for medium and small cities go up substantially, while adoption rates for big cities are essentially unchanged. However, because these cities are so small, the policy has negligible effects on the aggregate adoption rate, which goes up by only 0.1 percentage points.

In contrast, if the government aimed to increase aggregate adoption rates, the policy of subsidizing adoption everywhere would be very effective, as shown with the green line in Figure VI. Under this approach, the aggregate adoption rate increases by 2.3 percentage points. However, even though the subsidy is the same everywhere, adoption rates increase much more in big cities than in small cities, exacerbating the patterns of the Great Divergence. Panel (b) shows that the policy increases the skill premium everywhere while also amplifying the divergence in the skill premium across cities. Consequently, relative to the policy of subsidizing adoption only in small cities, adoption rates in small cities decline since the relative price of skilled labor, a key determinant of adoption as shown in Section 5.1, is now higher. These counterfactuals suggest that policies that subsidize the fixed costs of using new technologies are effective at increasing aggregate adoption rates, though at the expense of growing geographic differences in adoption, skill intensity, and the skill premium.

6.2 Subsidizing housing

Next, I consider the effect of a federal subsidy for constructing new buildings in big cities. Specifically, I assume that the national government imposes a lump sum tax of 1% of the average wage and uses it to fund additional buildings in cities with above-median population. I assume the government can turn the tax revenue into housing with the same production function as the landlord described in Section 3.5. The results of this policy for rents and the number of firms that adopt the new technology are shown in Figure VII.

The subsidy equalizes rents across cities, prompting both low- and high-income households to move to the big cities. However, the effect on adoption is ambiguous. The right panel shows that the number of firms adopting the new technology increases in the big city and falls in the small city, reflecting the shift in population. Overall, the number of adopters increases substantially; the mass of adopters rises by 15.2 percent. As adoption increases, the skill premium rises, causing adoption in small cities to fall. However, because rents do not increase as much in the big cities, the small unproductive firms find it profitable to remain in the market: selection declines, and the *share* of firms that adopt increases by only 0.07 percentage points. This underscores the role of selection in determining the share of firms that are adopting the new technology, as shown in Equation (15).

7 Conclusion

Since 1980, big and small cities have diverged on several important dimensions. The relationships between skill intensity and city size, the skill premium and city size, and business dynamism and city size all increased between 1980 and 2018. This paper examines the extent to which these facts can be jointly explained by the uneven diffusion of a new skill-biased technology.

I embed a rich model of firm dynamics into a benchmark spatial equilibrium model allowing the joint consideration of technology adoption, relative wage inequality, and business dynamism across cities. I calibrate the model to match the salient features of the 1980 data.

I 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 big city that is, ex-ante, better at using high-skilled labor and has abundant amenities for high-skilled labor. Cities with more adoption see an increase in high-skilled wages and rents. Smaller, less-productive firms that do not find it profitable to adopt the new technology exit the market, amplifying the differences in adoption rates across cities and changing the cross-sectional relationship between city size and business dynamism. Furthermore, the increase in selection puts additional downward pressure on low-skilled wages since the exiting firms have a comparative advantage in using low-skilled labor.

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 a significant share of the divergence between big and small cities. Specifically, it can account for all of the increasing relationship between the skill premium and city size, all of the increasing relationship between skill intensity and city size, and just over half of the increasing relationship between dynamism and city size. Further, assuming that adopting firms will use ICT technologies more intensively, this mechanism can rationalize the fact that firms in big cities spend more per worker on ICT investments than firms in small cities and devote a higher share of their investment budget to ICT expenses. A decomposition of TFP growth into a share from technology adoption and a share from selection suggests that selection played a significant role in amplifying the Great Divergence.

Finally, I use the model to examine the effects of policies aimed at increasing aggregate adoption by subsidizing the cost of adoption or by subsidizing the cost of construction. I show that policymakers face a trade-off: they can increase adoption rates and amplify the features of the Great Divergence, or they can work towards equalizing outcomes across cities with little effect on aggregate adoption.

Data Availability

All data and code required to replicate the results in this paper are publicly available at Zenodo: https://doi.org/10.5281/zenodo.15611657.

References

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2021): "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," *The Review of Economic Studies*, 89, 45–86.

Albouy, D. (2016): "What Are Cities Worth? Land Rents, Local Productivity, and the Total Value of Amenities," *The Review of Economics and Statistics*, 98, 477–487.

Arkolakis, C. (2016): "A Unified Theory of Firm Selection and Growth," *The Quarterly Journal of Economics*, 131, 89–155.

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., 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.

Baum-Snow, N., M. Freedman, and R. Pavan (2018): "Why Has Urban Inequality Increased?" *American Economic Journal: Applied Economics*, 10, 1–42.

Baumol, W. J. (1986): "Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show," *The American Economic Review*, 76, 1072–1085.

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.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014): "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy*, 122, 507–553.

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.

Combes, P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012): "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection," *Econometrica*, 80, 2543–2594.

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.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044.

Eckert, F. (2019): "Growing Apart: Tradable Services and the Fragmentation of the U.S. Economy," Ph.D. thesis, Yale University.

Eckert, F., S. Ganapati, and C. Walsh (2019): "Skilled Tradable Services: The Transformation of U.S. High-Skill Labor Markets," Working Paper 25, Minneapolis Federal Reserve Opportunity and Inclusive Growth Institute.

Eeckhout, 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.

Fajgelbaum, P. D., E. Morales, J. C. S. Serrato, and O. Zidar (2019): "State Taxes and Spatial Misallocation," *Review of Economic Studies*, 86, 333–376.

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.

Jiao, Y. and L. Tian (2019): "Geographic Fragmentation in a Knowledge Economy," .

Karahan, F., B. Pugsley, and A. Şahin (2019): "Demographic Orgins 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.

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.

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.

Sampson, T. (2015): " Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth *," *The Quarterly Journal of Economics*, 131, 315–380.

Stokey, N. L. (2008): *The Economics of Inaction: Stochastic Control Models with Fixed Costs*, Princeton University Press.

Syverson, C. (2004): "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 112, 1181–1222.

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.

Online Appendix

A Data Appendix

This appendix provides robustness checks and additional results for the main empirical findings in the paper. First, I present facts related to the changing relationship between the skill premium, the supply of skilled workers, and city size. These trends have become known as the Great Divergence, and I reproduce them here since they are important inputs to my quantitative analysis. 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

To document facts on the city-size wage premium and skill supply by city size, I use data from the 1980 Decennial Census and the 2018 American Community Survey (ACS) compiled by IPUMS (Ruggles et al., 2019).[17] 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.

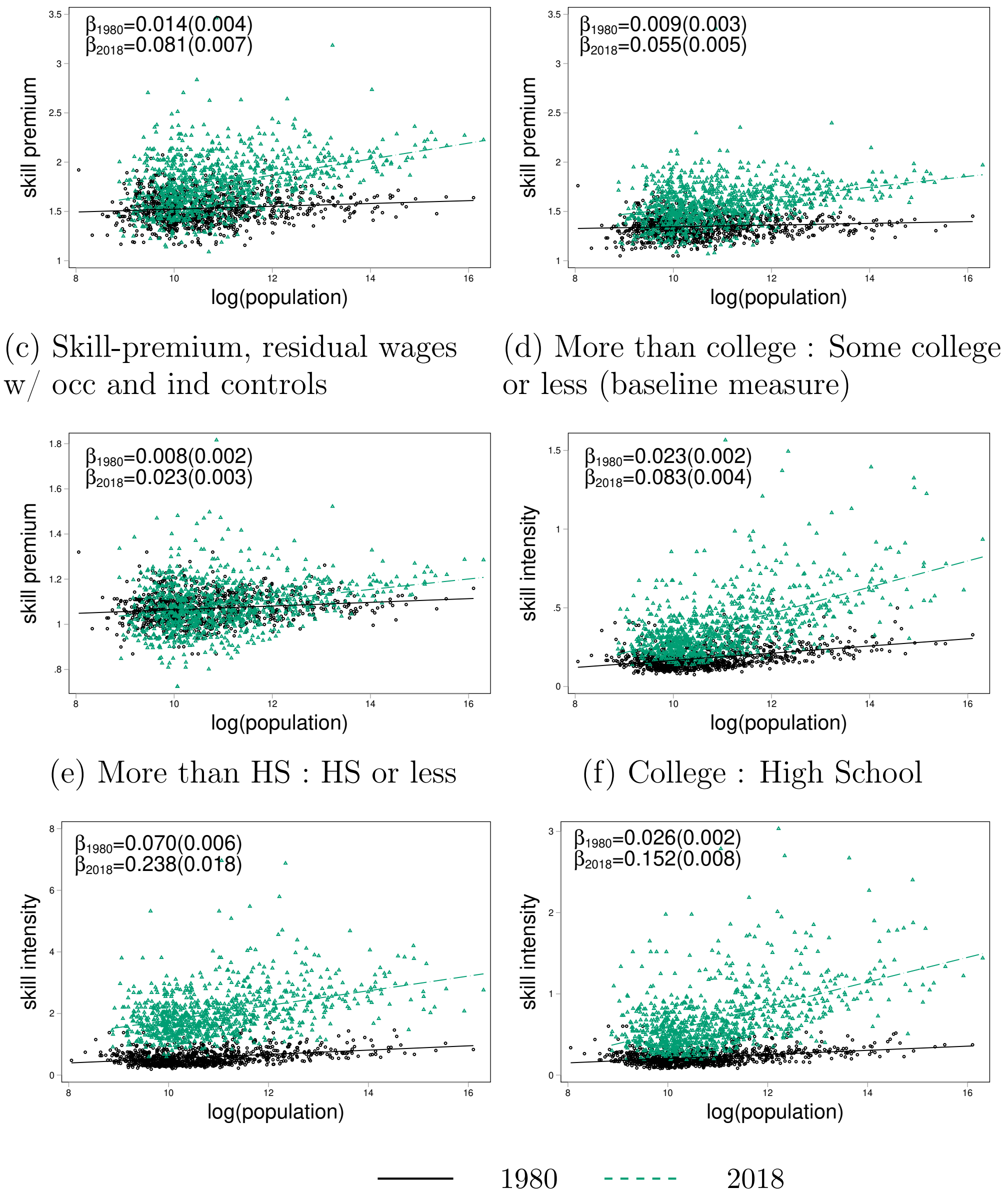
I limit the sample to civilian workers aged 16 to 64 who work at least 30 hours a week. I use the occupation codes from Autor and Dorn (2013) and the 1990 industry codes provided by IPUMS. I group the detailed race categories into five groups that are consistent over time: white non-Hispanic, Hispanic, Black, Asian, and other. 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. I show below that 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 of 12 times the monthly rent for renters and 5 percent of their house's value for homeowners. I use estimates of housing supply elasticity from Saiz (2010).

Figure A.I shows the relationship between the skill premium and city size by year. Panel (a) shows the correlation using average wages for high- and low-skilled workers, respectively, while Panel (b) uses residual wages from a standard Mincer regression controlling for age, age squared, sex, race, and foreign-born status. Panel (c) adds controls for industry and occupation. With each specification, the relationship between the skill-premium and city size increases between 1980 and 2018.

Figure A.I shows the relationship between skill-intensity and city size for 1980 and 2018. Figure

[17] I stop my analysis in 2018 because, at the time of writing, that is the last year the BDS is available.

Figure A.I: Skill-premium and skill-intensity by city size



(a) Baseline skill-premium

(b) Skill-premium, residual wages

(c) Skill-premium, residual wages w/ occ and ind controls

(d) More than college : Some college or less (baseline measure)

(e) More than HS : HS or less

(f) College : High School

Source: Wages and skill supply are from the 1980 Decennial Census and 2018 ACS. Population is working-age population (ages 20-64) from the Intercensal Population Estimates. Figures display the relationship between the skill premium and skill intensity with city size. Panel (a) shows the skill-premium using average wages. Panel (b) shows the skill-premium using wages adjusted for demographics including age, age squared, gender, and race dummies. Panel (c) adds further controls for occupation and industry fixed effects. Panel (d) defines H as those with 4+ years of college and L as those with less than four years of college. Panel (e) defines H as those with some college or more and L as those with a high school diploma or less. Panel (f) defines H as those with exactly four years of college and L as those with exactly a high school diploma.

A.I includes three definitions of skill intensity. First, panel (d) shows the baseline measure of skill intensity, those with four or more years of college divided by those with some college or less. Panel (e) uses an alternative definition: those with some college or more divided by those with high school or less. Finally, panel (f) uses those with exactly four 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 increased between 1980 and 2018.

A.2 Descriptive facts robustness checks: Firm facts

Additional measures of dynamism

Here, I present the main facts using several alternate measures of dynamism, including the firm 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 $er_t = \frac{\text{firm deaths}_t}{(\text{firms}_t + \text{firms}_{t-1})/2}$. Job creation and destruction are employment increases and decreases at expanding or contracting firms, respectively. Specifically, the job creation rate 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 the job destruction rate is

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

I take five-year moving averages of all dynamism rates to smooth out spurious year-to-year fluctuations in the data due to small city-industry bins and imputation error from Economic Census years.

Table A.I shows the relationship between dynamism and city size for 1980 and 2018, using data from the public Business Dynamic Statistics, controlling for a full set of sector-year fixed effects. In 1980, there was no relationship, or even a negative relationship, between measures of dynamism and city size, while by 2018, big cities have become more dynamic than small cities. Panel (b) shows the same coefficients β_t over time, but with further controls for the lagged 5 year annualized 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 dynamism. The same conclusions hold—the cross-sectional correlation between dynamism and city size has increased between 1980 and 2018.

Finally, Table A.II shows the age-conditional exit rates against city size, controlling for the sector-by-year fixed effects. Consistent with selection becoming tougher, the relationship between the probability of exit for establishments that are part of young firms and city size is growing over time. For firms aged 1 to 5, the likelihood that an establishment exits was strongly decreasing in city size in 1980 and 1989. By 2018, there was no significant relationship between the probability of exit for young firms and city size. On the other hand, for older firms aged 11+, there was a positive relationship between the probability of exit and city size in 1992.[18] Thus, while young firms used to be more likely to exit in big cities, old firms used to be less likely to exit, consistent with the model in which selection increases more in big cities than in small cities. In my model, the relationship between exit and city size for old firms will not become weaker over time, but the change in the slope will be smaller.

[18] 1992 is the first year we can measure the five-year average (1988-1992) of the age 11+ exit rate. In the LBD and BDS, age is imputed from the first year an establishment is observed in the dataset, which begins in 1977.

Table A.I: Dynamism and city size, controlling for sector composition

Panel A. Industry Controls

	esr	eer	fsr	fer	jcr	jdr
$log(pop_{1980})$	-0.305***	-0.234***	-0.269***	-0.137***	-0.883***	-0.547***
	(0.0366)	(0.0229)	(0.0450)	(0.0257)	(0.0661)	(0.0488)
$log(pop_{2018})$	0.423***	0.169***	0.663***	0.327***	0.293***	-0.0872***
	(0.0208)	(0.0151)	(0.0220)	(0.0153)	(0.0305)	(0.0288)
Naics 3-Year FES	Y	Y	Y	Y	Y	Y
Demogr. Ctrls.	N	N	N	N	N	N
Observations	9068	9068	9068	9068	9068	9068
R^2	0.695	0.772	0.694	0.727	0.716	0.729

Panel B. Adding Demographic Controls

	esr	eer	fsr	fer	jcr	jdr
$log(pop_{1980})$	0.0269	-0.139***	0.126***	-0.0166	-0.427***	-0.421***
	(0.0289)	(0.0226)	(0.0345)	(0.0244)	(0.0605)	(0.0511)
$log(pop_{2018})$	0.260***	0.126***	0.462***	0.264***	0.116***	-0.172***
	(0.0226)	(0.0175)	(0.0236)	(0.0179)	(0.0351)	(0.0349)
Sector-Year FES	Y	Y	Y	Y	Y	Y
Demogr. Ctrls.	Y	Y	Y	Y	Y	Y
Observations	9068	9068	9068	9068	9068	9068
R^2	0.790	0.791	0.800	0.756	0.761	0.737

Note: Regression of dynamism on population, controlling for sector-year fixed effects. Source: Business Dynamic Statistics and author calculations. Population is working-age population (ages 20-64) from the Intercensal Population Estimates. Unit of observation is a CBSA-Sector-Year. Demographic controls include annualized working-age population growth between year t-5 and t and prime age population share. I drop sector-cbsa pairs that have missing values of dynamism in 1980 or 2018.

Table A.II: Age-conditional exit rates and city size

	Establishment exit rates		
	ages 1-5	ages 6-10	ages 11+
$log(pop_{1980})$	-0.448***		
	(0.0497)		
$log(pop_{1989})$	-0.101**	-0.105**	
	(0.0432)	(0.0487)	
$log(pop_{1992})$	0.240***	0.159***	0.331***
	(0.0377)	(0.0455)	(0.0765)
$log(pop_{2018})$	0.00605	0.156***	0.0445
	(0.0351)	(0.0567)	(0.0284)
Observations	12424	4803	2064
R^2	0.685	0.651	0.744

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

Source: Regression of age-conditional exit rates on population, controlling for sector-year fixed effects. Source: Business Dynamic Statistics and author calculations. Population is working-age population (ages 20-64) from the Intercensal Population Estimates. Unit of observation is a CBSA-Sector. I drop CBSA-Sector pairs with missing values of dynamism in any year. Values for firms aged 6-10 and those aged 11+ are missing in the early years due to data availability.

Robustness checks in the Longitudinal Business Database

In Table A.III, I present the correlation between dynamism and city size over time controlling for industry composition at the three-digit NAICS level. To do this, I compute dynamism measures within a city-industry cell[19] and then estimate the following specification for each year

$$D_{ict} = \alpha_{it} + \sum_t \beta_t log(pop_{ct})\mathbb{I}[year = t] + \epsilon_{ict}.$$

Table A.III shows the coefficients β_t over time. Panel (b) of Table A.III shows the same coefficients β_t over time, but with further controls for the lagged ten-year population growth in the city and the prime age worker share.

Decline in dynamism versus initial city size

Next, I show that cities that were small in 1980 exhibited larger declines in dynamism than 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 A.II shows the relationship between the decline in dynamism versus 1980 city size in the public BDS data. In cities twice as large, the establishment start-up rate fell by about .3 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.

To control for industry composition, I estimate the following specification:

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

where $\Delta_{2014,1980} D_{ic}$ is the change in dynamism between 1980 and 2014 in city c and industry i, α_i is a full set of industry fixed effects and pop_{c1980} is a city's 1980 working-age population. I estimate this at the city by three-digit NAICS level. I look at the change between 1980 and 2014 instead of 2018 because 2014 is the latest year I have available in the LBD. The results are presented in Table A.IV.

[19] I note that results from the LBD all use 3-year instead of 5-year averages of dynamism rates. This is because the previous version of this paper used three year moving averages. It was later pointed out that Five years would be better because of imputation error from Census years. However, due to the cumbersome disclosure review process, I did not update the results in the LBD. These results are only included as robustness checks in the appendix.

Table A.III: Dynamism and city size, controlling for industry composition

Panel A. Industry Controls

	esr	eer	fsr	fer	jcr	jdr
$log(pop_{1980})$	0.184**	-0.270***	0.347***	0.124***	-1.146***	-0.994***
	(0.0741)	(0.0479)	(0.0721)	(0.0376)	(0.103)	(0.0621)
$log(pop_{2014})$	0.584***	0.103***	0.711***	0.268***	-0.224***	-0.578***
	(0.0398)	(0.0326)	(0.0356)	(0.0282)	(0.0509)	(0.0426)
Naics 3-Year FES	Y	Y	Y	Y	Y	Y
Demogr. Ctrls.	N	N	N	N	N	N
Observations	950,000	950,000	950,000	950,000	950,000	950,000
R^2	0.165	0.110	0.225	0.114	0.167	0.129

Panel B. Adding Demographic Controls

	esr	eer	fsr	fer	jcr	jdr
$log(pop_{1980})$	0.186***	-0.271***	0.363***	0.116***	-1.089***	-0.972***
	(0.0570)	(0.0498)	(0.0567)	(0.0417)	(0.0865)	(0.0684)
$log(pop_{2014})$	0.414***	0.107**	0.569***	0.269***	-0.520***	-0.595***
	(0.0484)	(0.0434)	(0.0406)	(0.0346)	(0.0733)	(0.0512)
Naics 3-Year FES	Y	Y	Y	Y	Y	Y
Demogr. Ctrls.	Y	Y	Y	Y	Y	Y
Observations	950,000	950,000	950,000	950,000	950,000	950,000
R^2	0.184	0.112	0.247	0.116	0.181	0.130

Note: Regression of dynamism on population, controlling for 3-digit NAICS fixed effects. Source: Longitudinal Business Database and author calculations. Population is working-age population (ages 20-64) from the Intercensal Population Estimates. Unit of observation is a CBSA-Naics3-Year. Demographic controls include annualized working-age population growth between year t-10 and t and prime age population share. This regression was run on every year between 1980 and 2014 but I suppress the coefficients on other years for readability.

Table A.IV: Decline in Dynamism and City Size, industry controls

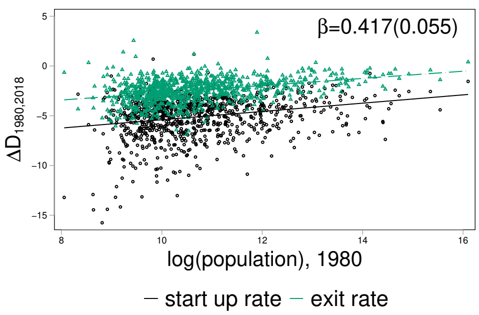
	Δ 1980 to 2014					
	sr	er	esr	eer	jcr	jdr
$log(pop_{1980})$	0.349***	0.0979***	0.322***	0.286***	0.780***	0.308***
	(0.0557)	(0.0387)	(0.0586)	(0.0444)	(0.0822)	(0.0567)
Observations	57,500	57,500	57,500	57,500	58,500	58,500
R^2	0.141	0.066	0.101	0.070	0.073	0.061

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

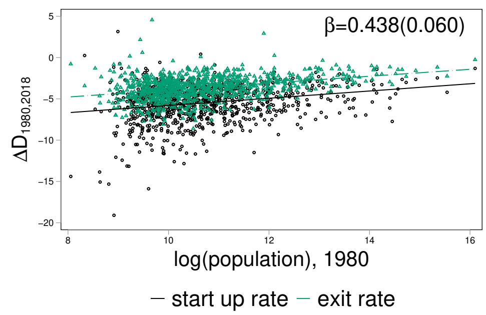
Source: LBD and author calculations. Working-age population is from the Intercensal Population Estimates. The table displays the relationship between the decline in dynamism, as measured by the establishment start-up and exit rate, the firm start-up and exit rate and job creation and destruction between 1980 and 2014 and city size. The unit of observation is a CBSA × 3-digit NAICS. The 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.

Figure A.II: Decline in dynamism versus initial city size

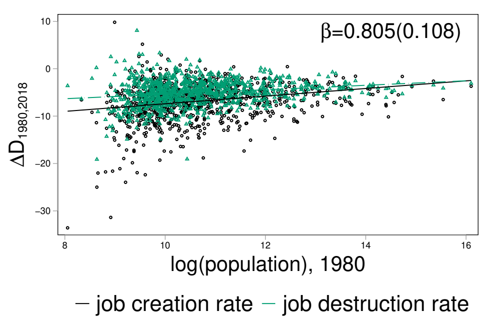
(a) Firm start-up and exit rate



(b) Establishment start-up and exit rate



(c) Job creation and job destruction rate



Source: Business Dynamic Statistics and author calculations. Working-age population is from the Intercensal Population Estimates. The figure displays the relationship between the 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 2018 and city size.

Table A.V: Matchers versus non-matchers

	average ICT investment share	average ICT investment	average total investment
non-matchers	0.316	16.02	96.27
	(0.35)	(873)	(5,326)
matchers	0.314	99.27	518.90
	(0.35)	(13,880)	(62,690)

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.

A.3 Annualized Capital Expenditure Survey Data Appendix

In this section, I provide 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, 88.8% of them match to a firm in the LBD. Since the ACES over-samples large firms; it is more relevant that 92.0% of the total sampling weight has a match in the LBD. In Table A.V, I show some summary statistics on the firms that match versus those that do not. While both matchers and non-matchers have similar ICT investment shares, 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, multiple observations in the ACES may 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 unavailable, 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.

A.4 Robustness of ICT results to firm location

In this section, I perform two robustness exercises to show that the relationship between ICT spending and city size is robust to the way in which firm location is assigned in the ACES dataset. First, in panel (a) of Table A.VI, I show that the results are robust to using single establishments firms only. For these establishments, there is no ambiguity as to the location of the firm—firm location is simply the location of their single establishment.

Second, in panel (b), I use an alternative dataset of ICT expenditures that is available at the establishment level rather than the firm level—the Census of Manufacturers (CMF), which contains data on computer purchases. The drawback to this dataset is that it only covers manufacturing firms. However, because in the CMF, computer purchases vary at the establishment level, I add a further specification controlling for firm fixed effects.

Across all specifications, the results are the same, ICT spending per employee and the ICT share of investment is increasing in city size. Even within a firm, computer spending is higher at establishments located in big cities.

Table A.VI: Robustness of ICT results to firm location

	Panel A: ACES Single Establishment Sample					
	$log(\frac{\text{ICT investment}}{\text{Employment}})$			$\frac{\text{ICT investment}}{\text{Total investment}}$		
log(pop)	0.0525***	0.0548***		0.0116***	0.0118***	
	(0.0102)	(0.00787)		(0.00136)	(0.00137)	
log(emp)		-0.464***			-0.0116***	
		(0.00954)			(0.00179)	
NAICS 4 FE	Y	Y		Y	Y	
Year FE	Y	Y		Y	Y	
Age FE	N	Y		N	Y	
Firm FE	N	N		N	N	
Observations	89000	89000		111000	111000	
R^2	0.227	0.364		0.173	0.175	
	Panel B: CMF Sample					
	$log(\frac{Computers}{Employment})$			$\frac{\text{Computer spending}}{\text{Total investment}}$		
log(pop)	0.0330***	0.0217***	0.0328***	0.00165***	0.00183***	0.00252***
	(0.00496)	(0.00460)	(0.00584)	(0.000309)	(0.000284)	(0.000485)
log(emp)		-0.229***	-0.348***		0.00381***	-0.000886
		(0.00432)	(0.0109)		(0.000452)	(0.000652)
NAICS 4 FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Age FE	N	Y	Y	N	Y	Y
Firm FE	N	N	Y	N	N	Y
Observations	486000	486000	486000	675000	675000	675000
R^2	0.087	0.168	0.637	0.054	0.056	0.559

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

B Derivations for firm problems

B.1 Value function

The problem of the firm is to choose the exit threshold below which they will exit the market. They choose the exit threshold to solve

$$V(z) = \max_{z_{xj} \le z} \mathbb{E}_z \int_0^{T(z_{xj})} e^{-\rho t} \left(\pi_j(z(t))\right) dt$$

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

Defining $s = \ln z^{\sigma-1}$ and using a change of variable, the value function can be rewritten as

$$V(s) = \max_{s_{xj} \le s} \mathbb{E}_s \int_0^{T(s_{xj})} e^{-\rho t} \left[Z_j^k e^s - F^k\right] dt$$

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

where $\tilde{\mu} = (\sigma - 1)\left(\mu - \frac{1}{2}\Psi^2\right)$ and $\tilde{\Psi} = \Psi(\sigma - 1)$ and $Z_j^k = \left(\frac{C_j^k}{\psi_j^k}\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Y_j P_j^\sigma \frac{1}{\sigma}$. The stochastic process for s is derived using Ito's Lemma[20] and follows a standard Brownian motion. Assuming $R_2 = \frac{-\tilde{\mu}+\sqrt{\tilde{\mu}^2+2\tilde{\Psi}^2\rho}}{\tilde{\Psi}^2} > 1$, then following Proposition 6.3 from Stokey (2008), s_x solves

$$\Pi(s_{xj}) = R_2 \int_0^\infty e^{-R_2\xi} \pi(s_{xj} + \xi) d\xi = 0$$

Integrating

$$\begin{aligned}
\Pi(s_{xj}) =& R_2 \int_0^\infty e^{-R_2\xi} \pi(s_{xj} + \xi) d\xi \\
&= R_2 \int_0^{s_a} e^{-R_2\xi} \left[Z^n e^{s_x+\xi} - F\right] d\xi + R_2 \int_{s_a}^\infty e^{-R_2\xi} \left[Z^a e^{s_x+\xi} - F^a\right] d\xi = 0 \\
&= R_2 \left[\frac{Z^n e^{s_x+s_a-R_2 s_a}}{1-R_2} + \frac{F e^{-R_2 s_a}}{R_2} - \frac{Z^n e^{s_x}}{1-R_2} - \frac{F}{R_2}\right] - R_2 \left[\frac{Z^a e^{s_x+s_a-R_2 s_a}}{1-R_2} + \frac{F^a e^{-R_2 s_a}}{R_2}\right].
\end{aligned}$$

Finally, setting this equal to zero gives

$$\Pi(s_{xj}) = 0 \implies e^{s_x} = \frac{R_2 - 1}{R_2} \frac{\left(F^a - F^n\right) e^{-R_2 s_a} + F^n}{\left(Z^a - Z^n\right) e^{s_a - R_2 s_a} + Z^n}.$$

[20] 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 $dy = df(x) = \left(\mu(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x)\right) dt + \sigma(x)f'(x)dW$. Thus,

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

B.2 Firm size distribution

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

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

In steady state, the distribution is unchanging, meaning $\frac{\partial g(s)}{dt} = 0$ This implies $-\delta\frac{dg(s)}{ds} = \frac{d^2g(s)}{ds^2}$ 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 are four boundary conditions that we can use to solve for $\{C_1, C_2, C_3, C_4\}$:

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

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

   $$\begin{aligned} \Delta N &\approx (1-p)h\left(g^+(s_e) + g^{+\prime}(s_e)h + g^{+\prime\prime}(s_e)h^2\right) \\ &\approx ph\left(g^-(s_e) + g^{-\prime}(s_e)(-h) + g^{-\prime\prime}(s_e)h^2\right) \end{aligned}$$

   Divide both sides by $\sqrt{\Delta}$ and take the limit as $\Delta \to 0$ 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) $(C_4 - C_2)\, e^{-\delta s_e} + C_2 e^{-\delta s_x} = 0 \implies C_4 = C_2\frac{\left(e^{-\delta s_e} - e^{-\delta s_x}\right)}{e^{-\delta s_e}}$
   (b) This gives $g(.)$ as a function of C_2

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

4. To solve for C_2, we can use that the mass of exiters is

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

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

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

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 as being an equilibrium of the model in 1980. The vector of fundamentals is given by $\mathbb{P}^c = \{\gamma, \psi, A_h, A_l, f^e, f^c, \phi, \eta\}$; that is, the weight on high-skilled labor, γ, the city specific productivity term, ψ, high- and low-skilled amenities, A_h and A_l, entry costs, f_e, fixed costs, f_c, the productivity of the building sector ϕ, and the elasticity of building supply, η. Table IV in the main text 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 (10), is determined.

2. Solve for γ_j using the two labor market clearing conditions. The two labor market clearing conditions are
$$H_j = h_j^d \left(Y_j + M_j f_j^c + E_j f_j^e\right)$$
$$L_j = l_j^d \left(Y_j + M_j f_j^c + E_j f_j^e\right),$$
where h_j^d and l_j^d are the Hicksian demand functions for high- and low-skilled labor, respectively. 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 γ 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 ψ_j. In city 1, ψ_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 $c_j(z) = \frac{C_j^n}{\psi_j^n z}$ and Ω_j was pinned down in step 1 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 ψ_j.
$$\psi_j = \left(\int_{\Omega_j} \left(\frac{C_j^n}{\psi_j^n z}\frac{\sigma}{\sigma-1}\right)^{1-\sigma} d\omega\right)^{\frac{1}{1-\sigma}}$$

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

entry cost f^e.

$$\text{final goods: } \beta w_{Hj}H_j + \beta w_{Lj}L_j + \left(\Pi_j^v - M_j^n C_j^n f_j^n - M_j^a C_j^a f_j^a - E_j C_j^n f_j^e\right) + \Pi_j^B + c_j^B = Y_j$$

$$\text{free entry: } V_j(z_e) = C_j^n f_j^e$$

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

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

$$\frac{\rho\frac{(1-\xi^-)}{-\xi^-}\frac{Z_j}{\rho-\tilde{\mu}-\frac{1}{2}\tilde{\Psi}^2}e^{s_x}}{C_j^n} = f^c.$$

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 η_j. Then use the building market clearing condition to solve for b_j

$$B_j^d + (1-\beta)H_j\frac{w_{hj}}{r_j} + (1-\beta)L_j\frac{w_{lj}}{r_j} = \phi_j r^{\eta_j}$$

where B_j^d gives the demand for buildings from firms.

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

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

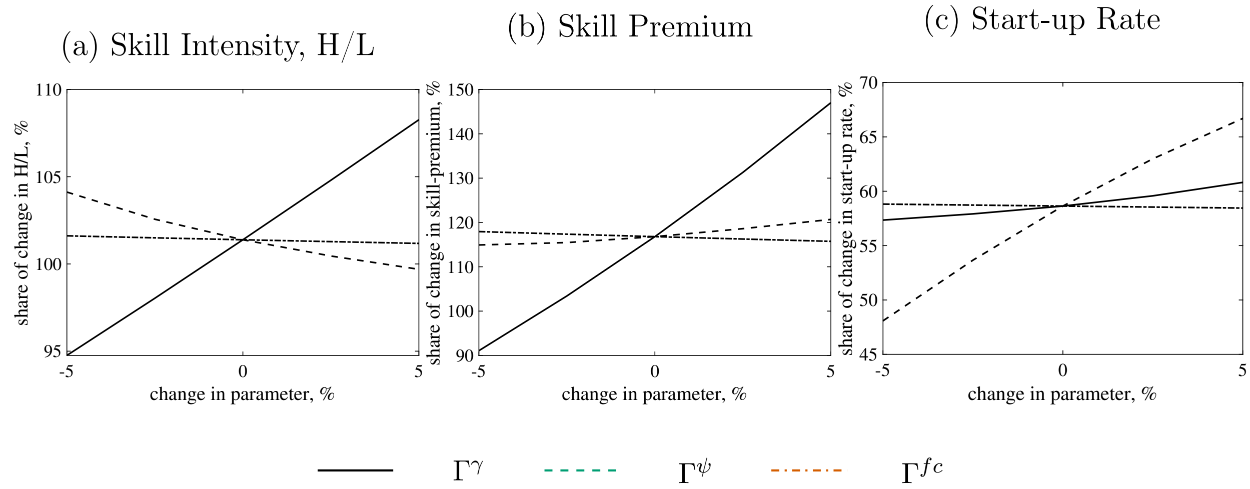
In city 1, the amenities are normalized to be 1, so $\bar{V}_{\tau 1}$ is known. Further, the $\pi_{\tau j}$s are known from the data. Define a vector $x = [-\pi_{\tau 2}\bar{V}_{\tau 1}, ..., -\pi_{\tau J}\bar{V}_{\tau 1}]$ and a matrix

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

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

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

Figure C.I: Robustness to the parameters of the new technology



Source: Model-generated output. Figures (a) - (c) show how the main results change in response to the parameters of the new technology. (a) shows the share of the cross-sectional change in skill intensity that is matched by the introduction of the new technology. (b) and (c) show the skill-premium and the start-up rate, respectively.

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

C.2 Robustness

In this section, I perform two additional robustness checks examining the sensitivity of the main results to alternative calibrations of the model. These join the robustness checks already presented in Section 4.3 that show how the main results would change when relaxing the assumption that all aggregate wage growth is attributed to the introduction of the new technology. First, I examine the impact of varying the parameters of the new technology within 5% of their calibrated values. Second, I recalibrate the model under three alternative assumptions, such as denominating the fixed costs in units of the final good instead of land and labor, and under alternative values for the parameters σ and ν.

Robustness to the parameters of the new technology. Figure C.I shows how the main cross-sectional results change as the parameters of the new technology are increased or decreased by 5%. Panel (a) shows the share of the cross-sectional change in skill intensity matched by the introduction of the new technology. In the baseline calibration, the model matches 101% of the changing relationship between skill intensity and city-size. This number fluctuates between 95 and 110 percent as the parameters of the new technology are increased or decreased by 5%. Panels (b) and (c) show the extent to which the model can match the changing cross-sectional relationship between the skill-premium and city size and the start-up rate and city size, respectively. In each case, the main results are qualitatively similar within a range of the Γs around the baseline calibration.

Table C.I: Robustness Exercises

	Share of cross-sectional change explained by model in:		
	Skill-premium	Skill-intensity	Start-up rate
Baseline	116.8	101.4	58.6
Fixed costs in output	123.9	105.6	42.1
Alternative σ	101.4	101.1	39.8
Alternative ν	125.4	156.2	62.9

Notes: Table shows the share of the cross-sectional changes explained by the model. The first row gives the results for the baseline calibration. Rows 2-4 show the robustness of the main results to alternative calibrations.

Robustness to alternative model calibrations. In this section, I show that the results are robust to alternative model calibrations. Specifically, I show that the main findings are qualitatively similar when: (1) the fixed costs are denominated in the final good instead of the factor shares; (2) the elasticity of substitution between varieties σ is set to 4 (another commonly used value in the literature) instead of 6.5; (3) alternative values of shape parameter of the Fréchet distribution for the idiosyncratic taste shocks across cities, ν, which governs the migration elasticity. In the baseline, I set ν to 1.5 following Fajgelbaum et al. (2019). In the alternative calibration, I set $\nu = 4.98$ for high-skilled workers and $\nu = 3.26$ for low-skilled workers, following Diamond (2016). In all three cases, I recalibrate the city-level fundamentals and the parameters of the new high-skilled technology as described in Section 4.

The results are shown in rows 2 to 4 of Table C.I. I examine the extent to which the introduction of the new technology can account for the cross-sectional changes observed in the Great Divergence. For comparison, the baseline results are reprinted in Row (1).

Under all three alternative calibrations, the model continues to match the main features of the Great Divergence. The main difference is in the extent to which the model can match the changing relationship between the start-up rate and city size. In the baseline calibration, the model predicts 59 percent of the changing slope. When the fixed cost is denominated in the final good, the model matches 42 percent of the changing slope, and when $\sigma = 4$, the model matches 40 percent of the changing slope. When the fixed costs are denominated in the factors of production instead of the final good, the increase in factor costs amplifies the changes in selection across cities. The fixed costs for the small non-adopters increase because of the change in factor prices, raising the exit threshold. Similarly, when σ, which governs the elasticity of substitution between varieties, is higher, small firms face more competition from the larger adopters. With more elastic demand, consumers more readily switch to goods produced by the big adopting firms. The small firms lose market share and are more likely to exit. Consequently, with lower σ, the model accounts for a smaller proportion of the cross-sectional change between the start-up rate and city size. The remaining results are similar under the alternative calibrations.

Table C.II: Robustness for Policy Experiments

	Adoption			Skill Premium		
	Small cities	Big cities	Aggregate	Small cities	Big cities	Aggregate
Baseline						
No Policy Intervention	1.94	3.80	3.66	1.93	2.12	2.12
Adoption subsidy small cities	3.40	3.78	3.75	1.99	2.13	2.12
Adoption subsidy everywhere	3.05	6.18	5.94	2.01	2.22	2.22
Building subsidy big cities	1.88	3.82	3.73	1.95	2.17	2.17
Target 50% of wage growth						
No Policy Intervention	0.26	0.78	0.74	1.63	1.84	1.83
Adoption subsidy small cities	0.47	0.78	0.76	1.67	1.84	1.84
Adoption subsidy everywhere	0.42	1.34	1.27	1.68	1.92	1.91
Building subsidy big cities	0.25	0.80	0.77	1.65	1.89	1.89
Allow change in old tech.						
No Policy Intervention	32.91	45.38	44.47	2.05	2.15	2.15
Adoption subsidy small cities	84.52	56.43	58.49	2.09	2.17	2.16
Adoption subsidy everywhere	78.05	95.51	94.23	2.11	2.22	2.22
Building subsidy big cities	32.33	45.62	45.03	2.08	2.20	2.20

C.3 Robustness for policy experiments

Section 4.4 presents robustness checks under two alternative assumptions for calibrating the parameters of the new technology. First, I assume that only 50% of the wage growth between the two steady states is attributable to technology adoption. Second, I calibrate the parameters of the new technology while also allowing for technical change in the old technology.

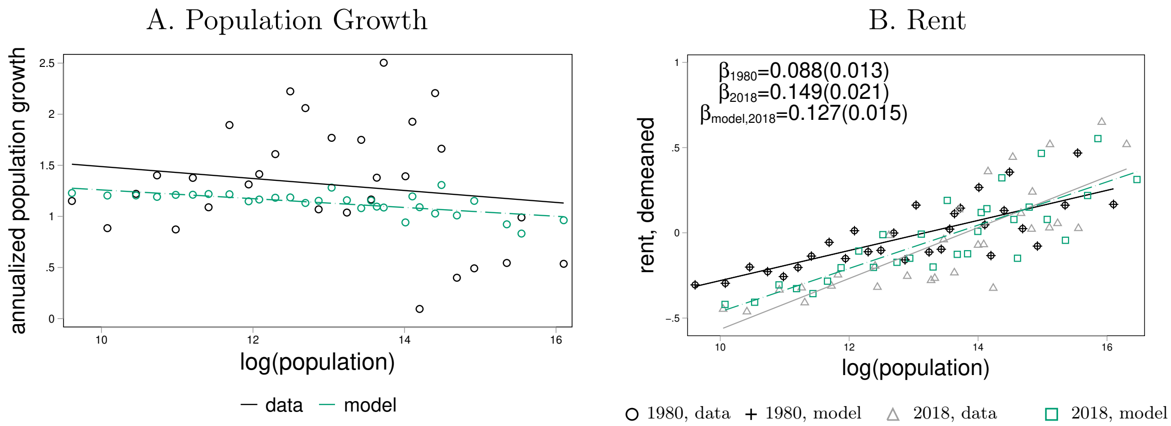
This section assesses the robustness of the policy results under each of the alternative calibrations. The results are presented in Table C.II. As in Section 6, I evaluate three policies: an adoption subsidy in small cities (cities 1-15), an adoption subsidy in all cities, and a building subsidy in big cities (cities 16-30). For succinctness, I present the results grouping cities into two bins: small cities (1-15) and big cities (16-30). Table C.II reports two key outcomes: the adoption rate and the skill premium.

The first four rows labeled “Baseline” reproduce the results from Section 6 under the baseline calibration. The policy of subsidizing adoption in small cities is effective in narrowing the gap between big and small cities in adoption rates and the skill premium. However, it has a limited aggregate impact, increasing the overall adoption rate by just 0.1 percentage point. In contrast, policies such as subsidizing adoption in all cities or subsidizing the building supply in big cities exacerbate the differences between big and small cities in adoption and the skill premium, but also have the biggest aggregate impact on adoption.

The next four rows, labeled “Target 50% of wage growth,” present the same policy interventions under an alternative calibration in which only half of the aggregate wage growth is attributed to technology adoption. The “No Policy Intervention" row shows the aggregates under this alternative calibration with no policy intervention. The adoption rates are much lower because less wage growth is being attributed to technology adoption. However, the impact of the policies is qualitatively similar. Subsidizing adoption in small cities equalizes the differences between big and small cities, but has a small aggregate impact. In contrast, subsidizing adoption everywhere and subsidizing building supply in big cities exacerbate the differences between big and small cities but generate the largest aggregate impact.

Finally, the bottom four rows show the policy interventions under a calibration that allows for technical change in the old technology alongside adoption of the new technology. The adoption rate is much higher under this alternative calibration, but the difference between the old technology and the new technology is much smaller. In the baseline calibration, the new technology scales up the weight on high-skilled labor by 24% while under the alternative calibration, it scales it by only 7%. This is because the old technology has also become more skill-biased. As a result, though adoption is more widespread, it has a smaller impact on the firm's skill-bias. Nevertheless, the qualitative impact of the policies is similar. Subsidizing adoption in the small city equalizes the differences between big and small cities. Unlike in the baseline calibration, it is much more effective at increasing adoption in the small city, but still has a small impact on the aggregate skill premium. In contrast, subsidizing adoption everywhere and subsidizing the building supply in the big cities exacerbates the differences between big and small cities while also having the biggest impact on the aggregate skill premium.

Figure D.I: Rent and Population, model vs. data



Source: Population growth and population are working-age population (ages 20-64) from the Intercensal Population Estimates. Rent is computed from the 1980 Decennial Census and 2018 ACS. Additional data from model output. Panel A displays the relationship between population growth and initial city size in the model and the data. The black line shows population growth between 1980 and 2018 in the data, and the grey line shows population growth between the initial and final steady state in the model. Panel B 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. Grey gives the 2018 data and green the 2018 steady state in the model. The unit of observation is one of 30 city-size categories.

D Additional Model Results

D.1 Changes in population: model vs. data

In this section, I verify that the relationship between population growth and initial population in the model is consistent with the data. Figure D.I shows annualized population growth between 1980 and 2018 across cities in the model and in the data. In both the model and the data there is a slight negative relationship between population growth and initial city size, though it is not statistically significant.

This relationship is counter-intuitive since we would expect big cities with more adoption to grow faster. However, these cities also have a lower housing supply elasticity. Thus, even though they adopt more, congestion effects limit 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 between population growth and initial city size. In particular, Rappaport (2018) finds that big cities grow less than less constrained and less crowded medium-sized cities.

D.2 Changes in rent: model vs data

In this section, I show that the model can reproduce the changing relationship between rent and city size in the data. In Figure D.I, 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 grey, I show the relationship between rents and city size in the 2018 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 .088 to .149, while in the model, it increases from .088 to .127. Thus, the model explains 63% of the increasing relationship between rent and city size in the data.

D.3 Drivers of heterogeneity in adoption: additional results

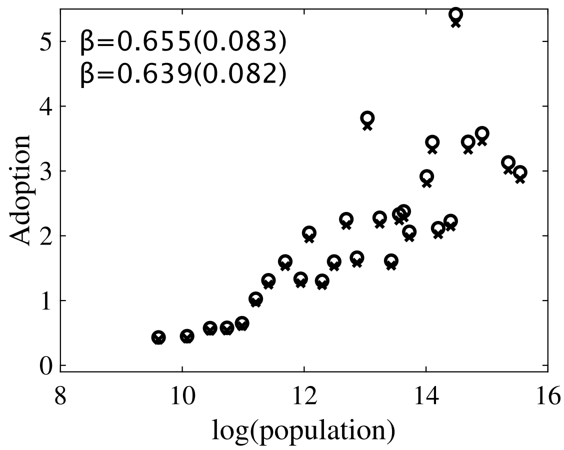
In Section 5.1, I show the main sources of city-level heterogeneity that drive differences in technology adoption. Here, I show the effect of additional city parameters that do not significantly affect adoption. While these parameters are essential for other dimensions of heterogeneity, like the level of wages, rents, and firm size, they are not the primary drivers of differences in adoption rates.

The counterfactuals shutting down the heterogeneity in ψ_j, in panel (a), and heterogeneity in the housing market, ϕ_j and η_j, in panel (b), demonstrate the effect of the market size. Equalizing ψ decreases the size of the biggest markets, reducing adoption, while equalizing the parameters of the housing market increases the size of the biggest markets, increasing adoption.

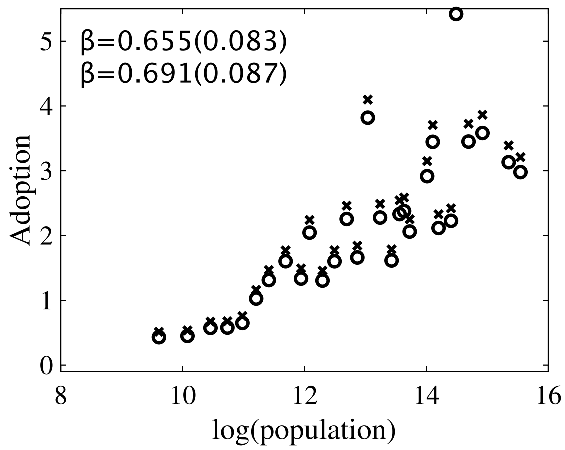
The role of fixed costs and entry costs are shown in panel (c). While these parameters matter for heterogeneity across cities in the start-up rate, firm size, and selection, they are not important factors in driving the heterogeneity in adoption. Panel (c) shows that when the fixed and entry costs are equalized, there is almost no change in cross-sectional pattern of adoption.

Figure D.II: Counterfactual adoption rates

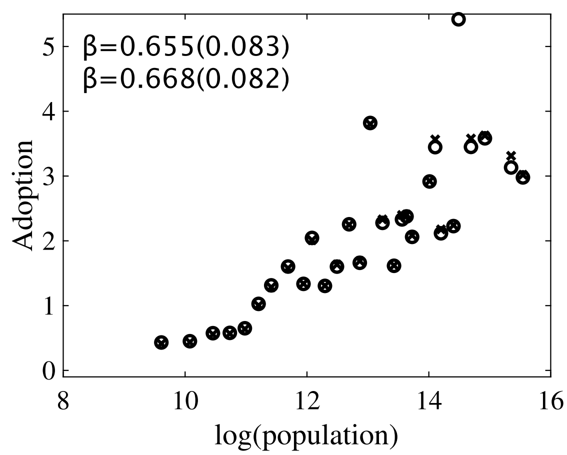
(a) ψ



(b) ϕ and η



(c) f^c and f^e



○ Baseline × Counterfactual

Note: Figure displays counterfactual adoption rates when parameters defining a city are equalized.