DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Essays in Local Labor Economics

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation	Diamond, Rebecca. 2013. Essays in Local Labor Economics. Doctoral dissertation, Harvard University.
Accessed	April 17, 2018 5:02:42 PM EDT
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:12362593
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

Essays in Local Labor Economics

A dissertation presented

by

Rebecca Randolph Diamond

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

April 2013

©2013 Rebecca Randolph Diamond

All rights reserved.

Disseration Advisor: Lawrence Katz Rebecca Randolph Diamond

Essays in Local Labor Economics

Abstract

This dissertation consists of three independent chapters. Chapter 1 examines the determinants and welfare implications of the increased geographic of workers by skill from 1980 to 2000. I estimate a structural spatial equilibrium model of local labor demand, housing supply, labor supply, and amenity levels. The estimates indicate that cross-city changes in firms' demands for high and low skill labor were the underlying forces driving the increase in geographic skill sorting. I find that the combined effects of changes in cities' wages, rents, and endogenous amenities increased well-being inequality between high school and college graduates by a significantly larger amount than would be suggested by the increase in the college wage gap alone.

Chapter 2 examines the abilities of state and local governments to extract rent from private sector workers by charging high tax rates and paying government workers high wages. Using a spatial equilibrium model where private sector workers are free to migrate across government jurisdictions, I show that variation in areas' housing supply elasticities differentially restrains governments' abilities to extract rent from private sector workers. Governments in less housing elastic areas can charge higher taxes without worry of shrinking their tax bases. I test the model's predictions by using worker wage data from the CPS-MORG. I find the public-private sector wage gap is higher in areas with less elastic housing supplies.

Chapter 3 studies the standard practice in regression analyses to allow for clustering in the error covariance matrix when an explanatory variable varies at a more aggregate level than the units of observation. However, the structure of the error covariance matrix may be more complex, with correlations not vanishing for units in different clusters. I show that with equal-sized clusters, if the covariate of interest is randomly assigned at the cluster level, only accounting for non-zero covariances at the cluster

level, and ignoring correlations between clusters as well as differences in within-cluster correlations, leads to valid confidence intervals. However, in the absence of random assignment of the covariates, ignoring general correlation structures may lead to biases in standard errors.

Acknowledgements

I would like to thank my advisors, professors Lawrence Katz, Ariel Pakes, and Edward Glaeser for their advice, help, and support throughout my graduate studies. Their guidance has been crucial to the completion of this dissertation and shaping my abilities to produce economic research.

In the past five years I have benefitted from many interactions with other economics graduate students and professors at Harvard University. In particular, I would like to acknowledge my peers Nikhil Agarwal and Adam Guren. Their thoughts and advice from many long discussions have helped me through every step of the research process. I would also like to thank my co-authors Guido Imbens, Michal Kolesar, and Thomas Barrios. The third chapter of this disseration is our joint work which my co-authors heavily helped produce. This chapter has been published in the Journal of the American Statistical Association, volume 107, 2012.

Contents

Introduction

This dissertation consists of three independent chapters all related to local labor market economics. Chapter 1 studies the causes and welfare consequences of the increase in geographic sorting of workers by skill from 1980 to 2000. During this time period, the substantial rise in the U.S. college-high school graduate wage gap coincided with an increase in geographic sorting as college graduates increasingly concentrated in high wage, high rent metropolitan areas, relative to lower skill workers. The increase in wage inequality may not reflect a similar increase in well-being inequality because high and low skill workers increasingly paid different housing costs and consumed different local amenities.

This chapter examines the determinants and welfare implications of the increased geographic skill sorting. I estimate a structural spatial equilibrium model of local labor demand, housing supply, labor supply, and amenity levels. The model allows local amenity and productivity levels to endogenously respond to a city's skill-mix. I identify the model parameters using local labor demand changes driven by variation in cities' industry mixes and interactions of these labor demand shocks with determinants of housing supply (land use regulations and land availability). The GMM estimates indicate that cross-city changes in firms' demands for high and low skill labor were the underlying forces of the increase in geographic skill sorting. An increase in labor demand for college relative to non-college workers increases a city's college employment share, which then endogenously raises the local productivity of all workers and improves local amenities. Local wage and amenity growth generates in-migration, driving up rents. My estimates show that low skill workers are less willing to pay high housing costs to live in high-amenity cities, leading them to elect more affordable, low-amenity cities. I find that the combined effects of changes in cities' wages, rents, and endogenous amenities increased well-being inequality between high school and college graduates by a significantly larger amount than would be suggested by the increase in the college wage gap alone.

Chapter 2 examines the abilities of state and local governments to extract rent from private sector workers by charging high tax rates and spending the revenue on non-social desirable projects, such as excessive government worker wages. Using a spatial equilibrium model where private sector workers are free to migrate across government jurisdictions, I show that private sector workers' migration elasticity with respect to local taxes determines the magnitude of rent extraction by rent seeking state and local governments. Since private sector workers "vote with their feet" by migrating out of rent extractive areas, governments trade off the benefits a higher tax rate with the cost of a smaller population to tax.

Variation in areas' housing supply elasticities differentially restrains governments' abilities to extract rent from private sector workers. The incidence of a tax increase falls more on local housing prices in a less housing elastic area, leading to less out-migration. Thus, governments in less housing elastic areas can charge higher taxes without worry of shrinking their tax bases. I test the model's predictions by using worker wage data from the CPS-MORG. I find the public-private sector wage gap is higher in areas with less elastic housing supplies. This fact holds both within state across metropolitan areas for local government workers and between states for state government workers.

In Chapter 3, which is joint work with Guido Imbens, Michal Kolesar, and Thomas Barrios, we examine the standard practice in regression analyses of allowing for clustering in the error covariance matrix when the explanatory variable of interest varies at a more aggregate level (e.g., the state level) than the units of observation (e.g., individuals). Often, however, the structure of the error covariance matrix is more complex, with correlations not vanishing for units in different clusters. Here we explore the implications of such correlations for the actual and estimated precision of least squares estimators. Our main theoretical result is that with equal-sized clusters, if the covariate of interest is randomly assigned at the cluster level, only accounting for non-zero covariances at the cluster level, and ignoring correlations between clusters as well as differences in within-cluster correlations, leads to valid confidence intervals. However, in the absence of random assignment of the covariates, ignoring general correlation structures may lead to biases in standard errors. We illustrate our findings using the 5% public use census data. Based on these results we recommend that researchers as a matter of routine explore the extent of spatial correlations in explanatory variables beyond state level clustering.

Chapter 1
The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000

1.1 Introduction

The dramatic increase in the wage gap between high school and college graduates over the past three decades has been accompanied by a substantial increase in geographic sorting of workers by skill.[1] Metropolitan areas which had a disproportionately high share of college graduates in 1980 further increased their share of college graduates from 1980 to 2000. Increasingly high skill cities also experienced higher wage and housing price growth than less skilled cities (Moretti (2004b), Shapiro (2006)). Moretti (2012) coins this phenomenon "The Great Divergence."

These facts call into question whether the increase in the college wage gap reflects a similar increase in the college well-being gap. Since college graduates increasingly live in areas with high housing costs, local price levels might offset some of the consumption benefits of their high wages, making the increase in wage inequality overstate the increase in consumption or well-being inequality (Moretti (2011b)). Alternatively, high housing cost cities may offer workers desirable amenities, compensating them for high house prices, and possibly increasing the well-being of workers in these cities. The welfare implications of the increased geographic skill sorting depend on why high and low skill workers increasingly chose to live in different cities.

This paper examines the determinants of high and low skill workers' choices to increasingly segregate themselves into different cities and the welfare implications of these choices. By estimating a structural spatial equilibrium model of local labor demand, housing supply,

[1] This large increase in wage inequality has led to an active area of research into the drivers of changes in the wage distribution nationwide. See Goldin and Katz (2007) for a recent survey.

labor supply, and amenity levels in cities, I show that changes in firms' relative demands for high and low skill labor across cities, due to local productivity changes, were the underlying drivers of the differential migration patterns of high and low skill workers.[2] Despite local wage changes being the initial cause of workers' migration, I find that cities which attracted a higher share of college graduates endogenously became more desirable places to live and more productive for both high and low skill labor. The combination of desirable wages and amenities made college workers willing to pay high housing costs to live in these cities. While lower skill workers also found these areas' wages and amenities desirable, they were less willing to pay high housing costs, leading them to choose more affordable cities. Overall, I find that the welfare effects of changes in local wages, rents, and endogenous amenities led to an increase in well-being inequality between college and high school graduates which was *significantly larger* than would be suggested by the increase in the college wage gap alone.

To build intuition for this effect, consider the metropolitan areas of Detroit and Boston. The economic downturn in Detroit has been largely attributed to decline of auto manufacturing (Martelle (2012)), but the decline goes beyond the loss of high paying jobs. In 2009, Detroit public schools had the lowest scores ever recorded in the 21-year history of the national math proficiency test (Winerip (2011)). Historically, the Detroit school district had not always been in such a poor state. In the early 20th century, when manufacturing was booming, Detroit's public school system was lauded as a model for the nation in urban education (Mirel (1999)).

By comparison, Boston has increasingly attracted high skill workers with its cluster of biotech, medical device, and technology firms. In the mid 1970s, Boston public schools were declining in quality, driven by racial tensions from integrating the schools (Cronin (2011)). In 2006, however, the Boston public school district won the Broad Prize, which honors the urban school district that demonstrates the greatest overall performance and improvement

[2] Work by Berry and Glaeser (2005) and Moretti (2011b) come to similar conclusions. Berry and Glaeser (2005) consider the role of entrepreneurship in cities. Moretti (2011b) analyzes the differential labor demands for high and low skill workers across industries.

in student achievement while reducing achievement gaps among low-income and minority students. Similar patterns can be seen in the histories of the Detroit and Boston Symphony Orchestras.[3] The prosperity of Boston and decline of Detroit go beyond jobs and wages, directly impacting the amenities and quality-of-life in these areas.

I illustrate these mechanisms more generally using U.S. Census data by estimating a structural spatial equilibrium model of cities. The setup shares features of the Rosen (1979) and Roback (1982) frameworks, but I extend the model to allow workers to have heterogenous preferences for cities. The fully estimated model allows me to assess the importance of changes in cities' wages, rents, and amenities in differentially driving high and low skill workers to different cities.

I use a static discrete choice setup to model workers' city choices.[4] The model allows workers with different demographics to differentially trade off the relative values of cities' characteristics, leading them to make different location decisions.[5] Workers maximize their utility by living in the city which offers them the most desirable bundle of wages, housing rent, and amenities.

Firms in each city use capital, high skill labor, and low skill labor as inputs into production. High and low skill labor have a constant elasticity of substitution in firms' production functions. I assume capital is sold in a national market, while labor is hired locally in a perfectly competitive labor market. Housing markets differ across cities due to heterogeneity in their elasticity of housing supply.

The key distinguishing worker characteristic is skill, as measured by graduation from a 4-year college. Cities' local productivity levels differ across high and low skill workers, and

[3]The Detroit Symphony Orchestra was one of the top in the nation during the 1950s. More recently, it has defaulted on loans, and is facing a labor dispute over wage cuts driven by decreased ticket sales and corporate donations. (Bennett (2010)) The Boston Symphony Orchestra, however, continues to be one of the best in the world.

[4]The model could be extended to allow for dynamics, as done by Kennan and Walker (2011) and Bishop (2010). However, panel data is needed to estimate a model of this nature. I focus on the role of preference heterogeneity in determining long run migration patterns, while Kennan and Walker (2011) and Bishop (2010) focus exclusively on high-school graduates and life-cycle migration patterns.

[5]Estimation of spatial equilibrium models when households have heterogeneous preferences using hedonics have been analyzed by Epple and Sieg (1999).

the productivity levels of both high and low skill workers within a city are endogenously impacted by the skill-mix in the city. Thus, changes in the skill-mix of a city will impact local wages both by moving along firms' labor demand curves and by directly impacting worker productivity.

A city's skill-mix is also allowed to influence local amenity levels, both directly, as more educated neighbors may be desirable, and indirectly by improving a variety of city amenities (Chapter 5 in Becker and Murphy (2000)). Indeed, observable amenities such as bars and restaurants per capita, crime rates, and pollution levels improve in areas with larger college populations and decline in areas with larger non-college populations. I use the ratio of college to non-college employees in each city as a unidimensional index for all amenities that endogenously respond to the demographics of cities' residents.

Workers' preferences for cities are estimated using a two-step estimator, similar to the methods used by Berry, Levinsohn, and Pakes (2004) and the setup proposed by McFadden (1973). In the first step, a maximum likelihood estimator is used to identify how desirable each city is to each type of worker, on average, in each decade, controlling for workers' preferences to live close to their state of birth. The utility levels for each city estimated in the first step are used in the second step to estimate how workers trade off wages, rents, and amenities when selecting a location to live. The second step of estimation uses a simultaneous equation non-linear GMM estimator. Moment restrictions on workers' preferences are combined with moments identifying cities' labor demand and housing supply curves. These moments are used to simultaneously estimate local labor demand, housing supply, and labor supply to cities.

The model is identified using local labor demand shocks driven by the industry mix in each city and their interactions with local housing supply elasticities. Variation in productivity changes across industries differentially impact cities' local labor demand for high and low skill workers based on the industrial composition of the city's workforce (Bartik (1991)). I measure exogenous local productivity changes by interacting cross-sectional differences in industrial employment composition with national changes in industry wage levels separately for high and low skill workers.

I allow cities' housing supply elasticities to vary based on geographic constraints on developable land around a city's center and land-use regulations (Saiz (2010), Gyourko, Saiz, and Summers (2008)). A city's housing supply elasticity will influence the equilibrium wage, rent, and population response to the labor demand shocks driven by industrial labor demand changes.

Workers' migration responses to changes in cities' wages, rents, and endogenous amenities driven by the Bartik labor demand shocks and the interactions of these labor demand shocks with housing supply elasticity determinants identify workers' preferences for cities' characteristics. Housing supply elasticities are identified by the response of housing rents to the Bartik shocks across cities.

The interaction of the Bartik productivity shocks with cities' housing markets identifies the labor demand elasticities. The wage differences, driven by the productivity shocks, induce workers to migrate to cities which offer more desirable wages. The migration drives demand in the local housing markets, which impacts house prices, as determined by the elasticity of housing supply. Heterogeneity in housing supply elasticity leads to differences in population changes, in response to a given Bartik shock. For a given size labor demand shock, fewer workers will migrate to a city with a less elastic housing supply because rents increase more than in a more elastic city. Thus, the *interaction* of Bartik shocks with measures of housing supply elasticity creates variation in high and low skill local populations which is independent of unobserved local productivity changes, which can identify labor demand elasticities.

The parameter estimates of workers' preferences show that while both college and non-college workers find higher wages, lower rents, and higher amenity levels desirable, high skill workers' demand is relatively more sensitive to amenity levels and low skill workers' demand is more sensitive to wages and rents.[6] The labor demand estimates show that increases

[6]These results are consistent with a large body of work in empirical industrial organization which finds substantial heterogeneity in consumers' price sensitivites. A Consumer's price sensitivity is also found to be

in the college employment ratio leads to productivity spillovers on both college and non-college workers. Combining the estimates of firms' elasticity of labor substitution with the productivity spillovers, I find that an increase in a city's college worker population raises *both* local college and non-college wages. Similarly, an increase in a city's non-college worker population decreases college and non-college wages.

Using the estimated model, I decompose the changes in cities' college employment ratios into the underlying changes in labor demand, housing supply, and labor supply to cities. I show that when a city's productivity gap between high and low skill workers exogenously increases, the local wage gap between these workers increases. If the migration responses to these wage changes lead to an increase in the local share of college workers, the wages of all workers will further increase beyond the initial effect of the productivity change due to the combination of endogenous productivity changes and shifts along firms' labor demand curves.

In addition to raising wages, an increase in a city's college employment ratio leads to local amenity improvements. The combination of desirable wage and amenity growth for all workers causes large amounts of in-migration, as college workers are particularly attracted by desirable amenities, while low skill workers are particularly attracted by desirable wages. The increased housing demand in high college share cities leads to large rent increases. Since low skill workers are more price sensitive, the increases in rent disproportionately discourage low skill workers from living in these high wage, high amenity cities. Lower skill workers are not willing to pay the "price" of a lower real wage to live in high amenity cities. Thus, in equilibrium, college workers sort into high wage, high rent, high amenity cities.

I use the model estimates to quantify the change in well-being inequality. I find the welfare impacts due to wage, rent, and endogenous amenity changes from 1980 to 2000 led to an increase in well-being inequality equivalent to *at least* a 24 percentage point increase in the college wage gap, which is 20% *more* than the actual increase in the college wage gap.

closely linked to his income. See Nevo (2010) for a review of this literature.

In other words, the additional utility college workers gained from of being able to consume more desirable amenities made them better off relative to high school graduates, *despite* the high local housing prices.

This paper is related to several literatures. Most closely related to this paper is work studying how local wages, rents, and employment respond to local labor demand shocks (Topel (1986), Bartik (1991), Blanchard and Katz (1992), Saks (2008), Notowidigdo (2011). See Moretti (2011a) for a review.) Traditionally, this literature has only allowed local labor demand shocks to influence worker migration through wage and rents changes.[7] My results suggest that endogenous local amenity changes are an important mechanism driving workers' migration responses to local labor demand shocks.

A small and growing literature has considered how amenities change in response the composition of an area's local residents (Chapter 5 in Becker and Murphy (2000), Bayer, Ferreira, and McMillan (2007), Card, Mas, and Rothstein (2008), Guerrieri, Hartley, and Hurst (2011), Handbury (2012)). Work by Handbury (2012) studies the desirability and prices of grocery products for sale across cities. Her work finds that higher quality products (an amenity) are more available in cities with higher incomes per-capita, but these areas also have higher prices for groceries. The paper finds higher income households are more willing to pay for grocery quality, leading them to prefer the high price, high quality grocery markets, relative to lower income households. I find a similar relationship for amenities and local real wages.

My findings also relate to the literature studying changes in the wage structure and inequality within and between local labor markets (Berry and Glaeser (2005), Beaudry, Doms, and Lewis (2010), Moretti (2011b), Autor and Dorn (2012), Autor, Dorn, and Hanson (2012)). Most related to this paper is Moretti (2011b), who is the first to show the importance of accounting for the diverging location choices of high and low skill workers when measuring both real wage and well-being inequality changes. Another strand of this literature, most

[7] Notowidigdo (2011) allows government social insurance programs in a city to endogenously respond to local wages, which is one of many endogenous amenity changes.

specifically related to my labor demand estimates, studies the impact of the relative supplies of high and low skill labor on high and low skill wages (Katz and Murphy (1992), Card and Lemieux (2001), Card (2009)). Card (2009) estimates the impact of local labor supply on local wages in cities. My paper presents a new identification strategy to estimate city-level labor demand and allows for endogenous productivity changes.

This paper is also related to the literature on the social returns to education (Acemoglu and Angrist (2001), Moretti (2004c), Moretti (2004a)) and work studying the determinants of economic growth in cities (Glaeser et al. (1992), Glaeser, Scheinkman, and Shleifer (1995), Shapiro (2006)). By using the interaction of local labor productivity shocks with housing supply elasticities as instruments for education differences across cities, I provide a new identification strategy for measuring the impact of an increase in a city's education level on the wages for all workers. Further, my findings show that an increase in a city's education level also spills over onto all workers' well-being through endogenous amenity changes.

The labor supply model and estimation draws on the discrete choice methods developed in empirical industrial organization to estimate consumers demand for products (McFadden (1973), Berry, Levinsohn, and Pakes (1995), Berry, Levinsohn, and Pakes (2004)). These methods have been applied to estimate households' preferences for neighborhoods by Bayer, Ferreira, and McMillan (2007). My paper adapts these methods to estimate the determinants of workers' labor supply to cities.[8]

The rest of the paper proceeds as follows. Section 2 discusses the data. Section 3 presents reduced form facts. Section 4 lays out the model. Section 5 discusses the estimation techniques. Section 6 presents parameter estimates. Section 7 discusses the estimates. Section 8 analyzes the determinants of cities' college employment ratio changes. Section 9 presents welfare implications. Section 10 concludes.

[8]Similar methods have been used by Bayer, Keohane, and Timmins (2009), Bishop (2010), and Kennan and Walker (2011) to estimate workers' preferences for cities. However, these papers do not allow local wages and rents to be freely correlated with local amenities. Bayer, Keohane, and Timmins (2009) focuses on the demand for air quality, while Bishop (2010) and Kennan and Walker (2011) study the dynamics of migration over the life-cycle exclusively for high school graduates.

1.2 Data

The paper uses the 5 percent samples of the U.S. Censuses from the 1980, 1990, and 2000 Integrated Public Use Microdata Series (IPUMS) (Ruggles et al. (2010)). These data provide individual level observations on a wide range of economic and demographic information, including wages, housing costs, and geographic location of residence. All analysis is restricted to 25-55 year-olds who report positive wage earnings. The geographical unit of analysis is the metropolitan statistical area (MSA) of residence, however I interchangeably refer to MSAs as cities. The Census includes 218 MSAs consistently across all three decades of data. Rural households are not assigned to an MSA in the Census. To incorporate the choice to live in rural areas, all areas outside of MSAs within each state are grouped together and treated as additional geographical units.[9]

The IPUMS data are also used to construct estimates of local area wages, population, and housing rents in each metropolitan and rural area. The advantage of the Census data is the ability to construct MSA-level measures disaggregated by education level and other demographics. A key city characteristic I focus on is the local skill mix of workers. I define high skill or college workers as workers with positive wage earnings and who have completed at least 4 years of college, while all other workers with positive wage earnings are classified as low skill or non-college. Throughout the paper, the local college employment ratio is measured by the ratio of college employees to non-college employees working within a given MSA. I use a two skill group model since the college/non-college division is where the largest divide in wages across education is seen, as found by Katz and Murphy (1992) and Goldin and Katz (2008).

Cities' high and low skill wages are measured by the average local hourly wage for each skill level in each city. The sample used to measure local wages is restricted to workers employed at least 35 hours a week and 48 weeks per year to get a standardized wage mea-

[9] Households living in MSAs which the census does not identify in all 3 decades are included as residents of states' rural areas.

sure. Local rents are measured by the average household rent in the city, where rent is measured using both reported rents, as well as imputed from reported house price values.[10] For additional city characteristics, I supplement these data with Saiz (2010)'s measures of geographic constraints and land use regulations to measure differences in housing supply elasticities. Table 1.1 reports summary statistics for these variables.

When estimating workers' city choices, I assume heads of household make city decisions for the entire household. If a household has multiple working adults between ages 25 and 55, I assume all household members migrate with the household head and supply labor to the labor market of the household's residence.[11] Thus, local labor supply and a city's college employment ratio are calculated using all household members who have positive wage earnings and are between ages 25 and 55, while the sample used to estimate workers' preferences for cities is restricted only to the household heads. See Appendix Table 1.1 for summary statistics of this head-of-household sample. Appendix A contains remaining data and measurement details.

[10] In Section 6 I consider defining the local wage and rents measures in a number of ways such as using hedonic adjustments to wages and rents, dropping imputed housing rents, and allowing high and low skill workers to face different local rents within a city. I re-estimate the model using these alternative definitions.

[11] While I abstract away from the role joint location decisions of dual earner households, work by Costa and Kahn (2000) shows that couples where both partners have high powered careers are particularly attracted to large, high skill cities. The labor markets of these cities are more likely to offer good jobs for both household members' careers. Costa and Kahn (2000) find that the increase in women's labor supply over this period has also contributed to increase in geographic skill sorting as large, high skill cities are increasingly attractive to high skill couples as the number of dual career couples increases.

Table 1.1: Summary Statistics

	Obs	Mean	Std. Dev.	Min	Max
A. Levels					
1980					
Ln College Wage	268	6.657	0.102	6.433	7.066
Ln Non-College Wage	268	6.326	0.127	5.919	6.703
Ln Rent	268	6.544	0.163	6.109	7.104
Ln College Employment Ratio	268	-1.247	0.364	-2.092	-0.236
1990					
Ln College Wage	268	6.792	0.119	6.470	7.287
Ln Non-College Wage	268	6.369	0.126	5.927	6.697
Ln Rent	268	6.535	0.286	6.033	7.569
Ln College Employment Ratio	268	-1.113	0.403	-2.061	0.066
2000					
Ln College Wage	268	6.847	0.131	6.536	7.585
Ln Non-College Wage	268	6.390	0.113	5.939	6.699
Ln Rent	268	6.609	0.247	6.142	7.721
Ln College Employment Ratio	268	-1.006	0.431	-1.903	0.500
B. Changes					
	Obs	Mean	Std. Dev.	Min	Max
Δ Ln College Wage	536	0.095	0.077	-0.127	0.339
Δ Ln Non-College Wage	536	0.032	0.070	-0.215	0.305
Δ Ln Rent	536	0.032	0.170	-0.417	0.660
Δ Ln College Emp Ratio	536	0.142	0.133	-0.304	0.624
Δ Ln College Population	536	0.314	0.277	-0.820	1.478
Δ Ln Non-College Population	536	0.172	0.273	-1.052	1.249
College Bartik	386	-0.003	0.007	-0.027	0.013
Non-College Bartik	386	0.018	0.008	-0.001	0.058
Total Bartik	386	0.013	0.007	-0.003	0.052
C. Housing Supply Elasticity Measures					
Land Unavailability	193	0.254	0.215	0.005	0.860
Land Use Regulation	193	-0.032	0.733	-1.677	2.229

Notes: Summary statistics for changes pool decadal changes in wages, rents, population from 1980-1990 and 1990-2000. The Bartik shocks are also measured across decades. The sample reported for MSAs' wages, rents, and population include a balanced panel of MSAs and rural areas which the 1980, 1990, and 2000 Censuses cover. The sample used for statistics on the Bartik shocks and housing supply elasticity characteristics are the balanced panel of MSAs which also contain data on housing supply elasticity characteristics. Wages, rents, and population are measured in logs. Bartik shocks use national changes in industry wages weighted by the share of a cities work force employed in that industry. College Bartik uses only wages and employment shares from college workers. Non-College Bartik uses non-college workers. Aggregate Bartik combines these. Land Unavailability measures the share of land within a 50Km radius of a city's center which cannot be developed due to geographical land constraints. Land use regulation is an index of Land-Use regulation policies within an MSA. College employment ratio is defined as the ratio of number of employed workers in the city with a 4 year college degree to the number of employed lower skill workers living in the city. See data appendix for further details.

1.3 Reduced Form Facts

From 1980 to 2000, the distribution of college and non-college workers across metropolitan areas was diverging. Specifically, a MSA's share of college graduates in 1980 is positively associated with larger growth in its share of college workers from 1980 to 2000. Figure 1.1 shows a 1% increase in a city's college employment ratio in 1980 is associated with a .22% larger increase in the city's college employment ratio from 1980 to 2000. The diverging college employment ratios across cities can also be seen in Figure 1.2, which shows the distribution of college employment ratios across cities. The standard deviation was 0.116 in 1980 and increased to 0.192 in 2000. These facts have also been documented by Moretti (2004b), Berry and Glaeser (2005), and Moretti (2011b).

The distribution and divergence of worker skill across cities are strongly linked to cities' wages and rents. Figure 1.3 shows that a higher college employment ratio is positively associated with higher rents across cities in both 1980 and 2000. Additionally, as the distribution of skill-mix across cities has spread out from 1980 to 2000, Figure 1.3 shows that there has been a similar spreading out of the distribution of rents. The third panel of Figure 1.3 plots changes in local rents against changes in college employment ratios from 1980 to 2000. A 1% increase in local college employment ratio is associated with a .63% increase in local rents. Further, the relationship between rent and college employment ratio is extremely tight. In 2000, variation in cities' college employment ratios can explain 74% of the variation of rent across cities. Overall, these figures show that college graduates are increasingly paying higher housing costs than lower skill workers and that local housing costs are strongly related to a city's skill-mix.

Figure 1.1: ΔCollege Employment Ratio: 1980-2000 vs. College Employment Ratio: 1980

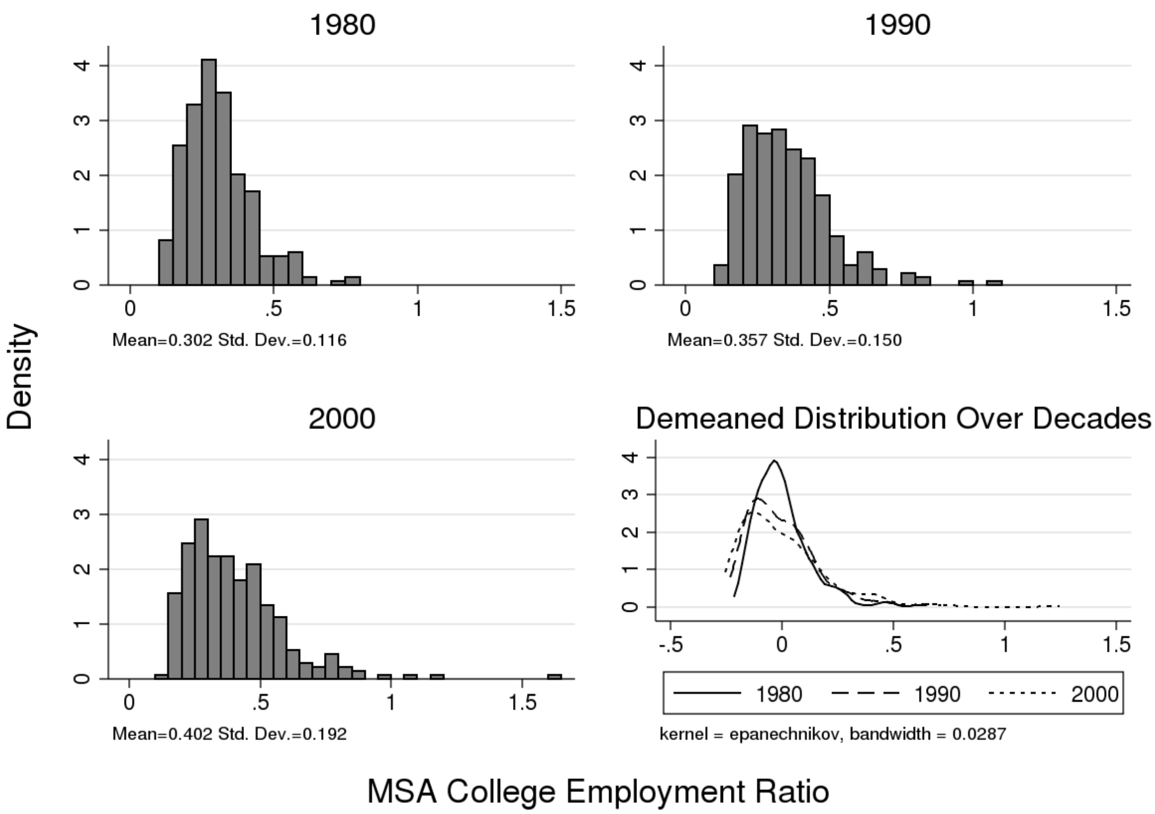


Figure 1.2: Distribution of Cities' College Employment Ratios: 1980-2000

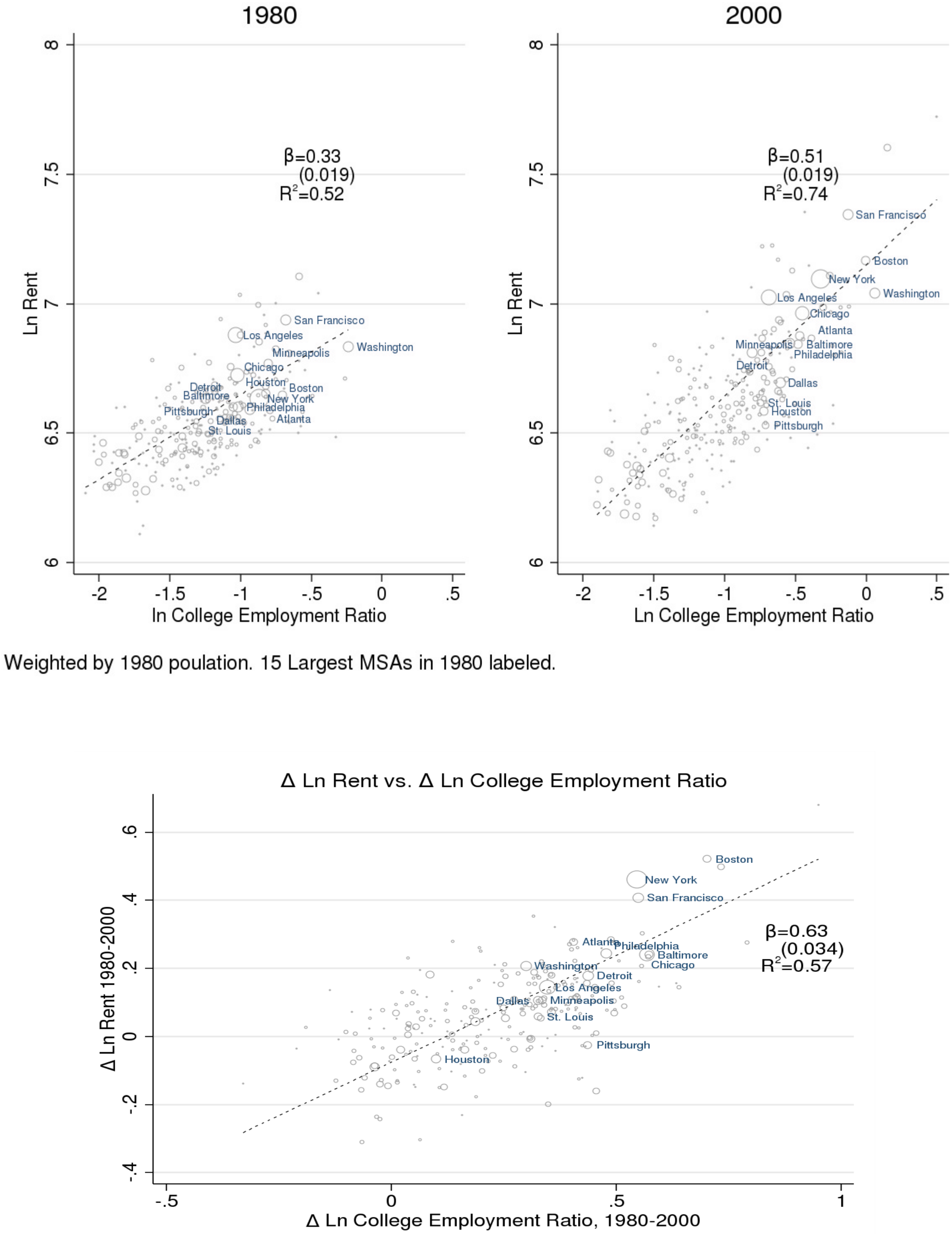


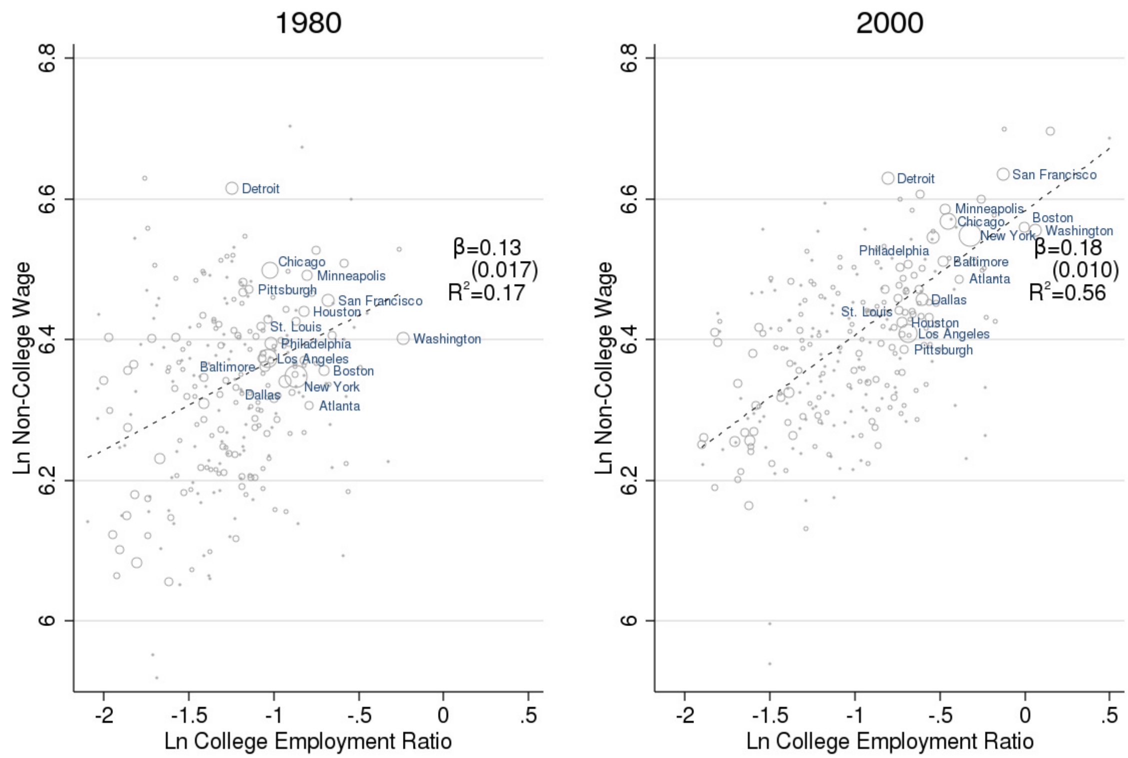
Figure 1.3:Rents vs. College Employment Ratios: Levels & Changes

Cities' local wages have a similar, but less strong relationship with the local college employment ratio. Figure 1.4 shows that local non-college wages are positively associated with local college employment ratios in both 1980 and 2000. The third panel of Figure 1.4 plots changes in local college employment ratios against changes in local non-college wages from 1980 to 2000. A 1% increase in college employment ratio is associated with a 0.18% increase in non-college wages. In 2000, the variation in college employment ratios across cities can explain 56% of the variation in local non-college wages. These figures show that low skill workers were both initially and increasingly concentrating in low wage cities.

Similarly, Figure 1.5 shows that college wages are higher in high college employment ratio cities in both 1980 and 2000. Looking at this relationship in changes, panel three of Figure 1.5 shows that a 1% increase in a city's college employment ration is associated with a 0.29% increase in college wages. Additionally, college employment ratio can explain 68% of the variation in local college wages in 2000. College workers are increasingly concentrating in high wage cities and high skill wages are closely linked to a city's skill-mix. Moretti (2011b) has also documented this set of facts and refers to them as "The Great Divergence" in Moretti (2012).

The polarization of skill across cities coincided with a large, nationwide increase in wage inequality. Table 1.9, along with a large body of literature, documents that the nationwide average college-high school graduate wage gap has increased from 42% in 1980 to 62% in 2000.[12] Moretti (2011b) points out that the increase in geographic skill sorting calls into question whether the rise in wage inequality represents a similar increase in well-being or "utility" inequality between college and high school graduates. Since college workers are increasingly living in high cost cities, the high local prices may diminish the consumption value of their wages relative to local prices faced by lower skilled workers. Looking only at changes in workers' wages and rents, it appears the differential increases in housing costs across cities disproportionately benefited low skill workers.

[12] This is estimated by a standard Mincer regression using individual 25-55 year old full time full year workers' hourly wages and controls for sex, race dummies, and a cubic in potential experience.



Weighted by 1980 poulation. 15 Largest MSAs in 1980 labeled.

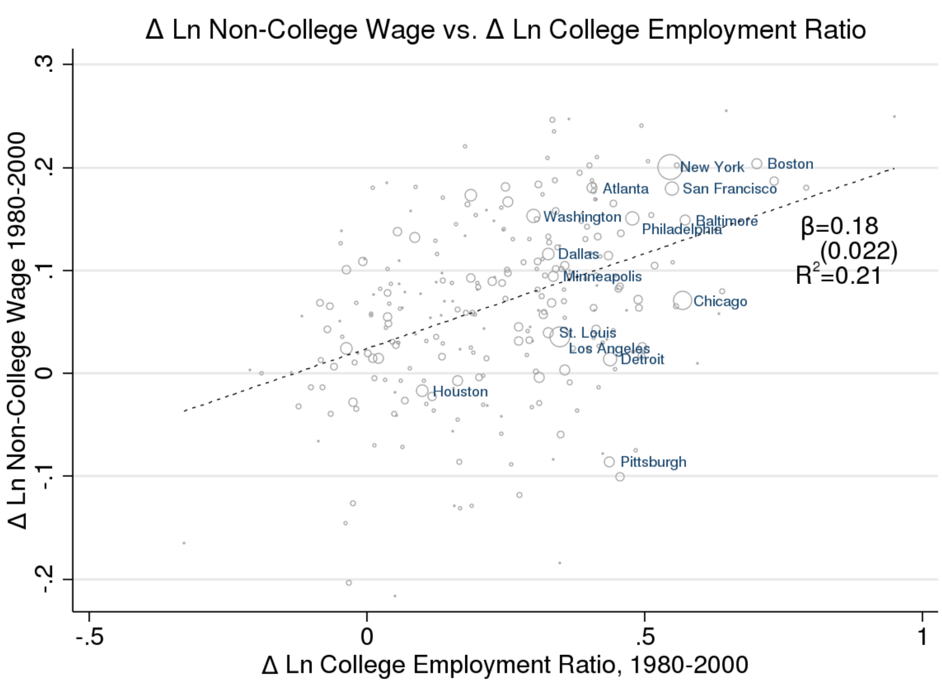
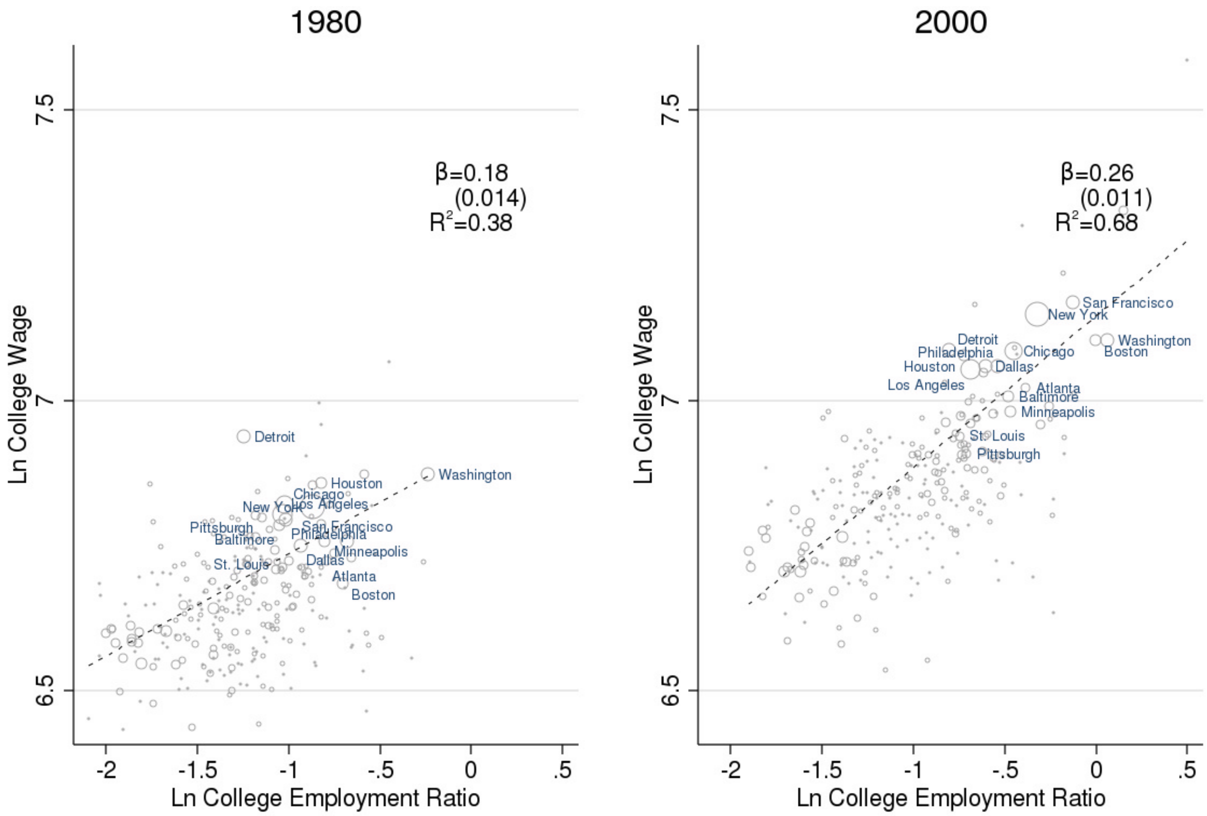


Figure 1.4: Non-College Wages vs. College Employment Ratios: Levels & Changes



Weighted by 1980 poulation. 15 Largest MSAs in 1980 labeled.

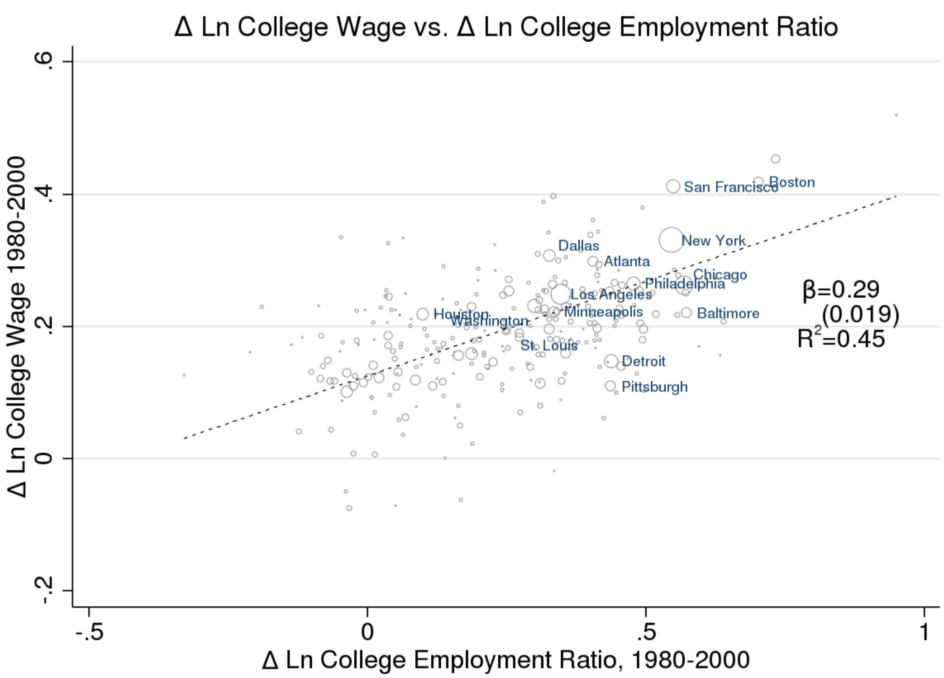


Figure 1.5: College Wages vs. College Employment Ratios: Levels & Changes

However, high skill workers were free to live in more affordable cities, but they chose not too. As Moretti (2011b) notes, the welfare impacts of the changes in rents across cities depends crucially on why high and low skill workers elected to live in high and low housing price cities.

While wage differences across cities are a possible candidate for driving high and low skill workers to different cities, it is possible that the desirability of cities' local amenities differentially influenced high and low skill workers' city choices. If college workers elected to live in high wage, high housing cost cities because they found the local amenities desirable, then the negative welfare impact of high housing costs would be offset by the positive welfare impact of being able to consume amenities.

Table 1.2 presents the relationships between changes in cities' college employment ratios from 1980 to 2000 and changes in a large set of local amenities.[13] These results show that increases in cities' college employment ratios are associated with larger increases in apparel stores per capita, eating and drinking places per capita, dry cleaners per capita, and movie theaters per capita, as well as larger decreases in pollution levels. There are similar point estimates for book stores per capita, museum and art galleries per capita, and property crime rates, but the estimates are not statistically significant. Changes in grocery stores per capita are negatively associated with change in a city's college employment ratio and property crime rates are positively associated, however these estimates are not statistically significant. While this is not an exhaustive set of amenities, it appears that the cities which increased their skill-mix not only experienced larger increases in wages and rents, but also had larger increases in amenities. Additionally, stores per capita, crime, and air quality could be endogenous outcomes. Just as wages and rents, are endogenously determined in the labor and housing markets, amenity levels could also respond to characteristics of a city.

[13] Appendix Table 2 presents similar regressions of observable amenity changes on changes in cities' college and non-college populations. These regressions show that when high and low skill population changes are seperatedly measured, amenities tend to improve with high skill population growth and decline with low skill population growth.

Table 1.2: MSA College Ratio Changes on Amenity Changes: 1980-2000

	(1)	(2)	(3)	(4)	(5)
	Grocery Stores per 1000 Residents	Apparel Stores per 1000 Residents	Eating and Drinking Places per 1000 Residents	Book Stores per 1000 Residents	Dry Cleaners per 1000 Residents
Δ College Emp Ratio	-0.0835	0.430***	0.199***	0.122	0.402***
	[0.0866]	[0.0898]	[0.0524]	[0.159]	[0.138]
Constant	0.236***	-0.432***	0.212***	0.153***	-0.0116
	[0.0300]	[0.0311]	[0.0182]	[0.0552]	[0.0479]
Observations	218	218	218	218	218
R-Squared	0.004	0.095	0.062	0.003	0.037
	(6)	(7)	(8)	(9)	(10)
	Movie Thearters per 1000 Residents	Museums and Art Galleries per 1000 Residents	Propery Crimes per 1000 Residents	Violent Crimes Per 1000 Residents	EPA Air Quality Index
Δ College Emp Ratio	0.331*	0.391	-0.177	0.139	-0.300*
	[0.172]	[0.240]	[0.122]	[0.161]	[0.169]
Constant	-0.865***	0.679***	0.200***	-0.0425	-0.0540
	[0.0596]	[0.0864]	[0.0426]	[0.0562]	[0.0614]
Observations	218	174	215	215	177
R-Squared	0.017	0.015	0.010	0.003	0.018

Notes: Standard erorrs in brackets. Changes measured between 1980-2000. All variables are measured in logs. College employment ratio is defined as the ratio of number of employed college workers to the number of employed lower skill workers living in the city. Retail and local service establishments per capita data come from County Business Patters 1980, 2000. Crime data is from the FBI. Air Quality Index in from the EPA. Higher values of of the air quality index indicate more pollution. *** p<0.01, ** p<0.05, * p<0.1

The amenity regressions in Table 1.2 do not tell us whether the link between amenities and skill-mix is driven by high skill workers disproportionately migrating to high amenity cities or amenities endogenously improving when a city's college employment ratio increases.

To understand why college workers elected to live in high wage, high rent, high amenity cities, one needs causal estimates of workers' migration elasticities with respect to each one of these city characteristics. Further, the impact of changes in high and low skill worker populations on wages, rents, and amenities depends on the elasticities of local housing supply, local labor demand, and amenity supply. To understand how this set of supply and demand elasticities interact and lead to equilibrium outcomes, it useful to view these elasticities through the lens of a structural model. Further, using a utility microfoundation of workers' city choices allows migration elasticities to be mapped to utility functions. The estimated parameters can then be used to quantify the welfare impacts of changes in wage, rents, and amenities.

1.4 An Empirical Spatial Equilibrium Model of Cities

This section presents a spatial equilibrium model of local labor markets that captures how wages, housing rents, amenities, and population are determined in equilibrium. The setup shares many features of the Rosen (1979) and Roback (1982) frameworks, but I enrich the model to more flexibly allow for heterogeneity in workers' preferences, cities' productivity levels, and cities' housing supplies. I also allow city productivity and amenities to be endogenously determined by the types of workers that choose to live in the city.

The model admits workers of different types based on their education level, race, immigrant status, and state of birth. Workers of different types differentially trade off the relative value of city characteristics, leading them to make different location decisions. Workers maximize their utility by living in the city which offers them the most desirable bundle of wages, housing rent, and amenities.

The key distinguishing worker characteristic is skill. Cities' local productivity levels differ across high and low skill workers, and workers of different skill are imperfect substitutes into production. Further, the productivity levels of both high and low skill workers within a city are endogenously influenced by the skill-mix in the city.

The skill mix of cities also partially determines cities' amenity levels. Many amenities likely respond to the college employment ratio in the city, such as education quality, the quality of the local goods and services markets, as well as crime. I use the college employment ratio as an index to represent the overall level of all of these amenities.

Housing markets differ across cities due to heterogeneity in their elasticity of housing supply. MSAs' housing supply elasticities differ based on geographic constraints on developable land around a city center, such as bodies of water or wetlands (Saiz (2010)). Additionally, land-use regulations also play an important role in housing supply elasticities by restricting new construction, leading to less new construction in response to population increases (Gyourko, Saiz, and Summers (2008), Saks (2008)).

The sections below describe the setup for labor demand, housing supply, worker labor supply to cities, and how they jointly determine the spatial equilibrium across cities.

1.4.1 Labor Demand

Each city, indexed j, has many homogeneous firms, indexed by d, in year t.[14] [15] They produce a homogenous tradeable good using high skill labor (H_{djt}), low skill labor (L_{djt}), and capital

[14] Autor and Dorn (2012) model local labor demand using a two sector model, where one sector produces nationally traded goods, and the other produces local goods. My use of a single tradable sector allows me to derive simple expressions for city-wide labor demand. I do not mean to rule out the importance of local goods production, which is surely an significant driver of low skill worker labor demand.

[15] I model firms as homogenous to derive a simple expression for the city-wide aggregate labor demand curves. Alternatively, one could explicitly model firms' productivities differences across industries to derive an aggregate labor demand curve.

(K_{djt}) according to the production function:

$$\begin{aligned} Y_{djt} &= N_{djt}^{\alpha} K_{djt}^{1-\alpha}, \quad (1.1) \\ N_{djt} &= \left(\theta_{jt}^{L} L_{djt}^{\rho} + \theta_{jt}^{H} H_{djt}^{\rho}\right)^{\frac{1}{\rho}} \\ \theta_{jt}^{L} &= \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma_L} \exp\left(\varepsilon_{jt}^{L}\right) \quad (1.2) \\ \theta_{jt}^{H} &= \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma_H} \exp\left(\varepsilon_{jt}^{H}\right) \quad (1.3) \end{aligned}$$

The production function is Cobb-Douglas in the labor aggregate N_{djt} and capital, K_{djt}.[16] This setup implies that the share of income going to labor is constant and governed by α.[17] The labor aggregate hired by each firm, N_{djt}, combines high skill labor, H_{djt}, and low skill labor, L_{djt}, as imperfect substitutes into production with a constant elasticity of substitution, where the elasticity of labor substitution is $\frac{1}{1-\rho}$. The large literature on understanding changes in wage inequality due to the relative supply of high and low skill labor uses this functional form for labor demand, as exemplified by Katz and Murphy (1992).

Cities' production functions differ based on productivity. Each city's productivity of high skill workers is measured by θ_{jt}^{H} and low skill productivity is measured by θ_{jt}^{L}. Equations (1.2) and (1.3) show that local productivity is determined by exogenous and endogenous factors. Exogenous productivity differences across cities and worker skill are measured by $\exp\left(\varepsilon_{jt}^{L}\right)$ and $\exp\left(\varepsilon_{jt}^{H}\right)$. Exogenous differences in productivity across cities could be proximity to a port or coal mine, as well as differences in industry mix of firms in the area.

Additionally, productivity is endogenously determined by the skill mix in the city. Equa-

[16] The model could be extended to allow local housing (office space) to be an additional input into firm production. I leave this to future work, as it would require a more sophisticated model of how workers and firms compete in the housing market. Under the current setup, if office space is additively separable in the firm production function, then the labor demand curves are unchanged.

[17] Ottaviano and Peri (2012) explicitly consider whether Cobb-Douglas is a good approximation to use when estimating labor demand curves. They show that the relative cost-share of labor to income is constant over the long run in the US. This functional form is also often used by the macro growth literature since the labor income share is found to be constant across many countries and time. See Ottaviano and Peri (2012) for further analysis.

tions (1.2) and (1.3) show that the ratio of high to low skill labor working in the city, $\frac{H_{jt}}{L_{jt}}$ differentially impacts high skill and low skill productivity, as measured by γ_H and γ_L, respectively. The literature on the social returns to education has shown that areas with a higher concentration of college workers could increase all workers' productivity through knowledge spillovers. For example, increased physical proximity with educated workers may lead to better sharing of ideas, faster innovation, or faster technology adoption.[18] Productivity may also be influenced by endogenous technological changes or technology adoption, where the development or adoption of new technologies is targeted at new technologies which offer the most profit (Acemoglu (2002), Beaudry, Doms, and Lewis (2010)).

Since there are a large number of firms and no barriers to entry, the labor market is perfectly competitive and firms hire such that wages equal the marginal product of labor. A frictionless capital market supplies capital perfectly elastically at price κ_t, which is constant across all cities. Thus, each firm's demand for labor and capital is:

$$\begin{aligned} W_{jt}^{H} &= \alpha N_{djt}^{\alpha-\rho} K_{djt}^{1-\alpha} H_{djt}^{\rho-1} \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma_H} \exp\left(\varepsilon_{jt}^{H}\right), \\ W_{jt}^{L} &= \alpha N_{djt}^{\alpha-\rho} K_{djt}^{1-\alpha} L_{djt}^{\rho-1} \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma L} \exp\left(\varepsilon_{jt}^{L}\right), \\ \kappa_t &= N_{djt}^{\alpha} K_{djt}^{-\alpha} \left(1-\alpha\right). \end{aligned}$$

Note that the productivity spillovers are governed by the city-level college employment ratio, so the hiring decision of each individual firm takes the city-level college ratio as given when making their hiring decisions.

Since capital is in equilibrium, it can freely adjust to changes in the labor quantities within cities, over time.[19] Firm-level labor demand translates directly to city-level aggregate labor

[18] See Moretti (2011a) for a literature review of these ideas.

[19] An alternative assumption would be to assume that capital is fixed across areas, leading to downward slopping aggregate labor demand within each city. Ottaviano and Peri (2012) explicitly consider the speed of capital adjustment to in response to labor stock adjustment across space. They find the annual rate of capital adjustment to be 10%. Since my analysis of local labor markets is across decades, I assume capital is in equilibrium.

demand since firms face constant returns to scale production functions and share identical production technology. Substituting for equilibrium levels of capital, the city-level log labor demand curves are:

$$w_{jt}^{H}=\ln W_{jt}^{H}=c_{t}+(1-\rho)\ln N_{jt}+(\rho-1)\ln H_{jt}+\gamma_{H}\ln\left(\frac{H_{jt}}{L_{jt}}\right)+\varepsilon_{jt}^{H} \tag{1.4}$$

$$\begin{aligned}
w_{jt}^{L} &= \ln W_{jt}^{L}=c_{t}+(1-\rho)\ln N_{jt}+(\rho-1)\ln L_{jt}+\gamma_{L}\ln\left(\frac{H_{jt}}{L_{jt}}\right)+\varepsilon_{jt}^{L} && (1.5)\\
N_{jt} &= \left(\exp\left(\varepsilon_{jt}^{L}\right)\left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma_{L}}L_{jt}^{\rho}+\exp\left(\varepsilon_{jt}^{H}\right)\left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma_{H}}H_{jt}^{\rho}\right)^{\frac{1}{\rho}} && (1.6)\\
c_{t} &= \ln\left(\alpha\left(\frac{(1-\alpha)}{\kappa_{t}}\right)^{\frac{1-\alpha}{\alpha}}\right).
\end{aligned}$$

1.4.2 Housing Supply

Local prices, R_{jt}, are set through equilibrium in the housing market. The local price level represents both local housing costs and the price of a composite local good, which includes goods such as groceries and local services which have their prices influenced by local housing prices. See Appendix A.6 for a full micro-foundation of the local goods market. Inputs into the production of housing include construction materials and land. Developers are price-takers and sell homogenous houses at the marginal cost of production.

$$P_{jt}^{\text{house}}=MC\left(CC_{jt},LC_{jt}\right).$$

The function $MC\left(CC_{jt},LC_{jt}\right)$ maps local construction costs, CC_{jt}, and local land costs, LC_{jt}, to the marginal cost of constructing a home. In the asset market steady state equilibrium, there is no uncertainty and prices equal the discounted value of rents. Local rents are:

$$R_{jt}=\iota_{t}*MC\left(CC_{jt},LC_{jt}\right),$$

where ι_t is the interest rate. Housing is owned by absentee landlords who rent the housing to local residents.

The cost of land LC_{jt} is a function of the population size of the city. As more people move into the city, the developable land in the city becomes more scarce, driving up the price of land.[20]

I parameterize the log housing supply equation as:

$$r_{jt} = \ln(R_{jt}) = \ln(\iota_t) + \ln(CC_{jt}) + \gamma_j \ln(H_{jt} + L_{jt}), \tag{1.7}$$

$$\gamma_j = \gamma + \gamma^{geo} \exp\left(x_j^{geo}\right) + \gamma^{reg} \exp\left(x_j^{reg}\right). \tag{1.8}$$

The elasticity of rent with respect to population $(H_{jt} + L_{jt})$, varies across cities, as measured by γ_j. House price elasticities are influenced by characteristics of the city which impact the availability of land suitable for development. Geographic characteristics, which make land in the city undevelopable, lead to a less elastic housing supply. With less available land around to build on, the city must expand farther away from the central business area to accommodate a given amount of population. x_j^{geo} measures the share of land within 50 km of each city's center which is unavailable for development due to the presence of wetlands, lakes, rivers, oceans, and other internal water bodies as well as share of the area corresponding to land with slopes above 15 percent grade. This measure was developed by Saiz (2010). In equation (1.8), γ^{geo} measures how variation in $\exp\left(x_j^{geo}\right)$ influences the inverse elasticity of housing supply, γ_j.

Local land use regulation has a similar effect by further restricting housing development. Data on municipalities' local land use regulation was collected in the 2005 Wharton Regu-

[20] A full micro-foundation of this assumption can be derived from the Alonso-Muth-Mills model (Brueckner (1987)) where housing expands around a city's central business district and workers must commute from their house to the city center to work. Within-city house prices are set such that workers are indifferent between having a shorter versus longer commute to work. Average housing prices rise as the population grows since the houses on the edge of the city must offer the same utility as the houses closer in. As the city population expands, the edge of the city becomes farther away from the center, making the commuting costs of workers living on the edge higher than those in a smaller city. Since the edge of the city must offer the same utility value as the center of the city, housing prices rise in the interior parts of the city.

lation Survey. Gyourko, Saiz, and Summers (2008) use the survey to produce a number of indices that capture the intensity of local growth control policies in a number of dimensions. Lower values in the Wharton Regulation Index, can be thought of as signifying the adoption of more laissez-faire policies toward real estate development. Metropolitan areas with high values of the Wharton Regulation Index have zoning regulations or project approval practices that constrain new residential real estate development. I use Saiz (2010)'s metropolitan area level aggregates these data as my measure of land use regulation x_j^{reg}. See Table 1.1 for summary statistics of these measures. In equation (1.8), γ^{reg} measures how variation in $\exp\left(x_j^{reg}\right)$ influences the inverse elasticity of housing supply γ_j. γ measures the "base" housing supply elasticity for a city which has no land use regulations and no geographic constraints limiting housing development.

1.4.3 Amenity Supply

Cities differ in the amenities they offer to their residents. I define amenities broadly as all characteristics of a city which could influence the desirability of a city beyond local wages and prices. This includes the generosity of the local social insurance programs as well as more traditional amenities like annual rainfall. All residents within the city have access to these amenities simply by choosing to live there.[21] Some amenity differences are due to exogenous factors such as climate or proximity to the coast. I refer to exogenous amenities in city j in year t by the vector x_{jt}^{A}.

I also consider the utility value one gets from living in a city in or near one's state of birth to be an amenity of the city. Define x_j^{st} as a 50x1 binary vector where each element k is equal to 1 if part of city j is contained in state k. Similarly, define x_j^{div} as a 9x1 binary vector where each element k is equal to 1 if part of city j is contained within Census division k.

[21] See appendix A.6 in which amenities are partially determined by the quality of the local retail market. "Access" to the city's amenities then depends on the purchase choices of the household in the local retail market. In this case, the amenity value represents the indirect utility function which measures the quality of the products purchased by the worker in the local retail market.

Additionally, city amenities endogenously respond to the types of residents who choose to live in the city. I model the level of endogenous amenities to be determined by cities' college employment ratio, $\frac{H_{jt}}{L_{jt}}$. Using the college employment ratio as an index for the level of endogenous amenities is a reduced form for the impact of the distribution of residents' education levels, as well as the impact of residents' incomes on local amenities. The vector of all amenities in the city, A_{jt}, is:

$$A_{jt} = \left(x_{jt}^{A}, x_{j}^{st}, x_{j}^{\text{div}}, \frac{H_{jt}}{L_{jt}} \right).$$

This setup is motivated by work by Guerrieri, Hartley, and Hurst (2011), Handbury (2012), and Bayer, Ferreira, and McMillan (2007). Guerrieri, Hartley, and Hurst (2011) shows that local housing price dynamics suggest local amenities respond to the income levels of residents. Bayer, Ferreira, and McMillan (2007) show that at the very local neighborhood level, households have direct preferences for the race and education of neighboring households. Handbury (2012) shows that cities with higher income per capita offer wider varieties of high quality groceries. The quality of the products available within a city are an amenity.

I approximate these forces by cities' college employment ratios as an index for local endogenous amenity levels. Regressions of changes in observable amenities over time discussed earlier in section 1.3 suggest that amenities are positively associated with a city's college employment, which motivates this setup. For a full microfoundation of how amenities endogenously respond to the skill mix of the city through changes in the desirability of the local retail market, see Appendix A.6.

1.4.4 Labor Supply to Cities

Each head-of-household worker, indexed by i, chooses to live in the city which offers him the most desirable bundle of wages, local good prices, and amenities. Wages in each city differ between college graduates and lower educated workers. A worker of skill level edu living in city j in year t inelastically supplies one unit of labor and earns a wage of W_{jt}^{edu}.

The worker consumes a local good M, which has a local price of R_{jt} and a national good O, which has a national price of P_t, and gains utility from the amenities A_{jt} in the city. The worker has Cobb-Douglas preferences for the local and national good, which he maximizes subject to his budget constraint:

$$\max_{M,O} \ln\left(M^{\zeta_i}\right) + \ln\left(O^{1-\zeta_i}\right) + s_i\left(A_{jt}\right) \tag{1.9}$$
$$s.t.\ P_t O + R_{jt} M \leq W_{jt}^{edu}.$$

Workers differ in the relative taste for national versus local goods, which is governed by ζ_i, where $0 \leq \zeta_i \leq 1$. Workers' relative value of the local versus national good is a function of their demographics, z_i:[22]

$$\zeta_i = \beta^r z_i.$$

z_i is a vector of the workers's demographics, which includes his skill level, as well as his skill level interacted with his race and whether the worker immigrated to the US.[23] Thus, preferences for the local and national good are heterogenous between workers with different demographics, but homogenous within demographic group.

Workers are also heterogenous in how much they desire the local non-market amenities. The function $s_i\left(A_{jt}\right)$ maps the vector of city amenities, A_{jt}, to the worker's utility value for them.

The worker's optimized utility function can be expressed as an indirect utility function

[22] The demographics which determine a household's city preferences are those of the household head. I abstract away from the impact of preferences of non-head-of-household workers on a household's city choice.

[23] Since I focus on the location choices of head-of-households, I do not include gender as characteristic influencing preferences for city. My sample of household heads is mechanically strongly male dominated.

for living in city j. If the worker were to live in city j in year t, his utility V_{ijt} would be:

$$\begin{aligned} V_{ijt} &= \ln\left(\frac{W_{jt}^{edu}}{P_t}\right) - \beta^r z_i \ln\left(\frac{R_{jt}}{P_t}\right) + s_i\left(A_{jt}\right), \\ &= w_{jt}^{edu} - \beta^{\zeta} z_i r_{jt} + s_i\left(A_{jt}\right), \end{aligned}$$

where $w_{jt}^{edu} = \ln\left(\frac{W_{jt}^{edu}}{P_t}\right)$ and $r_{jt} = \ln\left(\frac{R_{jt}}{P_t}\right)$. Since the worker's preferences are Cobb-Douglas, he spends $\beta^r z_i$ share of his income on the local good, and $(1 - \beta^r z_i)$ share of his income on the national good. The price of the national good is measured by the CPI-U index for all goods excluding shelter and measured in real 2000 dollars.

Consumption of amenities is only determined by the city a worker chooses to live in because all workers living in city j have full access to the amenities A_{jt}. Worker i's value of amenities A_{jt} is:

$$s_i\left(A_{jt}\right) = \beta_i^A x_{jt}^A + \beta_i^{\text{col}} \frac{H_{jt}}{L_{jt}} + \beta_i^{st} x_j^{st} + \beta_i^{\text{div}} x_j^{\text{div}} + \sigma_i \varepsilon_{ijt} \quad (1.10)$$

$$\beta_i^A = \beta^A z_i$$

$$\beta_i^{\text{col}} = \beta^{\text{col}} z_i \quad (1.11)$$

$$\beta_i^{st} = \beta^{st} z_i st_i$$

$$\beta_i^{\text{div}} = \beta^{\text{div}} z_i \,\text{div}_i \quad (1.12)$$

$$\sigma_i = \beta^{\sigma} z_i \quad (1.13)$$

$$\varepsilon_{ijt} \sim \text{Type I Extreme Value.}$$

Worker i's marginal utility of the exogenous amenities β_i^A, endogenous amenities β_i^{col}, and birthplace amenities β_i^b, are each a function of his demographics z_i. st_i and div_i are a 50x1 and 9x1 binary vectors, respectively where each element is equal to 1 if the worker was born in the state or Census division.

The endogenous amenities are a function of the college employment ratio of the overall city, which workers take as fixed when evaluating the desirability of the amenities. Each worker is small and cannot individually impact amenity levels.

Each worker also has an individual, idiosyncratic taste for cities' amenities, which is measured by ε_{ijt}. ε_{ijt} is drawn from a Type I Extreme Value distribution.[24] The variance of workers' idiosyncratic tastes for each city differs across demographic groups, as shown in equation (1.12). Explicitly incorporating iid error terms in the utility function is important because households surely have idiosyncratic preferences beyond those based on their demographics, and iid errors allow a demographic group's aggregate demand elasticities for each city with respect to cities' characteristics to be smooth and finite.

To simplify future notation and discussion of estimation, I re-normalize the utility function by dividing each workers' utility by $\beta^{\sigma} z_i$. Using these units, the standard deviation of worker idiosyncratic preferences for cities is normalized to one. The magnitudes of the coefficient on wages, rents, and amenities now represent the elasticity of workers' demand for a small city with respect to its local wages, rents, or amenities, respectively.[25] With a slight abuse of notation, I redefine the parameters of the re-normalized utility function using the same notation of the utility function measured in wage units. The indirect utility for worker i of city j is now represented as:

$$V_{ijt} = \beta^{w} z_i w_{jt}^{edu} - \beta^{r} z_i r_{jt} + \beta^{A} z_i x_{jt}^{A} + \beta^{\text{col}} z_i \frac{H_{jt}}{L_{jt}} + \beta^{st} z_i st_i x_j^{st} + \beta^{\text{div}} z_i \,\text{div}_i\, x_j^{\text{div}} + \varepsilon_{ijt}.$$

To show how the utility function determines the population distribution of workers across cities, I introduce some additional notation. The preferences of different workers with identical demographics z for a given city differ only due to workers' birth states and divisions (st_i, div_i) and their idiosyncratic taste for the city, ε_{ijt}. I define δ_{jt}^{z} as utility value of the

[24] I also normalize ε_{ijt} such that is has a mean of zero. The standard Type I extreme value distribution has a mean of 0.57. Thus, I make ε_{ijt} distributed by a Type I value distribution, minus .57. This simply amounts to subtracting 0.57 from all workers' utility functions.

[25] Due to the functional form assumption for the distribution of workers' idiosyncratic tastes for cities, the elasticity of demand of workers with demographics z for a city j with respect to local rents, for example, is:$(1 - s_{jz})\, \beta^{r} z$. s_{jz} is the share of all workers of type z in the nation, living in city j. For a small city, where the share of all type z workers living in city j is close to zero, the demand elasticity for rent is simply $\beta^{r} z$.

components of city j which all workers' of type z value identically:

$$\delta_{jt}^{z} = \beta^{w} z w_{jt}^{edu} - \beta^{r} z r_{jt} + \beta^{A} z x_{jt}^{A} + \beta^{\text{col}} z \frac{H_{jt}}{L_{jt}}.$$

Economically, δ_{jt}^{z} measures workers of type z who are not born in city $j's$ state or census division's average utility for a city. I will refer to δ_{jt}^{z} as the mean utility of city j for workers of type z. Rewriting the utility function in terms of δ_{jt}^{z} gives:

$$V_{ijt} = \delta_{jt}^{z} + \beta^{st} z_i st_i x_j^{st} + \beta^{\text{div}} z_i \,\text{div}_i\, x_j^{\text{div}} + \varepsilon_{ijt}.$$

Since each worker elects to live in the city which gives him the highest utility level, different city choices made by workers with identical demographics and birth states must be due to differences in their idiosyncratic tastes, ε_{ijt}.

Aggregate population differences of workers of a given type z across cities represent differences in these workers' mean utility values for these cities. To illustrate this, I use a simple example of 2 cities and no demographic differences in the population. Workers' utilities for cities 1 and 2 are:

$$\begin{aligned} U_{i1} &= u_1 + \varepsilon_{i1} \\ U_{i2} &= u_2 + \varepsilon_{i2}. \end{aligned}$$

Each worker chooses which city he prefers to live in. Without loss of generality, assume city 2 is on average nicer that city 1, $u_2 > u_1$. The share of the population that live in city 1 is the percent of people whose idiosyncratic preference for city 1 relative to city 2 $(\varepsilon_{i1} - \varepsilon_{i2})$ is larger than the difference in average utility of city 2 versus city 1:

$$(\varepsilon_{i1} - \varepsilon_{i2}) > u_2 - u_1.$$

While this relationship holds for any distributional assumption on ε_{ij}, the expected population in each city is analytic when ε_{ij} is assumed to be drawn from a Type I Extreme Value distribution. Under this assumption, the expected population shares for each city are:

$$\begin{aligned}\frac{Pop_1}{Pop_1 + Pop_2} &= \frac{\exp(u_1)}{\exp(u_1) + \exp(u_2)} \\ \frac{Pop_2}{Pop_1 + Pop_2} &= \frac{\exp(u_2)}{\exp(u_1) + \exp(u_2)}.\end{aligned}$$

Pop_1 and Pop_2 are the population sizes in cities 1 and 2, respectively. Dividing these and taking logs gives:

$$\ln(Pop_1) - \ln(Pop_2) = u_1 - u_2.$$

The difference in the cities' mean utilities equals the difference in log population of the cities.

In the general case, there are many cities, and workers have heterogeneous preferences based on demographics. The probability worker i chooses to live in city j is:

$$\Pr(V_{ijt} > V_{i-jt}) = \frac{\exp(\delta_{jt}^{z_i} + \beta^{st} z_i st_i x_j^{st} + \beta^{\text{div}} z_i \, \text{div}_i \, x_j^{\text{div}})}{\sum_k^J \exp(\delta_{kt}^{z_i} + \beta^{st} z_i st_i x_k^{st} + \beta^{\text{div}} z_i \, \text{div}_i \, x_k^{\text{div}})}.$$

This setup is the conditional logit model, first formulated in this utility maximization context by McFadden (1973). The total expected population of city j is simply the probability each worker lives in the city, summed over all workers. Thus, the total high and low skill populations of city j are:

$$\begin{aligned}H_{jt} &= \sum_{i \in \mathcal{H}_t} \frac{\exp(\delta_{jt}^{z_i} + \beta^{st} z_i st_i x_j^{st} + \beta^{\text{div}} z_i \, \text{div}_i \, x_j^{\text{div}})}{\sum_k^J \exp(\delta_{kt}^{z_i} + \beta^{st} z_i st_i x_k^{st} + \beta^{\text{div}} z_i \, \text{div}_i \, x_k^{\text{div}})} \\ L_{jt} &= \sum_{i \in \mathcal{L}_t} \frac{\exp(\delta_{jt}^{z_i} + \beta^{st} z_i st_i x_j^{st} + \beta^{\text{div}} z_i \, \text{div}_i \, x_j^{\text{div}})}{\sum_k^J \exp(\delta_{kt}^{z_i} + \beta^{st} z_i st_i x_k^{st} + \beta^{\text{div}} z_i \, \text{div}_i \, x_k^{\text{div}})}.\end{aligned}$$

$\mathcal{H}_t$ and $\mathcal{L}_t$ are the set of high and low skill workers in the nation, respectively.

While population reflects a city's desirability, this relationship can be attenuated in the presence of moving costs, since households will be less willing to move to nicer cities and away from worse cities in the presence of moving costs. I have implicitly incorporated moving costs by allowing workers to prefer to live in or near their state of birth. This setup can be thought of as there being a childhood period of life before one's career. During childhood, workers are born into their birth locations, and as adults, they are allowed to move to a new city for their career. The utility value of living in or near one's birth state represents both the value of being near one's family and friends, as well as the psychic and financial costs of moving away. In a fully dynamic model, workers can elect to move every period, and they are no longer always moving away from their birth state. Panel data is needed to estimate a model of this nature, such as the NLSY used by Kennan and Walker (2011) and Bishop (2010). However, this dataset is significantly smaller and is not large enough to consistently estimate my model.[26] Further, recent work by Notowidigdo (2011) argues that mobility costs are reasonably modest (around 17% of annual income for the marginal migrant). He concludes that changes over 10 year horizons, as I analyze here, can be well approximated with a model of no mobility costs.

1.4.5 Equilibrium

Equilibrium in this model is defined by a menu of wages, rents and amenity levels, $\left(w_t^{L*}, w_t^{H*}, r_t^*, \frac{H_{jt}^*}{L_{jt}^*}\right)$ with populations $\left(H_{jt}^*, L_{jt}^*\right)$ such that:

- The high skill labor demand equals high skill labor supply:

$$\begin{aligned} H_{jt}^* &= \sum_{i \in \mathcal{H}_t} \frac{\exp(\delta_{jt}^{z_i} + \beta^{st} z_i st_i x_j^{st} + \beta^{\text{div}} z_i \, \text{div}_i \, x_j^{\text{div}})}{\sum_k^J \exp(\delta_{kt}^{z_i} + \beta^{st} z_i st_i x_k^{st} + \beta^{\text{div}} z_i \, \text{div}_i \, x_k^{\text{div}})} \qquad (1.14) \\ w_{jt}^{H*} &= c_t + (1-\rho) \ln N_{jt}^* + (\rho - 1) \ln H_{jt}^* + \gamma_H \ln \left(\frac{H_{jt}^*}{L_{jt}^*} \right) + \varepsilon_{jt}^H \end{aligned}$$

[26] I have considered more explicitly incorporating moving costs using the census' migration data in which MSA workers lived 5 years ago. Unfortunately, the confidentially screening makes this data offer little useful information. In all decades of the public use census data, significantly more MSAs (and a larger proportion of each MSA) are reported in the data covering MSA of current residence than the data reporting MSA residence 5 years earlier. As a result, implied 5 year population changes do not strongly track other official data on measured population changes across the 5 year periods. Further, only half of the 1980 dataset has migration data reported. Overall, this data is not helpful in estimating moving costs.

- The low skill labor demand equals low skill labor supply:

$$\begin{aligned}
L^*_{jt} &= \sum_{i\in\mathcal{L}_t} \frac{\exp(\delta^{z_i}_{jt} + \beta^{st} z_i st_i x^{st}_j + \beta^{\text{div}} z_i \,\text{div}_i\, x^{\text{div}}_j)}{\sum^J_k \exp(\delta^{z_i}_{kt} + \beta^{st} z_i st_i x^{st}_k + \beta^{\text{div}} z_i \,\text{div}_i\, x^{\text{div}}_k)} \qquad (1.15)\\
w^{L*}_{jt} &= c_t + (1-\rho)\ln N^*_{jt} + (\rho-1)\ln L^*_{jt} + \gamma_L \ln\left(\frac{H^*_{jt}}{L^*_{jt}}\right) + \varepsilon^L_{jt}
\end{aligned}$$

- Total labor supply equals housing supply:

$$r^*_{jt} = \ln(i_t) + \ln(CC_{jt}) + \left(\gamma + \gamma^{geo}\exp\left(x^{geo}_j\right) + \gamma^{reg}\exp\left(x^{reg}_j\right)\right)\ln\left(L^*_{jt} + H^*_{jt}\right).$$

- Endogenous amenities supplied equal college employment ratio:

$$\begin{aligned}
\delta^{z_{i\in H_t}}_{jt} &= \beta^w z_i w^{H*}_{jt} - \beta^r z_i r^*_{jt} + \beta^A z_i x^A_{jt} + \beta^{\text{col}} z_i \frac{H^*_{jt}}{L^*_{jt}}\\
\delta^{z_{i\in L_t}}_{jt} &= \beta^w z_i w^{L*}_{jt} - \beta^r z_i r^*_{jt} + \beta^A z_i x^A_{jt} + \beta^{\text{col}} z_i \frac{H^*_{jt}}{L^*_{jt}}.
\end{aligned}$$

The model does not allow me to solve for equilibrium wages and local prices analytically, but this setup is useful in estimation.

1.5 Estimation

1.5.1 Labor Demand

As discussed in the Section 1.4.1, a city's high and low skill labor demand curves determine the quantity of labor demanded by local firms as a function of local productivity and wages. In first differences over decades, these are:

$$\begin{aligned}
\Delta w^H_{jt} &= \Delta c_t + (1-\rho)\Delta\ln N_{jt} + (\rho-1)\Delta\ln H_{jt} + \gamma_H \Delta\ln\left(\frac{H_{jt}}{L_{jt}}\right) + \Delta\varepsilon^H_{jt} \qquad (1.16)\\
\Delta w^L_{jt} &= \Delta c_t + (1-\rho)\Delta\ln N_{jt} + (\rho-1)\Delta\ln L_{jt} + \gamma_L \Delta\ln\left(\frac{H_{jt}}{L_{jt}}\right) + \Delta\varepsilon^L_{jt} \qquad (1.17)
\end{aligned}$$

Changes over a decade in cities' high and low skill exogenous productivity levels, $\Delta\varepsilon_{jt}^{L}$ and $\Delta\varepsilon_{jt}^{H}$, shift the local labor demand curves, directly impacting wages. Changes in the productivity levels of the industries located within each city contribute to the city's productivity change. Variation in productivity changes across industries will differentially impact cities' local high and low skill productivity levels based on the industrial composition of the city's workforce (Bartik (1991)). I measure exogenous local productivity changes by interacting cross-sectional differences in industrial employment composition with national changes in industry wage levels, separately for high and low skill workers.[27] I refer to these as Bartik shocks. Formally, I define the Bartik shock for high and low skill workers, as:

$$\Delta B_{jt}^{H} = \sum_{ind} \left(w_{ind,-j,t}^{H} - w_{ind,-j,t-10}^{H} \right) \frac{H_{ind,jt-10}}{H_{jt-10}} \tag{1.18}$$

$$\Delta B_{jt}^{L} = \sum_{ind} \left(w_{ind,-j,t}^{L} - w_{ind,-j,t-10}^{L} \right) \frac{L_{ind,jt-10}}{L_{jt-10}},$$

where $w_{ind,-j,t}^{H}$ and $w_{ind,-j,t}^{L}$ represent the average log wage of high and low skill workers, respectively, in industry ind in year t, excluding workers in city j. $H_{ind,jt-10}$ and $L_{ind,jt-10}$ measure the number of high and low skill workers, respectively, employed in industry ind in city j, in year t-10.

Since I do not observe all contributors to local productivity changes, I rewrite the labor demand equations (1.16) and (1.17) by explicitly defining $\Delta\tilde{\varepsilon}_{jt}^{H}$ and $\Delta\tilde{\varepsilon}_{jt}^{L}$, as all high and low

[27] Other work has measured industry productivity changes by using national changes in employment shares of workers across industries, instead of changes in industry wages. (See Notowidigdo (2011), and Blanchard and Katz (1992).) They use the productivity shocks as an instrument for worker migration to cities. Thus, it makes sense to measure the shock in units of workers, instead of wages units. I focus on how these industry productivity shocks impact wages, which is why I measure the shock in wages units. Guerrieri, Hartley, and Hurst (2011) also constructs the instrument using industry wage changes.

skill local productivity changes uncorrelated with the Bartik shocks $\Delta B^H_{jt}, \Delta B^L_{jt}$:[28]

$$\begin{aligned}
\Delta w^H_{jt} &= \Delta c_t + (1-\rho)\,\Delta\ln N_{jt} + (\rho-1)\,\Delta\ln H_{jt} \qquad (1.19)\\
&\quad +\gamma_H \Delta\ln\left(\frac{H_{jt}}{L_{jt}}\right) + \beta^H_B \Delta B^H_{jt} + \Delta\tilde{\varepsilon}^H_{jt}\\
\Delta w^L_{jt} &= \Delta c_t + (1-\rho)\,\Delta\ln N_{jt} + (\rho-1)\,\Delta\ln L_{jt} \qquad (1.20)\\
&\quad +\gamma_L \Delta\ln\left(\frac{H_{jt}}{L_{jt}}\right) + \beta^L_B \Delta B^L_{jt} + \Delta\tilde{\varepsilon}^L_{jt}.
\end{aligned}$$

The direct effect of the Bartik shocks shift the local labor demand curves, directly influencing local wages. To identify the labor demand curve parameters, one needs variation in the quantities of high and low skill labor supplied to each city which is independent of unobserved changes in local productivity levels.

The interaction of the Bartik productivity shocks with cities' housing supply elasticities leads to variation in the migration response to the Bartik shocks, which can identify the labor demand parameters. To see this, consider an example of 2 cities which receive the same positive local productivity shocks. One city has a very elastic housing supply, while the housing supply of the other is very inelastic. As workers migrate into these cities to take advantage of the increased wages, they drive up the housing prices by increasing the local demand for housing. The housing inelastic city exhibits much larger rent increases in response to a given amount of migration than the elastic city. These rent increases lead to relatively less in-migration to the housing inelastic city because the sharp rent increase driven by a relatively small amount of in-migration offsets the desirability of high local wages.[29] While the direct effect of the Bartik shocks shifts local labor demand curves, the *interaction* of the labor demand shocks with city's housing supply elasticities causes variation in labor

[28]The parameters $\left(\beta^H_B, \beta^L_B\right)$ on the Bartik shocks in the high and low skill labor demand equations represent the projection of exogenous changes in local productivity on the Bartik shocks: $\Delta\varepsilon^H_{jt} = \beta^H_B \Delta B^H_{jt} + \Delta\tilde{\varepsilon}^H_{jt}$. Thus, β^H_B and β^L_B are not structural parameters. Since this is a projection, $\Delta\tilde{\varepsilon}^H_{jt}$ is mechanically uncorrelated with the Bartik shock ΔB^H_{jt}.

[29]Saks (2008) has also analyzed how labor demand shocks interact is local housing supply elasticities to influence equilbrium local wages, rents, and populations.

supply, which can be used as instruments for high and low skill population changes to identify the labor demand curve parameters.

As discussed in Section 1.4.2, land unavailable for housing development due to geographic features x_j^{geo} and land-use regulation x_j^{reg} impact local housing supply elasticity. The interaction of these housing supply elasticity measures with local Bartik shocks are used as instruments for quantities of labor within the city. The exclusion restriction assumes that land-unavailability and land-use regulation do not directly impact unobserved changes in local productivity.

An additional measure of how Bartik shocks influence housing price changes is the size of the aggregate Bartik shock. I measure the aggregate Bartik shock by:

$$\begin{aligned}\Delta B_{jt}^{all} &= \sum_{ind} \left(\begin{array}{c} \left(w_{ind,-j,t}^{H} - w_{ind,-j,t-10}^{H}\right) \frac{H_{ind,jt-10}}{L_{jt-10}+H_{jt-10}} \\ + \left(w_{ind,-j,t}^{L} - w_{ind,-j,t-10}^{L}\right) \frac{L_{ind,jt-10}}{L_{jt-10}+H_{jt-10}} \end{array} \right) \\ &= \Delta B_{jt}^{H} \frac{H_{jt-10}}{L_{jt-10}+H_{jt-10}} + \Delta B_{jt}^{L} \frac{L_{jt-10}}{L_{jt-10}+H_{jt-10}}. \end{aligned} \tag{1.21}$$

The aggregate Bartik shock adds additional information by weighting the high and low skill Bartik shocks by the share of each cities total population made up of high and low skill workers. Cities which have a disproportionate amount of high skill workers will have local house prices respond relatively more to high skill Bartik shocks than cities with smaller college population shares.

The shape of cities' labor demand curves are functions of firms' labor demand curves, which are governed by ρ, and the spillovers of the city's college employment ratio on the productivity of high and low skill workers, governed by γ_H and γ_L. Since I only observe changes in city-level aggregate amounts of labor, I cannot disentangle the shape of firms' labor demand curves from changes in city productivity spillovers without parametric assumptions. However, the combined effect of productivity spillovers and firms' labor demands on local wages can be identified. Exogenous changes in the supply of labor to a city can be used to measure the impact of labor supply to a city on local wages, but not how much of this effect is driven by productivity spillovers versus movement along firms' labor demand curves.

Under the assumption that firms' labor demand curves have a CES functional form, one can disentangle the productivity spillovers from shifts along firm's labor demand curves by the asymmetric response of high and low skill labor changes on high and low skill wages. Equations (1.16) and (1.17) show that if there were no productivity spillovers, the inverse elasticity of high skill labor demand with respect to high skill wages should be the same as the inverse elasticity of low skill labor demand with respect to low skill wages. Further, the cross price inverse elasticities should also be the same for high and low skill labor. By adding productivity spillovers, I allow the labor demand elasticities to be more flexible than the structure imposed by the CES functional form. However, separately identifying ρ from γ_H and γ_L is due to parametric assumptions.

A technical issue with estimating the labor demand equations (1.19) and (1.20) is that $\ln N_{jt}$ is a non-linear, non-seperable function of local productivity ε_{jt}^{L} and ε_{jt}^{H}, which is not directly observed by the econometrician, as seen in equation (1.5) . Due to the CES functional form, I show in Appendix A.2 that N_{jt} can be rewritten as non-linear function of data $(w_{jt}^{H}, w_{jt}^{L}, L_{jt}, H_{jt})$ and parameters $(\rho, \gamma_L, \gamma_H)$:

$$\hat{N}_{jt}(\rho, \gamma_L, \gamma_H) = \left(\frac{w_{jt}^{L} H_{jt}^{\rho-1+\gamma_H} L_{jt}^{\rho-\gamma_L} + w_{jt}^{H} L_{jt}^{\rho-1-\gamma_L} H_{jt}^{\rho+\gamma_H}}{w_{jt}^{H} L_{jt}^{\rho-1+\gamma_H-\gamma_L} + w_{jt}^{L} H_{jt}^{\rho-1+\gamma_H-\gamma_L}} \right)^{1/\rho}. \tag{1.22}$$

Equation (1.22) can be substituted in for N_{jt} in the labor demand equations:

$$\begin{aligned}
\Delta w_{jt}^{H} &= \Delta c_t + (1-\rho)\Delta \ln \hat{N}_{jt}(\rho, \gamma_L, \gamma_H) + (\rho - 1)\Delta \ln H_{jt} \\
&\quad + \gamma_H \Delta \ln \left(\frac{H_{jt}}{L_{jt}} \right) + \beta_B^{H} \Delta B_{jt}^{H} + \Delta \tilde{\varepsilon}_{jt}^{H} \\
\Delta w_{jt}^{L} &= \Delta c_t + (1-\rho)\Delta \ln \hat{N}_{jt}(\rho, \gamma_L, \gamma_H) + (\rho - 1)\Delta \ln L_{jt} \\
&\quad + \gamma_L \Delta \ln \left(\frac{H_{jt}}{L_{jt}} \right) + \beta_B^{L} \Delta B_{jt}^{L} + \Delta \tilde{\varepsilon}_{jt}^{L}.
\end{aligned}$$

Thus, for a given guess of the parameter estimates $(\rho, \gamma_L, \gamma_H)$, N_{jt} is known and does not depend on $\left(\varepsilon^L_{jt}, \varepsilon^H_{jt}\right)$. The local productivity terms $\left(\varepsilon^L_{jt}, \varepsilon^H_{jt}\right)$ are additively separable in the labor demand equation, given N_{jt}. See Appendix A.2 for further details on constructing the moment functions used in estimation.

These moment restrictions will be combined with the moments identifying cities' housing supply curves and workers' labor supply to cities. All parts of the model will be estimated jointly using two-step GMM estimation.

1.5.2 Housing Supply

I rewrite the housing supply curve in first differences over decades:

$$\Delta r_{jt} = \Delta \ln (i_t) + \left(\gamma + \gamma^{geo} \exp\left(x^{geo}_j\right) + \gamma^{reg} \exp\left(x^{reg}_j\right)\right) \Delta \ln (H_{jt} + L_{jt}) + \Delta\varepsilon^{CC}_{jt}.$$

$\Delta\varepsilon^{CC}_{jt}$ measures local changes in construction costs and other factors impacting housing prices not driven by population change. To identify the elasticity of housing supply, one needs variation in a city's total population which is unrelated to changes in unobserved factors driving housing prices. I use the Bartik shocks discussed above, which shift local wages leading to a migration response of workers, as instruments for housing demand. I combine these moments with the labor demand and labor supply moments to jointly estimate all the parameters of the model.

1.5.3 Labor Supply

Recall that the indirect utility of city j for worker i with demographics z_i is:

$$\begin{aligned} V_{ijt} &= \delta^z_{jt} + \beta^{st} z_i st_i x^{st}_j + \beta^{\text{div}} z_i \,\text{div}_i \, x^{\text{div}}_j + \varepsilon_{ijt} \\ \delta^{z_i}_{jt} &= \beta^w z_i w^{edu}_{jt} - \beta^\zeta z_i r_{jt} + \beta^A z_i x^A_{jt} + \beta^{\text{col}} z_i \frac{H_{jt}}{L_{jt}}. \end{aligned}$$

To estimate workers' preferences for cities, I use a two-step estimator similar to Berry, Levinsohn, and Pakes (2004).

In the first step, I use a maximum likelihood estimator, in which I treat the mean utility value of each city for each demographic group in each decade δ^z_{jt} as a parameter to be estimated. Recall the discussion from Section (1.4.4) that shows how differences in the mean utility value of cities leads to population differences across cities for a given type of worker. Observed population differences in the data for a given type of worker are what identify the mean utility estimates for each city. In the simple case where workers do not gain utility from living close to their birth state, the estimated mean utility levels for each city would exactly equal the log population of each demographic group observed living in that city. When workers' care about living close to their state of birth, the mean utility estimates represent what the population distribution would be across cities in the absence of birth state preferences. The maximum likelihood estimation measures the mean utility level for each city, for each demographic group, for each decade of data.[30]

The second step of estimation decomposes the mean utility estimates into how workers value wages, rents, and amenities. First differencing cities' mean utility estimates for workers with demographics z across decades:

$$\Delta\delta^z_{jt} = \beta^w z\Delta w^{edu}_{jt} - \beta^r z\Delta r_{jt} + \beta^A z\Delta x^A_{jt} + \beta^{\text{col}} z\Delta\left(\frac{H_{jt}}{L_{jt}}\right). \tag{1.23}$$

I observe changes in cities' wages, rents, and college employment ratios in the data. However, I do not observe the other amenity changes. Define $\Delta\xi^z_{jt}$ as the change in utility value of city $j's$ amenities unobserved to the econometrician across decades for workers with demographics z:

$$\Delta\xi^z_{jt} = \beta^A z\Delta x^A_{jt}.$$

[30] While one often worries about bias in estimating fixed effects using a non-linear objective function, such as maximum likelihood, I have a very large sample of individual level data (over 2 million observations for each decade), relative to about 1000 estimated fixed effects per decade. The asymptotics assume the number of individuals goes to infinity at a faster rate than the number of cities goes to infinity.

Plugging this into equation (1.23) gives:

$$\Delta\delta_{jt}^{z} = \beta^{w} z \Delta w_{jt}^{edu} - \beta^{\zeta} z \Delta r_{jt} + \beta^{\text{col}} z \Delta \left(\frac{H_{jt}}{L_{jt}}\right) + \Delta\xi_{jt}^{z}. \tag{1.24}$$

To identify workers' preferences for cities' wages, rents, and endogenous amenities, I instrument for these outcomes using the Bartik productivity shocks and their interaction with housing supply elasticity characteristics (land-use regulation and land availability). The Bartik shocks measure nationwide changes in industry wages and provide variation in local labor demand unrelated to unobserved changes in local amenities. Since workers will migrate to take advantage of desirable wages driven by the labor demand shocks, they will bid up rents in the housing market. Heterogeneity in cities' housing supply elasticities provides variation in the rental rate response to the induced migration. Thus, the interactions of housing supply elasticity characteristics with the Bartik shocks impact changes in rents (and wages) unrelated to unobserved changes in local amenities. To identify the migration elasticity of workers within a given skill group with respect to endogenous amenities (as measured by the college employment ratio), the Bartik shock to the *other* skill group is useful. For example, the low skill Bartik shock impacts the quantity of low skill workers living in a city, which leads to endogenous amenity changes by shifting the local college employment ratio. This shift in endogenous amenities will impact high skill workers' migration, identifying high skill workers' preference for endogenous amenities. While the low skill Bartik shocks also influence local prices and high skill workers' wages, jointly instrumenting for all three endogenous parameters simultaneously (wages, local prices, college employment ratio) allows all instruments to impact all endogenous outcomes and simultaneously identifies all three parameters. The exclusion restrictions assume that these instruments are uncorrelated with unobserved exogenous changes in the city's local amenities.

1.5.4 Summary of Estimating Equations

The key structural parameters to be estimated are workers' migration elasticities with respect to wages, local prices, and the endogenous amenity index (college employment ratio); city-wide labor demand elasticities, including firm's labor demand elasticities and endogenous productivity spillovers; and housing supply elasticities. All moment restrictions from the models of labor demand, housing supply, and labor supply to cities are stacked and all parameters are jointly estimated using a 2-step simultaneous equations GMM estimator. I summarize the estimating equations, moment restrictions, and instruments below.

Labor Demand

- Estimating Equations:

$$\begin{aligned}
\Delta w_{jt}^{H} &= \Delta c_t + (1-\rho)\,\Delta \ln \hat{N}_{jt}\,(\rho, \gamma_L, \gamma_H) + (\rho - 1)\,\Delta \ln H_{jt} \\
&\quad + \gamma_H \Delta \ln \left(\frac{H_{jt}}{L_{jt}}\right) + \beta_B^H \Delta B_{jt}^H + \Delta \tilde{\varepsilon}_{jt}^H \\
\Delta w_{jt}^{L} &= \Delta c_t + (1-\rho)\,\Delta \ln \hat{N}_{jt}\,(\rho, \gamma_L, \gamma_H) + (\rho - 1)\,\Delta \ln L_{jt} \\
&\quad + \gamma_L \Delta \ln \left(\frac{H_{jt}}{L_{jt}}\right) + \beta_B^L \Delta B_{jt}^L + \Delta \tilde{\varepsilon}_{jt}^L.
\end{aligned}$$

- Moment Restrictions:

$$E\left(\Delta \tilde{\varepsilon}_{jt}^H \Delta Z_{jt}\right) = 0$$

$$E\left(\Delta \tilde{\varepsilon}_{jt}^L \Delta Z_{jt}\right) = 0$$

$$\text{Instruments:} \Delta Z_{jt} \in \left\{ \begin{array}{c} \Delta B_{jt}^{all}, \Delta B_{jt}^{H}, \Delta B_{jt}^{L} \\ \Delta B_{jt}^{all} x_j^{reg}, \Delta B_{jt}^{H} x_j^{reg}, \Delta B_{jt}^{L} x_j^{reg} \\ \Delta B_{jt}^{all} x_j^{geo}, \Delta B_{jt}^{H} x_j^{geo}, \Delta B_{jt}^{L} x_j^{geo} \end{array} \right\}$$

Firm's labor demand elasticities and productivity spillovers due to a city's college employment ratio are identified by using the interaction of the Bartik labor demand shocks with local housing supply elasticity characteristics as instruments for changes in high and low skill worker populations. The key exclusion restriction is that the housing supply elasticity characteristics do not directly impact unobserved changes in local productivity.

Housing Supply

- Estimating Equation:

$$\Delta r_{jt} = \Delta \ln (i_t) + \left(\gamma + \gamma^{geo} \exp\left(x_j^{geo}\right) + \gamma^{reg} \exp\left(x_j^{reg}\right)\right) \Delta \ln (H_{jt} + L_{jt}) + \Delta\varepsilon_{jt}^{CC}.$$

- Moment Restrictions:

$$E\left(\Delta\varepsilon_{jt}^{CC} \Delta Z_{jt}\right) = 0$$

$$\text{Instruments:} \Delta Z_{jt} \in \left\{ \begin{array}{c} \Delta B_{jt}^{all}, \Delta B_{jt}^{H}, \Delta B_{jt}^{L} \\ \Delta B_{jt}^{all} x_j^{reg}, \Delta B_{jt}^{H} x_j^{reg}, \Delta B_{jt}^{L} x_j^{reg} \\ \Delta B_{jt}^{all} x_j^{geo}, \Delta B_{jt}^{H} x_j^{geo}, \Delta B_{jt}^{L} x_j^{geo} \end{array} \right\}$$

Housing supply elasticities are identified by using the Bartik shocks as instruments for population changes and observing the differential rental rate response across cities based on the measures of land-use regulation and land availability. The key exclusion restriction in that Bartik shocks only impact changes in local rents through changes in cities' populations.

Labor Supply

- Estimating Equation:

$$\Delta\delta_{jt}^{z} = \beta^{w} z \Delta w_{jt}^{edu} - \beta^{\zeta} z \Delta r_{jt} + \beta^{\text{col}} z \Delta \left(\frac{H_{jt}}{L_{jt}}\right) + \Delta\xi_{jt}^{z}.$$

- Moment Restrictions:

$$E\left(\Delta\xi^{z}_{jt}\Delta Z_{jt}\right) = 0$$

$$\text{Instruments:}\Delta Z_{jt} \in \left\{ \begin{array}{c} \Delta B^{all}_{jt}, \Delta B^{H}_{jt}, \Delta B^{L}_{jt} \\ \Delta B^{all}_{jt} x^{reg}_{j}, \Delta B^{H}_{jt} x^{reg}_{j}, \Delta B^{L}_{jt} x^{reg}_{j} \\ \Delta B^{all}_{jt} x^{geo}_{j}, \Delta B^{H}_{jt} x^{geo}_{j}, \Delta B^{L}_{jt} x^{geo}_{j} \end{array} \right\}$$

Changes in local wages, local prices, and college employment ratios are jointly instrumented for using Bartik shocks to local labor demand and their interactions with local housing supply elasticity characteristics. The key exclusion restriction assumes the Bartik shocks and housing supply elasticity characteristics do not directly impact changes in unobserved amenities.

See Appendix A.2 for further details on all steps of the estimation procedures.

1.6 Parameter Estimates

1.6.1 Worker Labor Supply

I estimate two specifications of the model. The main specification, which I refer to as the "full model," is the model exactly described in the previous sections. As an alternative specification, I estimate the model without endogenous amenities or endogenous productivity changes. I refer to this specification as the "no spillover" model. Parameter estimates for both models are reported in Table 1.3.

Panel A of Table 1.3 reports the estimates of workers' demand elasticities for cities with respect to wages, rents,and endogenous amenities. Focusing on the estimates of the full model, I find that both college and non-college workers prefer cities with higher wages, lower rent, and a higher college employment ratio. While both college and non-college workers find higher wages net of local prices and amenity levels desirable, the elasticity of college workers' demand for a city with respect to amenities is much higher than non-college workers' amenity demand elasticity. The improvement in amenities driven by a 1% increase in a cities' college employment ratio raises the college population within the city by 3.3%, while it only increases the non-college population by 1.3%.

Similarly, non-college workers' demand is much more elastic with respect to wages and rent than college workers. A 1% increases in local college wages increases the local college population by 1.4%, while a similar increase in low-skill worker wages leads to a 4.2% increase in the low-skill worker population. Similarly for rents, a 1% increase in local rent, leads to a 2.8% decrease in low skill worker population, while it only leads to a .95% decrease in the high skill population.

Table 1.3: GMM Estimates of Full Model & Model with No Spillovers

A. Worker Preferences for Cities					
College Workers:			Non-College Workers:		
	Full Model	No Spill-Overs		Full Model	No Spill-Overs
Wage	1.438***	2.326***	Wage	4.181***	4.673***
	[0.266]	[0.330]		[0.577]	[0.519]
Rent	-0.953***	-0.156	Rent	-2.832***	-2.464***
	[0.252]	[0.248]		[0.333]	[0.306]
Implied Local Expenditure			Implied Local Expenditure		
Share	0.663***	0.067	Share	0.677***	0.527
	[0.180]	[0.010]		[0.071]	[0.045]
College Emp Ratio Amenity	3.268***		College Emp Ratio Amenity	1.301***	-
	[0.531]	-		[0.482]	-
Differential Effects: Blacks			Differential Effects: Blacks		
Wage	3.182***	4.597***	Wage	3.351***	3.539***
	[1.202]	[0.870]		[0.761]	[0.687]
Rent	-1.956***	-1.152***	Rent	-0.994***	-0.887***
	[0.542]	[0.446]		[0.398]	[0.357]
College Emp Ratio Amenity	3.505***		College Emp Ratio Amenity	0.201	-
	[0.786]	-		[0.414]	-
Differential Effects: Immigrants			Differential Effects: Immigrants		
Wage	3.709***	3.424***	Wage	2.949***	2.872***
	[1.022]	[0.839]		[0.94]	[0.908]
Rent	0.866	0.080	Rent	-0.069	0.432
	[0.523]	[0.469]		[0.488]	[0.477]
College Ratio Amenity	-2.16***		College Ratio Amenity	0.802	-
	[0.499]	-		[0.515]	-

Table 1.3 (Continued)

B. Housing Supply					
	Full Model	No Spill-Overs	Predicted Inverse Housing Supply Elasticities (Full Model):		
Exp(Land Use Regulation)	0.062***	0.063***	Mean:	0.17	
	[0.01]	[0.009]			
Exp(Land Unavailability)	0.024***	0.019***	Standard Deviation:	0.16	
	[0.008]	[0.007]			
Base House Supply Elasticity	0.021	0.003	Minimum:	0.04	
	[0.082]	[0.074]	Maximum:	1.3	
C. Labor Demand					
	Full Model	No Spill-Overs		Full Model	No Spill-Overs
Rho	0.558***	0.499***			
	[0.043]	[0.048]			
Productivity Spillover on College Workers	0.553***	-	Productivity Spillover on Non-College Workers	0.182***	-
	[0.045]	-		[0.059]	-
Implied Average Wage Elasticity with respect to Labor, Full Model:					
College Wage wrt College Labor	0.517		Non-College Wages wrt College Labor	0.5895	
College Wage wrt Non-College Labor	-0.517		Non-College Wage wrt Non-College Labor	-0.5895	

Notes: Standard errors in brackets. Changes measured between 1980-1990 and 1990-2000. Black and immigrant estimates the differential preferences of these groups for each city characteristic, relative to base estimates for college and non-college workers. Magnitude of parameter estimates represent worker's demand elasticity with respect to the given city characteristic, in a small city. Sample is all heads of household with positive labor income. See text for model details. Standard errors clustered by MSA. *** p<0.01, ** p<0.05, * p<0.1

These results are consistent with a large body of work in empirical industrial organization which finds large amounts of heterogeneity in how sensitive consumers are to products' prices.[31]

The ratio of workers' demand elasticities for rents to wages measures their expenditure share on local goods. As derived in Section 1.4, since workers' preferences are Cobb-Douglas in the national and local good, the indirect utility value of rent measured in wage units represents the share of expenditure spent on locally priced goods. For both college and non-college workers, I estimate this to be 66%-68% (0.95/1.4=0.66 for college, 2.8/4.2=0.68 for non-college). Note that the estimation methods in no way mechanically restricted this to be less than one or similar across workers' skill groups. The expenditure share is inferred by workers' revealed preference on the trade off between wages and rents across cities. Albouy (2008) measures the relative utility impact of housing rents versus local wages to equal 65%.[32] While Albouy (2008) takes a very different estimation approach, it is reassuring we find very similar estimates.

The estimates of the "no spillover" model strongly differ from the full model. The no spillover model assumes that Bartik shocks and housing supply elasticity can only impact worker migration by influencing local wages and rent. Under this assumption, I find that college workers would have to spend only 7% of their expenditure on local goods, while non-college workers would spend 53%. To be able to rationalize the migration responses to the Bartik shocks through only wage and rent changes, college workers would have to be almost indifferent to local price levels in order to be willing to live in cities with such high housing prices. Data from the Consumer Expenditure Survey shows that housing expenditure makes

[31] For an example, see Berry, Levinsohn, and Pakes (1995) on consumer demand for cars. See Nevo (2010) for a review of this literature.

[32] He takes a different approach to estimating the relative utility value of wage earnings versus housing rent by explicitly accounting for a variety of mechanisms which influence how local wage earnings and housing prices impact workers' utility, such as non-labor income and tax rates. Since my methods use workers' revealed preferences on the relative utility value of local wages versus rents based on their city choices, I do not need to explicitly model all the possible mechanisms which wages and rent influence workers' utility levels. Workers will take into account all the ways which wages and rent impact their utility when selecting which city to live in. Moretti (2011b) uses a similar approach to Albouy, but does not explicitly consider taxes and non-labor income. He estimates local good expenditure to be 58%.

up 33% of the average college household's consumption and 32% of the average non-college household's consumption. These data make it clear that college workers' expenditure shares on local goods cannot be close to 7% and do not appear to be strongly non-homothetic.[33] This result is consistent with the reduced form evidence presented earlier in Section 1.3, which shows that in addition to local wage and rent changes, local amenity changes were related to changes in cities' college employment ratios.

Panel A of Table 1.3 shows that black college and non-college workers are more sensitive to wages, rents, and amenities. As explicitly derived in the model section, higher aggregate demand elasticities with respect to all city characteristics signifies a relatively low variance of idiosyncratic tastes across workers for each city. Since the black population is likely a more homogenous group than the combination of all other non-immigrant races and ethnicities lumped together in the non-black group, the variance of black worker's idiosyncratic tastes for each city is likely to be lower. Intuitively, if most workers agree on the utility value of each city, changes in local wages, rents, or amenities that make a given worker migrate to that city will also make many other similar workers migrate, since they all agree it offers the higher utility value.

Immigrant preferences differ strongly from the US non-immigrant population. I find that immigrants are much more sensitive to city wages. A 1% increase in local wages increases the low skill immigrant population by 7.1% and the high skill immigrant population by 5.2%. However, immigrants are much less sensitive to local good prices. A 1% increase in local rents lead to out migration of 0.09% of the high skill immigrant population, and 2.9% decrease in the low skill immigrant population. This relative wage-rent trade off implies that

[33] When looking at housing expenditure shares across households of different *incomes* in the CEX, one sees that households with higher incomes have a lower expenditure share on housing. However, Chetty and Szeidl (2007) show that housing consumption expenditure is relatively hard for a household to adjust year to year. Household income, on the other hand, has much more year to year fluctuation. High variance in income coupled with a low variance in housing consumption over time, within a household, will lead to the spurious cross-sectional correlation that the households with higher annual income spend a small fraction on housing. If this were evidence of non-homotheticity in housing consumption, then we should see a similar gradient across proxies for households' permanent income, such as education. Given that the consumption expenditure on housing is strikingly flat across education levels, one can conclude that housing consumption is close to homothetic.

immigrants must spend a smaller share of their expenditure on local goods. Additionally, high skill immigrants are less elastic to the college employment ratio driven endogenous amenities than non-immigrant college workers. Overall, these results show that immigrants elect their city to live in based almost exclusively on local wages. They spend much less on housing and local goods, which makes them able to afford to live in cities with high housing prices.

Table 1.4 reports estimates for workers' preferences to live in their own state of birth or Census Division of birth.[34] Non-college workers are 4.4 times more likely to live in a given MSA if it is located is his state of birth than if it is not, while college workers are only 3.5 times more likely. Both college and non-college workers are 2.2 times more likely to live in an MSA located in his Census Division of birth than an MSA farther away. These estimates are similar for blacks. Unlike the amenities influenced by a city's college employment ratio, the amenity of living near one's place of birth influences the city choices of low skill workers more than high skill. This is consistent with the migration literature that finds high skilled workers are more likely to move away from their place of birth.[35]

[34] I estimate decade-specific parameters for workers' preferences to live close to their state of birth. This is purely for computational convenience. Since these parameters are jointly estimated along with the mean utility levels for each city for each demographic group for each decade, estimating each decade's parameters in a separate optimization allowed for a signficant decrease in the computational memory requirements needed for estimation.

[35] See Greenwood (1997) for a review of this literature.

Table 1.4: Value of Living in Own Birth State & Division

A. Birth State

	1980		1990		2000	
	Base	Black	Base	Black	Base	Black
Non-College	3.413	-0.119	3.412	0.054	3.423	0.157
	[0.0022]	[0.0069]	[0.004]	[0.018]	[0.005]	[0.018]
College	2.537	0.225	2.525	0.252	2.629	0.206
	[0.0047]	[0.018]	[0.003]	[0.0105]	[0.0032]	[0.0105]

B. Birth Division

	1980		1990		2000	
	Base	Black	Base	Black	Base	Black
Non-College	1.286	-0.325	1.268	-0.529	1.216	-0.540
	[0.0034]	[0.0059]	[0.006]	[0.028]	[0.005]	[0.014]
College	1.199	-0.484	1.192	-0.510	1.140	-0.382
	[0.0047]	[0.018]	[0.002]	[0.0101]	[0.0042]	[0.0105]

Notes: Standard erorrs in brackets. Estimates from maximum likelihood of conditional logit model of city choice. Magnitudes represent the semi-elasticity of demand for an average sized city with respect to whether the city is located within one's birth state or division. Black estimates are relative to base estimates. Sample is all heads of household with positive labor income.

1.6.2 Housing Supply

Panel B of Table 1.3 presents the inverse housing supply elasticity estimates. Consistent with the work of Saiz (2010) and Saks (2008), I find housing supply is less elastic in areas with higher levels of land-use regulation and less land near a city's center available for real estate development. The inverse housing supply elasticity estimates do not differ much between the full model and the no spillover model. This is expected, since these two models have identical housing supply models. I use the parameter estimates to predict the inverse elasticity of housing supply in each city. Panel B of Table 1.3 shows that the average inverse housing supply elasticity is 0.17, with a standard deviation of 0.16. A regression of my inverse housing supply elasticity estimates on Saiz (2010)'s estimates yields a coefficient of 1.04 (0.12), suggesting we find similar amounts of variation in housing supply elasticities across cities. However, Saiz (2010)'s inverse housing supply estimates are higher than mine by 0.39, on average. The overall level of my estimates is governed by the "base" inverse housing supply term, γ. This parameter is the least precisely estimated of the housing supply elasticity parameters, with a point estimate of 0.021 (0.082), which could explain why I find lower inverse housing supply estimates overall. Further, Saiz's estimates are identified using a single, long run change in housing prices from 1970-2000, while I am looking at decadal changes from 1980 to 1990 and 1990 to 2000. Differences in time frame could impact these parameter estimates as well.

1.6.3 Labor Demand

Panel C of Table 1.3 presents parameter estimates for the local labor demand curves. In the no spillovers model, I estimate ρ to be 0.499, which implies a elasticity of labor substitution of 1.99. This estimate is very close to others in the literature, which tend to be between 1 and 3.[36] Work by Card (2009) estimates the elasticity of labor substitution at the MSA level and finds an elasticity of 2.5, which is not statistically distinguishable from my results.

[36] See Katz and Autor (1999) for a literature review of this work.

In the full model specification, I allow for a more flexible labor demand curve. The parametric assumptions allow me to disentangle the slopes of firms' labor demand curves from productivity spillovers due to changes in cities' college employment ratios. With this more general model, I estimate ρ to be 0.558, which implies an elasticity of labor substitution within firms of 2.26, slightly higher than the more restrictive model. Additionally, a 1% increase in the a city's college employment ratio increases the productivity of college workers by 0.55% and non-college workers by 0.18%.

These estimates suggest that the social returns to education are high. Combining the estimates of the elasticity of labor substitution with the productivity spillovers, I find that a 1% increase in a city's college worker population *raises* the local college wage by 0.52% and the local non-college wage by 0.59%. Similarly, a 1% increase in a city' non-college worker population decreases college wages by 0.52% and non-college wages by 0.59%. The wage increases due to local productivity spillovers, such as endogenous technology adoption or knowledge and ideas spillovers, overwhelm the wage decreases driven by movement along firms' labor demand curves from the increase supply of college graduate labor.

Moretti (2004a) also analyzes the impact of high and low skill worker labor supply on workers' wages within a city. He estimates a 1% increase in a city's college employment ratio leads to a .16% increase in the wages of high school graduates and a .10% increase in the wages of college graduates, both of which are smaller than my findings.[37] His estimates are identified off of cross-sectional variation in city's college shares, driven by the presence of a land grant college, while my estimates are estimated off of changes in skill-mix driven by housing supply elasticity heterogeneity. Additionally, my estimates explicitly combine the impact of

[37] Moretti (2004a)'s setup looks at the impact of a city's share of college graduates $\left(\frac{H_{jt}}{H_{jt}+L_{jt}}\right)$ on workers' wages by education level, while my setup measures the local education mix using the log ratio of college to non-college workers $\left(\ln\frac{H_{jt}}{L_{jt}}\right)$. To transform Moretti's estimates into the same units of my own, note that $\frac{H_{jt}}{H_{jt}+L_{jt}} = \frac{\frac{H_{jt}}{L_{jt}}}{1+\frac{H_{jt}}{L_{jt}}}$. Moretti estimates: $w_{jt} = \beta\frac{H_{jt}}{H_{jt}+L_{jt}}$. Thus, $\frac{\partial w_{jt}}{\partial\ln\left(\frac{H_{jt}}{L_{jt}}\right)} = \frac{\partial w_{jt}}{\partial\frac{H_{jt}}{H_{jt}+L_{jt}}}\frac{\partial\frac{H_{jt}}{H_{jt}+L_{jt}}}{\partial\ln\left(\frac{H_{jt}}{L_{jt}}\right)} = \beta * \left(\left(\frac{H_{jt}}{H_{jt}+L_{jt}}\right)\left(1-\frac{H_{jt}}{H_{jt}+L_{jt}}\right)\right)$. Plugging in the average college share in 1990, 0.25 gives: $\frac{\partial w_{jt}}{\ln\left(\frac{H_{jt}}{L_{jt}}\right)} = \beta*(0.1875)$. Thus, I scale Moretti's estimates by 0.1875 to make them in the same units as my own.

movement along firm's labor demand curves with endogenous productivity spillovers, while Moretti controls for labor demand variation (he includes the local unemployment rate as well as other city controls). He also uses the lagged age structure of the city as a instrument for changes in cities' skill mix. Using this identification strategy, he finds slightly larger effects (point estimates become 0.39 for high school graduate wages and 0.16 for college wages.) While my estimates of the social returns to education are slightly above Moretti's, we both conclude that they are large and substantial.[38]

In Appendix A.3, I consider whether the observed population distribution across cities in each decade is a stable equilibrium assuming a simple dynamic adjustment process, given the parameter estimates of the model. Using a simple dynamic adjustment process where 5% of the population can choose to migrate to a new city each period, and cities' wages, rents, and amenities instantaneously respond to population changes each period, I find that the equilibria observed in the data in 1980, 1990, and 2000 are stable under this assumed adjustment dynamic. See Appendix A.3 for further details.

1.6.4 Estimation Robustness

To assess whether the parameter estimates of the model are sensitive to ways that I have measured wages, rents, and Bartik shocks, I re-estimate the model using a variety of different variable definitions. These results are in Table 1.5. A particularly important reason to assess whether the parameter estimates are stable across different specifications is to indirectly test for whether my instruments are weak. I am not aware of a standard econometric test for weak instruments when estimating a multiple equation, non-linear GMM model where there are multiple endogenous variables.

[38] Ciccone and Peri (2006) also estimate the productivity spillovers of education. However, they focus on the social return to an additional year of average education, without diffentiating between college and non-college years of education. They also use lagged age structure of a city as an instrument for the local skill mix, but do not find any evidence of spillovers. Since they do not explicitly analyze spillovers due to college versus. non-college skill mix, it is hard to compare exactly why these estimates differ. Their analysis also does not include the 2000 census.

Table 1.5: Robustness Tests of Alternative Model Specifications

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Drop Total Bartik Shock	Hedonic Adjusted Wage & Rent	Only Actual Rents	Only Imputed Rents	Rents from College Workers	Rents from Non-College Workers	Use Endog Amenity Index	Reduced Form Labor Demand
		Household Preferences for Cities							
College Workers:	Wage	1.883	3.337	1.382	1.131	1.499	1.337	5.294	1.723
		[0.437]	[0.345]	[0.288]	[0.276]	[0.289]	[0.259]	[0.622]	[0.339]
	Rent	-1.229	-2.597	-0.646	-0.857	-0.901	-1.020	-0.439	-0.563
		[0.459]	[0.373]	[0.285]	[0.216]	[0.254]	[0.264]	[0.377]	[0.267]
	Endog Amenity	4.275	3.167	2.838	3.223	2.982	3.362	2.292	3.352
		[1.189]	[0.601]	[0.417]	[0.521]	[0.502]	[0.551]	[0.852]	[0.613]
Differential Effects, Blacks:	Wage	4.366	7.825	3.847	2.302	3.220	2.815	7.460	2.993
		[1.654]	[1.385]	[1.326]	[1.151]	[1.224]	[1.186]	[1.365]	[1.185]
	Rent	-2.336	-6.339	-2.136	-1.545	-1.866	-2.021	-1.533	-1.331
		[0.846]	[0.932]	[0.771]	[0.448]	[0.488]	[0.575]	[0.595]	[0.544]
	Endog Amenity	3.397	4.091	3.406	3.116	2.753	3.633	2.722	3.497
		[1.428]	[0.976]	[0.854]	[0.715]	[0.718]	[0.779]	[1.325]	[0.835]
Differential Effects, Immigrants:	Wage	5.497	8.424	2.591	5.535	5.583	5.231	2.959	3.882
		[0.918]	[0.99]	[1.143]	[0.642]	[0.683]	[0.691]	[1.189]	[1.049]
	Rent	0.578	-2.130	2.464	0.028	0.027	0.331	0.265	0.573
		[0.709]	[0.523]	[0.714]	[0.338]	[0.384]	[0.431]	[0.563]	[0.521]
	Endog Amenity	-4.609	-1.913	-3.835	-2.045	-2.189	-2.196	0.017	-1.957
		[1.899]	[0.653]	[0.646]	[0.514]	[0.5]	[0.5]	[0.792]	[0.515]
Non-College Workers:	Wage	4.361	6.757	6.676	2.678	3.847	3.888	4.344	4.327
		[0.821]	[0.554]	[0.765]	[0.466]	[0.547]	[0.586]	[0.612]	[0.564]
	Rent	-3.379	-5.940	-4.223	-2.090	-2.614	-2.855	-2.247	-2.587
		[0.65]	[0.508]	[0.56]	[0.241]	[0.279]	[0.365]	[0.348]	[0.312]
	Endog Amenity	2.829	2.267	1.043	1.114	0.659	1.243	0.413	1.105
		[1.195]	[0.591]	[0.459]	[0.405]	[0.447]	[0.473]	[0.609]	[0.546]
Differential Effects, Blacks:	Wage	2.672	4.289	3.492	3.037	3.392	3.371	3.055	3.354
		[0.982]	[0.74]	[0.897]	[0.651]	[0.686]	[0.767]	[0.698]	[0.861]

Table 1.5 (Continued)

Rent	-0.754	-2.244	-0.565	-0.775	-0.938	-0.926	-0.816	-0.822
	[0.614]	[0.516]	[0.615]	[0.307]	[0.347]	[0.417]	[0.379]	[0.431]
College Emp Ratio Amenity	0.222	0.498	-0.496	0.186	-0.106	0.082	-0.123	0.288
	[0.932]	[0.468]	[0.437]	[0.404]	[0.392]	[0.412]	[0.609]	[0.416]
Differential Effects, Immigrants: Wage	1.944	2.929	1.679	2.839	2.704	2.551	3.620	2.547
	[1.286]	[0.828]	[0.953]	[0.794]	[0.844]	[0.932]	[1.112]	[1.075]
Rent	0.465	0.379	0.610	-0.063	0.050	0.134	-1.018	0.246
	[0.72]	[0.559]	[0.53]	[0.371]	[0.427]	[0.511]	[0.621]	[0.553]
College Emp Ratio Amenity	0.108	-0.687	1.165	0.780	0.712	0.676	-3.193	1.487
	[0.97]	[0.469]	[0.57]	[0.541]	[0.537]	[0.504]	[0.75]	[0.677]
Housing Supply								
Land Use Regulation	0.066	0.044	0.036	0.080	0.062	0.059	0.066	0.044
	[0.013]	[0.008]	[0.007]	[0.012]	[0.01]	[0.01]	[0.01]	[0.01]
Land Unavailability	0.021	0.016	0.020	0.036	0.025	0.028	0.017	0.020
	[0.01]	[0.007]	[0.006]	[0.01]	[0.009]	[0.008]	[0.008]	[0.007]
Base House Supply Elasticity	0.029	0.197	0.170	-0.211	-0.028	-0.032	0.033	0.082
	[0.119]	[0.063]	[0.056]	[0.104]	[0.081]	[0.081]	[0.08]	[0.083]
Labor Demand								
Rho	0.466	0.405	0.568	0.569	0.552	0.570	0.564	
	[0.073]	[0.038]	[0.044]	[0.045]	[0.044]	[0.042]	[0.066]	
Productivity Spillover on College Workers	0.573	0.616	0.556	0.560	0.557	0.557	0.704	
	[0.096]	[0.037]	[0.047]	[0.045]	[0.045]	[0.046]	[0.047]	
Productivity Spillover on Non-College	0.131	-0.001	0.221	0.195	0.173	0.207	0.289	
	[0.129]	[0.053]	[0.059]	[0.06]	[0.06]	[0.061]	[0.094]	
Labor Demand Elasticity Estimates for Reduced for Labor Demand Model								
College Wage wrt College Labor								0.389
								[0.082]
College wage wrt NonCollege Labor								-0.116
								[0.094]
NonCollege wage wrt College Labor								0.587
								[0.096]
NonCollege wage wrt NonCollege Labor								-0.422
								[0.104]

Notes: Standard errors in brackets. Standard errors clustered by MSA. See text for details on all the alternative model specifications.

Stock, Wright, and Yogo (2002) say "If identification is weak, then GMM estimates can be sensitive to the addition of instruments or to changes in the sample. If these features are present in an empirical application, then they can be symptomatic of weak identification." I try to do the best I can to "test" for weak identification by changing which instruments I include and how I define the sample's variables.

I first assess the sensitivity of including the "aggregate" Bartik shock as an additional instrument. The model can be identified using only the high skill and low skill Bartik shocks, along with their interactions with housing supply elasticities. I included the aggregate Bartik shock because it likely contains information about how each individual low and high skill Bartik shock influences local housing prices. Column 1 of Table 1.5 reports the parameter estimates when all of the moments using the aggregate Bartik shocks are dropped. The parameter estimates are extremely similar between this specification and the main specification. The standard errors increase slightly, which is expected when there are fewer identifying restrictions used to estimate the model. No parameters change sign and all of the changes in the parameter estimates are within the confidence bands of my preferred specification.

As additional robustness tests, I try a variety of methods for measuring local wages and rent. I re-estimate the model where I hedonically adjust local wage and rent changes. Wages are adjusted for a more fine measure of education, a quadratic in experience, gender, and race. I adjust rents by number of rooms, number of bedrooms, and apartment units in the structure. The model estimates using these wage and rent measures are qualitatively the same. High and low skill workers are slightly more elastic to cities' wages and rents, but high skill workers remain significantly less price sensitive than low skill workers.

I also measure local rent changes in a variety of ways. I try using only rent data from housing units actually rented, dropping the imputed rents. I also try using only the imputed rents. Further, I estimate the model using only rents from college workers, as well as rents only from non-college workers. The results are very similar across all of these specifications. This shows that the variation of housing costs between cities does not strongly differ across home owners versus home rents or between the neighborhoods of the high skill workers or low skill workers. Overall, the parameter estimates are stable across a variety of specifications, which alleviates some concern about whether the model is weakly identified.

I also consider an alternative measure of cities' endogenous amenities using the observable amenity data. The main model specification uses the college employment ratio as the index of endogenous amenities within a city. This index captures both the direct benefits of having more educated neighbors, as well as the indirect benefits through amenities such as more restaurants, lower crime, or better air quality. To assess whether these observable amenities can capture the full extent of the increased desirability of high skill cities, I create an alternative endogeneous amenity index. I compute changes in the observable amenity index by measuring the average change in logs of each amenity measure in each decade.[39] Specifically, the amenity index is defined as:

$$\Delta amen\_index_{jt} = \frac{1}{K}\sum_{k}\Delta\ln\left(amen_{kjt}\right).$$

Column 7 of Table 1.5 shows the re-estimated model where this observable amenity index is used in place of the college employment ratio for workers' preferences for cities. Similar to the main model estimates, I find that elasticity of college workers' demand for a city with a high amenity index is substantially larger than lower skill workers' demand elastcity (2.29 (0.85) for college versus 0.413 (0.61) for non-college). However, the model estimates using this amenity index still show that college workers appear to allocate a very small share of their expenditure to housing and local goods, relative to non-college workers. These estimates would imply that college workers only spend 8% of their expenditure housing and local goods, while non-college workers spend 52%. This suggests that the college employment

[39]The amenity measures included are grocery stores per 1000 residents, apparel stores per 1000 residents, eating and drinking places per 1000 residents, book stores per 1000 residents, dry cleaners per 1000 residents, movie thearters per 1000 residents, property crimes per 1000 residents, violent crimes per 1000 residents, and the EPA air quality index. I do not include museums and art galleries because data is missing for a substantial number of MSAs. Since the air quality data is missing for a number of MSAs, espcially in 1980, I impute air quality by regressing non-missing air quality on the other amenities, as well as future period's airquality measures. I then impute the missing airquality data from this regression.

ratio is capturing desirable amenities of the city, over and beyond these observable measures. This is not surprising, since Table 1.2 showed that while the college employment ratio was positively correlated with many of these desirable amenities, the college employment ratio could explain only a very small fraction of the varition in any of these obsevable amenity measures. A high college share city appears to be desirable for reasons over and beyond these measures of observable amenities.

As a final robustness test, I consider an alternative functional form for the local labor demand curves. Using the standard constant elasticity of substitution production function restricts wages to only depend on the the relative supply of college and non-college workers, and not aggregate amount of labor in the city. To assess whether the model's labor demand elasticity estimates are being restricted due to the functional form assumptions, I estimate reduced form local labor demand curves. I replace the CES labor demand curves with the following log-linear labor demand curves:

$$\begin{aligned} \Delta w_{jt}^{H} &= \alpha_{t}^{H} + \beta^{HH} \Delta \log \left(H_{jt} \right) + \beta^{HL} \Delta \log \left(L_{jt} \right) + \Delta \varepsilon_{jt}^{H} \\ \Delta w_{jt}^{L} &= \alpha_{t}^{L} + \beta^{LH} \Delta \log \left(H_{jt} \right) + \beta^{LL} \Delta \log \left(L_{jt} \right) + \Delta \varepsilon_{jt}^{L}. \end{aligned}$$

Column 8 of Table 1.5 reports the model estimates using these reduced-form labor demand curves. None of the labor demand elasticity estimates change sign from the main model estimates. The reduced form estimates show that a 1% increase in non-college labor leads only to a 0.12% decline in college worker wages. The magnitude of this estimate is smaller than the main model estimate of a 0.59% decline in college worker wages. The other reduced form labor demand estimates are of similar magnitudes to those in the main model estimates. Overall it seems the functional form restrictions of the CES production function are not restraining the labor demand elasticities to be extremely different from those found in the reduced form estimates.

1.7 Amenities & Productivity Across Cities

Using the estimated parameters, one can infer the productivity of local firms and the desirability local amenities in each city. Comparing the model's predicted amenity and productivity levels to outside research and data on these city characteristics will assess model fit and provide a sanity check of whether these estimates appear plausible.

There is a large literature which attempts to estimate which cities offer the most desirable amenities using hedonic techniques.[40] The methodology used in this paper to infer which cities offer the most desirable amenities differs from the previous hedonic literature. Recalling equation (1.24) , the utility value of the amenities in a city to workers of a given demographic group ($\beta^{\text{col}} z \frac{H_{jt}}{L_{jt}} + \xi_{jt}^{z}$) is measured by the component of the workers' common utility level for each city, δ_{jt}^{z}, which is not driven by the local wage and rent level, $\beta^{w} z w_{jt}^{edu} + \beta^{r} z r_{jt}$. The utility workers of type z receive from the amenities in city j in year t, U_A^z, is thus:

$$U_A^z = \beta^{\text{col}} z \frac{H_{jt}}{L_{jt}} + \xi_{jt}^{z} = \delta_{jt}^{z} - \beta^{w} z w_{jt}^{edu} - \beta^{r} z r_{jt}.$$

Intuitively, amenities are inferred to be highest in cities which have higher population levels of a given demographic group than would be expected, given the city's wage and rent levels and workers' preferences for wages and rent. If many workers choose to live in a city that appears relatively undesirable based on its wages and rent, as compared to other cities, it must offer desirable amenities. Unlike the literature using hedonics to rank city amenities, these methods are able to rank the desirability of cities amenities separately for different types of workers.

A test of whether the model fits the data well is to asses whether the amenity rankings appear "intuitive." Of the largest 75 cities, as measured by their population in 1980, Appendix

[40] The hedonic methods infer a city's amenities by directly comparing local real wages across cities. In a model where workers have homogeneous preferences for cities, the equilibrium local real wages across cities must be set to equate all workers utility values in all cities. In equilibrium, the difference in real wages across cities is a direct measure of the amenity value of the city. A low amenity city must offer a high real wage in order to offer the same utility as a high amenity city. See Albouy (2008) for recent amenity estimates using these techniques.

Table 1.3 reports the top 10 cities with the most desirable and undesirable amenities for college and non-college workers in 1980 and 2000, as well as the cities with the largest improvements and declines in amenities during this time period. In 2000, Los Angeles-Long Beach, CA had the most desirable amenities for non-college workers, followed by Phoenix, AZ, Denver-Boulder, CO, Seattle-Everett, WA, and Boston, MA-NH. The cities with the most desirable amenities for college workers in 2000 were: Los Angeles-Long Beach CA, Washington, DC/MD/VA, Boston, MA-NH, Atlanta, GA, and San Francisco-Oakland-Vallejo, CA. These cities are known to have vibrant cultural scenes, desirable weather, and often considered to have high quality-of-life.

The least desirable city amenities for college workers in 2000 are located in Youngstown-Warren, OH-PA, which is followed by Allentown-Bethlehem-Easton, PA/NJ, Syracuse, NY, Harrisburg-Lebanon–Carlisle, PA, and Scranton-Wilkes-Barre, PA. Similarly, non-college workers find the least desirable amenities in Youngstown-Warren, OH-PA, followed by Toledo, OH/MI, Syracuse, NY, Buffalo-Niagara Falls, NY, and Allentown-Bethlehem-Easton, PA/NJ. All of these cities are located in America's Rust Belt, where the cities have historically had high levels of pollution due the concentration of manufacturing jobs. They have recently faced large declines in manufacturing jobs, population declines, and growing crime rates since the 1980s. It is not surprising these cities offer the least desirable amenities. Overall, the model seems to do well at capturing which cities are the most and least desirable.

A similar validation test can be done by analyzing which cities have the highest and lowest productivity levels. The productivity level in each city for each skill level of worker can be inferred from the residual of the labor demand equation:

$$\text{Prod}_{jt}^{H} = \ln w_{jt}^{H} - ((1-\rho)\ln N_{jt} + (\rho-1)\ln H_{jt}) \quad (1.25)$$

$$\text{Prod}_{jt}^{L} = \ln w_{jt}^{L} - ((1-\rho)\ln N_{jt} + (\rho-1)\ln L_{jt})\,. \quad (1.26)$$

Intuitively, productivity is inferred to be the highest in areas where wages are much higher than what would be predicted by the amount of labor supplied to an area. Appendix Table 1.4 reports the most and least productive cities in 1980 and 2000 for college and non-college workers. In 2000, the most productive city for college graduates was San Jose, CA, followed by San Francisco-Oakland-Vallejo, CA, New York-Northeastern NJ, Washington, DC/MD/VA, and Boston, MA-NH. These cities are the hubs of many of the most productive industries such as high tech in Silicon Valley, finance in New York, and biotech in Boston and San Francisco.

The most productive cities for lower skill workers are very different. In 2000, the city most productive for low skill workers was Riverside-San Bernardino,CA, followed by Detroit, MI, Las Vegas, NV, Tacoma, WA, and Ventura-Oxnard-Simi Valley, CA. Riverside-San Bernardino, CA is where many of the largest manufacturing companies have chosen to place their distribution centers. These centers transport finished goods and materials from the ports surrounding Los Angeles to destinations around the US. Shipping and distribution provide many relatively high paying jobs for low skill workers here. While Detroit, MI ranks as the second most productive area in 2000, it is also in the top 10 for cities which have experienced the largest productivity declines for low skill workers from 1980 to 2000. In 1980, Detroit, MI was the most productive city for low skill workers, but as American auto manufacturing has lost market share, wages and jobs have fallen here. However, Detroit still provides some of the most productive uses of low skill labor in 2000.

In Appendix A.4, I compare my productivity estimates with other work studying the determinants of productivity changes across cities. My estimates are consistent with work by Autor and Dorn (2012) who study the differences in skill biased technology change across cities driven by historical occupation differences. Work by Beaudry, Doms, and Lewis (2010) studying differences in technology adoption of firms across cities also lines up with my estimated productivity changes. Additionally, the cities which I find to have the largest declines in low skill productivity closely overlap with the cities' which had larger amounts of their labor force employed in industries that faced the most competition from the increased penetration of Chinese exports (Autor, Dorn, and Hanson (2012)).

There are striking difference in which cities have had the largest changes in their local high and low skill productivity from 1980 to 2000. Appendix Table 1.4 shows that three cities (San Jose, CA, Boston, MA-NH, and Chicago, IL) rank in the top 10 largest productivity *increases* for high skill workers and the top 10 largest productivity *decreases* for low skill workers. Table 1.6 presents a regression of the model's predicted change in a city's high skill productivity on its predicted change in low skill productivity. I find a slight negative relationship between local high skill productivity change, and local low skill productivity. Note that this negative relationship between changes in local high skill productivity and low skill productivity cannot be seen by simply comparing changes in local high skill wages with changes in local low skill wages. Table 1.6 shows that changes in high and low skill wages are strongly positively correlated, with an R squared of 0.50. Movement along local labor demand curves driven by migration masks the large differences local productivity changes by skill.

The differences in high and low skill workers' preferences for a city's amenities is unlikely to differ by the same magnitude. The preference estimates of the model showed that both high and low skill workers agreed that the amenity changes caused by an increase in a city's college employment ratio were desirable. Thus, one would expect that college workers' overall utility value for a city's amenities to be positively associated with non-college workers' utility value for the same city's amenities. Table 1.6 shows that the utility value of college and non-college amenity changes across cities are strongly positively correlated. Changes in non-college workers' utility due to changes in cities' amenities explains 58% of the variation in changes in college workers' utility for the same cities' amenities.

Table 1.6: Relations between Amenity and Productivity Changes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Δ College Amentity	Δ Non-College Amenity	Δ College Amentity	Δ Non-College Amenity	Δ College Amentity	Δ College Productivity	Δ College Wage
Δ Non-College Amenity					0.595***		
					[0.0310]		
Δ College Productivity	1.039***	-0.037					
	[0.274]	[0.360]					
Δ Non-College Productivity			0.615	-1.289**		-0.062	
			[0.402]	[0.512]		[0.0883]	
Δ Non-College Wage							0.6879***
							[0.0426]
Constant	-0.342***	-0.394***	-0.0898***	-0.429***	0.137***	0.229***	0.146***
	[0.0669]	[0.0880]	[0.0242]	[0.0307]	[0.0194]	[0.00529]	[0.0046]
Observations	267	267	267	267	267	268	268
R-squared	0.052	0.00	0.009	0.023	0.581	0.002	0.500

Standard errors in brackets. Changes in amenities and productivities are measured between 1980 and 2000. Cities' amenities and productivity levels are inferred from model estimates. See text for further details. *** p<0.01, ** p<0.05, * p<0.1

Table 1.6 shows that changes in college productivity are positively associated with changes in college amenities. However, changes in non-college productivity are negatively associated with non-college amenity changes. Industries which are the most productive for low skill workers may negatively impact the amenities of the city. For example, Kahn (1999) finds the decline of manufacturing in Rust Belt cities led to increases in local air-quality. The model allows for this effect through the college employment ratio. However, the industries most productive for college workers tend to be driven by research, ideas and innovation. Even though high skill production might not directly spillover onto a city's amenities, cities which attract high skill workers by offering high wages benefit from the endogenous amenity increases, driven by the presence of high skill workers living in the city.

The inferred local productivity and amenity changes across cities appear consistent with outside knowledge on these measures, and the relationships between productivity and amenities changes also appear intuitive. These tests of the model provide evidence that the model setup and parameter estimates capture the key underlying drivers of the determinants of local wages, rents, and amenities across cities.

1.7.1 The Determinants of Cities' College Employment Ratio Changes

To understand what drove the changes in cities' college employment ratios from 1980 to 2000, I use the estimated model to assess the contributions of productivity, amenities, and housing supply elasticities to the changes in cities' college employment ratios.

1.7.2 College Employment Ratio Changes and Productivity

I first consider how much of the observed changes in cities' college employment ratios can be explained by changes in cities' exogenous productivity levels. Changes in local productivity directly impact wages, but also influence local prices and endogenous amenities through migration. Before considering how local productivity changes are capitalized into equilibrium wages, rents, and endogenous amenities, I consider the direct effect of productivity changes on local wages. Using the exogenous productivity changes predicted by the model, I compute the direct effect of exogenous changes in productivity from 1980 to 2000 in each city on local high and low skill wages. These counterfactual college and non-college wages in 2000, $\hat{w}^H_{j2000}$ $\hat{w}^L_{j2000}$ are defined as:

$$
\begin{aligned}
\hat{w}^H_{j2000} &= \underbrace{c_{1980} + (1-\rho)\ln \hat{N}_{j2000} + (\rho-1)\ln H_{j1980} + \gamma_H \ln\left(\frac{H_{j1980}}{L_{j1980}}\right)}_{\text{Labor Supply in 1980}} + \underbrace{\varepsilon^H_{j2000}}_{\text{Exog. Productivity in 2000}} \\
\hat{w}^L_{j2000} &= \underbrace{c_{1980} + (1-\rho)\ln \hat{N}_{j2000} + (\rho-1)\ln L_{j1980} + \gamma_L \ln\left(\frac{H_{j1980}}{L_{j1980}}\right)}_{\text{Labor Supply in 1980}} + \underbrace{\varepsilon^L_{j2000}}_{\text{Exog. Productivity in 2000}} \\
\hat{N}_{j2000} &= \left(\exp\left(\varepsilon^L_{j2000}\right)\left(\frac{H_{j1980}}{L_{j1980}}\right)^{\gamma_L} L^{\rho}_{j1980} + \exp\left(\varepsilon^H_{j2000}\right)\left(\frac{H_{j1980}}{L_{j1980}}\right)^{\gamma_H} H^{\rho}_{j1980}\right)^{\frac{1}{\rho}}.
\end{aligned}
$$

The counterfactual wages only reflect the shifts in local labor demand curves driven by the exogenous changes in local productivity from 1980 to 2000, but not the movement along cities' labor demand curves or endogenous productivity changes due to migration.

Using these counterfactual year 2000 wages, while holding rents and amenity levels fixed at their 1980 levels, I use the model to predict where workers would have chosen to live if they had to choose among this set of hypothetical cities. Specifically, worker i's utility for hypothetical city j is:

$$V_{ijt} = \beta^w z_i \hat{w}^{edu}_{j2000} - \beta^r z_i r_{j1980} + \xi^z_{j1980} + \beta^{\text{col}} z_i \frac{H_{j1980}}{L_{j1980}} + \beta^{st} z_i st_i x^{st}_j + \beta^{\text{div}} z_i \,\text{div}_i\, x^{\text{div}}_j + \varepsilon_{ij80}.$$

The predicted cities' college employment ratios from this hypothetical world are then compared to those observed in the data. This counterfactual scenario assesses whether the cities which became disproportionately productive for college, relative to non-college workers, were also the cities which experienced disproportionate growth in their college versus non-college populations. Figure 1.6 panel A plots the observed college employment ratio changes against these predicted counterfactual changes.

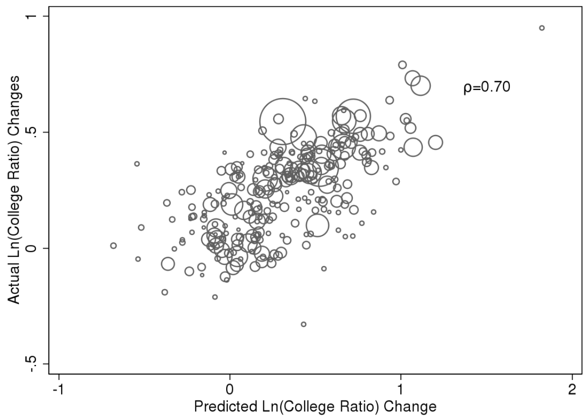


Figure 6.A Predicted change in college ratio due to local wage changes from productivity changes

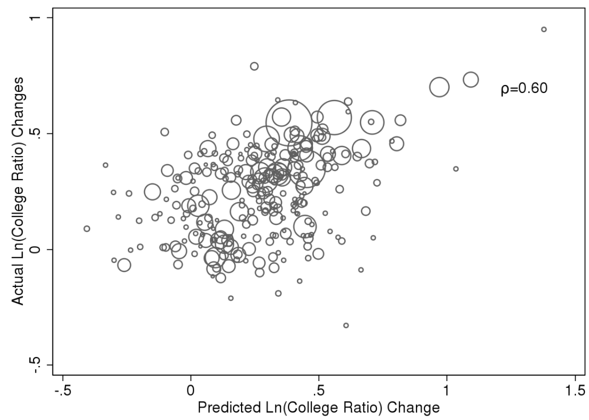


Figure 6.B Predicted change in ln college ratio due to observed changes in local wages & rent

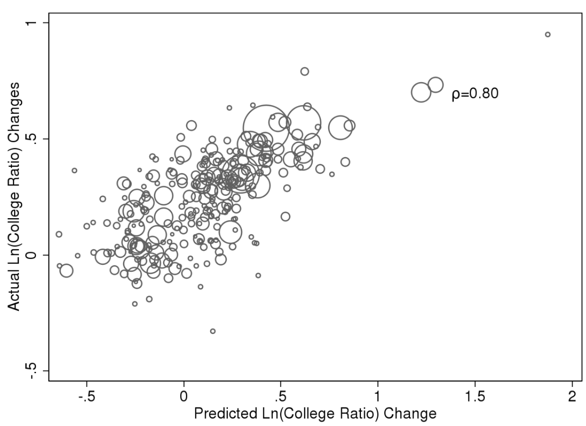


Figure 6.C Predicted change in ln college ratio due to changes in observed changes in wages, rent, and endogenous amenities

Figure 1.6: Predicted Changes in Ln College Employment Ratio: 1980-2000

The predicted and actual changes are strongly correlated with a correlation coefficient of 0.70. This shows local productivity changes explain a large share of the changes in cities' local college employment ratios from 1980 to 2000.

While simple shifts in cities' wages due to productivity changes explain a large amount of the variation in college employment ratio changes, workers' actual migration decisions depended on how local productivity changes influenced the overall desirability of cities's wages, rents, and amenities. For example, productivity growth in a city with very inelastic housing supply would attract many fewer workers than a similar city with elastic housing supply. Additionally, the amenities of the city will respond to the mix of college and non-college workers living in the city.

In a model where amenities are assumed to be exogenous, the only ways which productivity changes can influence workers' location decisions is by influencing local wages and rents. To measures the importance of endogenous amenity changes in influencing cities' college employment ratio changes, I use the model to predict workers' city choices in 2000, using only the observed changes in wages and rent. Holding amenities fixed at the 1980 levels, I set local wages and rents to the levels observed in 2000. Specifically, worker i's utility for hypothetical city j is:

$$V_{ijt} = \beta^{w} z_i w_{j2000}^{edu} - \beta^{r} z_i r_{j2000} + \xi_{j1980}^{z} + \beta^{\text{col}} z_i \frac{H_{j1980}}{L_{j1980}} + \beta^{st} z_i st_i x_j^{st} + \beta^{\text{div}} z_i \,\text{div}_i\, x_j^{\text{div}} + \varepsilon_{ij80}.$$

I predict where workers would have chosen to live if they had to choose from this set of counterfactual cities. If endogenous amenity changes were not an important factor in how productivity changes influenced cities' college employment ratio changes, then local wage and rent changes should be at least as strong of a predictor of college employment ratio changes as simply shifting wages by productivity changes. Figure 1.6 panel B plots the observed college employment ratio changes against these counterfactual predicted college employment ratio changes. The correlation of the predicted versus actual college employment ratio changes falls by 14% to 0.60. Local wage and rent changes have less predictive power in explaining cities' college employment ratio changes than the direct impact of productivity changes. This suggests endogenous amenity changes are an important component in changing cities' college employment ratios.

To test how much predictive power the endogenous amenity changes have in explaining college employment ratio changes, beyond wages and rent, I create a third set of counterfactual cities. These cities hold the exogenous amenities fixed at their 1980 levels, but allow wages, rents, and endogenous amenities to shift to the level observed in 2000. Specifically, worker i's utility for hypothetical city j is:

$$V_{ijt} = \beta^{w} z_i w^{edu}_{j2000} - \beta^{r} z_i r_{j2000} + \xi^{z}_{j1980} + \beta^{\text{col}} z_i \frac{H_{j2000}}{L_{j2000}} + \beta^{st} z_i st_i x^{st}_j + \beta^{\text{div}} z_i \, \text{div}_i \, x^{\text{div}}_j + \varepsilon_{ij80}.$$

I use the model to predict where workers would have chosen to live within this set of counterfactual cities. Figure 1.6 panel C plots actual college employment ratio changes against these predicted changes due to wages, rents, and endogenous amenities. The correlation coefficient is now 0.80, a 30% increase relative to the predictive power of wage and rent changes alone. The combination of wage, rent, and endogenous amenity changes have more predictive power than the productivity shifts alone, which shows that the endogenous amenity response was a key mechanism in how local productivity changes led to migration changes.

1.7.3 Corroborating Reduced Form Evidence

As an alternative method to assess the role of local productivity changes in driving local migration patterns, I analyze the reduced form relationship between the observable Bartik local labor demand shocks and changes in cities college employment ratios. To measure the labor demand impacts of the Bartik shocks directly on migration, I measure the Bartik shocks using changes in nationwide industry employment.[41] I regress changes in cities' college

[41] Since the Bartik shocks were used as instruments for wages in estimation, they offer the most power when measured in wage units. However, when measuring the reduced form impact of Bartik labor demand shocks on migration, it is useful to measure the industry shock in the same units as migration: counts of

employment ratios on the college and non-college Bartik local labor demand shocks. The results are in Table 1.7. Consistent with the findings of Moretti (2011b), I find that high skill labor demand shocks strongly predict increases in cities' college employment ratios, while low skill labor demand shocks are negatively predictive. Further, the R-squared of this regression shows that over 40% of the variation in changes in cities' college employment ratios can be explained by changes in industrial labor demand for high and low skill workers. Both these observable local productivity changes, as well as the model estimated local productivity changes, show that the local productivity changes were the underlying drivers of changes in cities' college employment ratios.

As an additional test for the importance of endogenous amenity changes, I analyze the relationship between the observed Bartik labor demand shocks and changes in local real wages. The model estimates showed that both college and non-college workers spent 67% of their expenditure on locally priced goods. Thus, I measure local real wages as:

$$\text{local real wage}_{jt}^{edu} = w_{jt}^{edu} - 0.67 * r_{jt}.$$

A regression of changes in college workers' local real wages on Bartik labor demand shocks in Table 1.7 show that an increase in the college Bartik labor demand *decreases* college workers' local real wages. As previously discussed, the college Bartik shocks strongly predicted *increases* in the local college employment ratio. It is hard to rationalize why college workers would disproportionately migrate to areas with decreases in local real wages, unless the Bartik shocks caused increases in local amenities. In contrast, Table 1.7 shows that increases in non-college Bartik shocks lead to *increases* in non-college workers' local real wages, which

workers. Additionally, Bartik measured in wage changes for each skill group should impact wages, and thus migration of both skill groups, due to imperfect labor substitution within firms. Bartik measured in quantities, thus, should give a more precise measure of the increase in quantities of labor demanded for high and low skill workers. Bartik shocks measured in quantities are defined as: $\Delta B_{jt}^{H,\text{quant}} = \sum_{ind} s_{jt-1}^{H,ind} \Delta \ln H_{(-j)t}^{ind}$, $\Delta B_{jt}^{L,\text{quant}} = \sum_{ind} s_{jt-1}^{L,ind} \Delta \ln L_{(-j)t}^{ind}$, where $s_{jt-1}^{H,ind}$ and $s_{jt-1}^{L,ind}$ are the share of college and non-college workers employed in industry *ind* in city j in 1980, respectively. $\Delta \ln H_{(-j)t}^{ind}$ measures the change between 1980 and 2000 in the log of number of jobs for college graduates in industry *ind* nationwide, excluding city j. $\Delta \ln L_{(-j)t}^{ind}$ is a similar measure for non-college workers.

shows the in-migration of low skill workers did not endogenously increase local amenities. Additionally, Table 1.7 shows the OLS relationship between changes in local college employment ratios and cities' local real wages for high and low skill workers. Increases in local college employment ratios are associated with *decreases* in local real wages for both college and non-college workers. Earlier reduced form facts presented in Section 3 showed changes in cities' college employment ratios are positively associated with increases in both wages and rents. However, the rent increases are so large, that they overwhelm the wage increases, leading to local real wage decreases. This further shows that local amenities changes played an important role in the observed migration patterns.

As a direct test of the impact of local Bartik labor demand shocks on endogenous amenity changes, Table 1.8 reports regressions of a set of observable amenity changes on high and low skill Bartik labor demand shocks. I find apparel stores per capita and eating and drinking places per capita positively respond to increases in labor demand for college workers. I also find property crime rates and pollution levels fall in response to increase in labor demand for college workers. Similarly, increase in non-college labor demand causes decreases in grocery stores per capita, apparel stores per capita, eating and drinking places per capita, and movie theaters per capita. These results show that in-migration of college workers leads to increases in a large set of amenities, while in-migration of low skill workers, leads to decreases of amenities. This further confirms the structural model estimates which showed workers' migration responses to Bartik shocks could not be rationalized without the college employment ratio directly influencing local amenities.

Table 1.7: Reduced Form Relationships between College Employment Ratios, Local Real Wages & Local Employment Shocks

	(1)	(2)	(3)	(4)
	Δ College Employment Ratio	Δ College Employment Ratio	Δ College Local Real Wage	Δ Non-College Local Real Wage
Δ College Local Real Wage	-0.500***			
	[0.183]			
Δ Non-College Local Real Wage	-0.783***			
	[0.147]			
College Bartik Employment Shock		2.375***	-0.651***	-0.780***
		[0.176]	[0.0745]	[0.0912]
Non-College Bartik Employment Shock		-1.588***	0.283***	0.240**
		[0.179]	[0.0757]	[0.0927]
Constant	0.405***	-0.632***	0.442***	0.398***
	[0.0259]	[0.0766]	[0.0324]	[0.0397]
Observations	268	268	268	268
R-squared	0.257	0.409	0.256	0.276

Notes: Standard errors in brackets. Changes measured between 1980-2000. Weighted by MSA population in 1980. College employment ratio is defined as the ratio of number of employed college workers to the number of employed lower skill workers living in the city. Δ Real Wage = Δ ln(Wage)-.67*ΔLn(Rent). Bartik employment shock measure uses national industry changes in log employment quantities, weighted by the share of a city's employees employed in the industry. See text for futher details. *** p<0.01, ** p<0.05, * p<0.1

Table 1.8: Amenity Responses to Bartik Employment Shocks: 1980-2000

	(1)	(2)	(3)	(4)	(5)
	Grocery Stores per 1000 Residents	Apparel Stores per 1000 Residents	Eating and Drinking Places per 1000 Residents	Book Stores per 1000 Residents	Dry Cleaners per 1000 Residents
Non-College Bartik Employment Shock	-0.768***	-0.958***	-0.716***	-0.265	0.399
	[0.277]	[0.301]	[0.171]	[0.520]	[0.457]
College Bartik Employment Shock	0.0131	1.269***	0.523**	-0.642	0.596
	[0.367]	[0.399]	[0.226]	[0.689]	[0.606]
Constant	0.418**	-0.767***	0.170	0.634*	-0.346
	[0.175]	[0.190]	[0.108]	[0.328]	[0.289]
Observations	218	218	218	218	218
R-squared	0.054	0.055	0.076	0.013	0.021
	(6)	(7)	(8)	(9)	(10)
	Movie Thearters per 1000 Residents	Museums and Art Galleries per 1000 Residents	Propery Crimes per 1000 Residents	Violent Crimes Per 1000 Residents	EPA Air Quality Index
Non-College Bartik Employment Shock	-1.594***	-0.326	-0.221	-0.525	0.599
	[0.557]	[0.826]	[0.375]	[0.520]	[0.560]
College Bartik Employment Shock	1.194	1.421	-2.040***	-1.068	-1.739**
	[0.738]	[1.056]	[0.497]	[0.689]	[0.761]
Constant	-1.008***	0.0680	1.387***	0.763**	0.695*
	[0.352]	[0.499]	[0.237]	[0.328]	[0.360]
Observations	218	174	215	215	177
R-squared	0.037	0.013	0.136	0.040	0.031

Notes: Standard errors in brackets. Changes measured between 1980-2000. Retail and local service establishments per capita data come from County Business Patters 1980, 2000. Crime data is from the FBI. Air Quality Index in from the EPA. Higher values of of the air quality index indicate more pollution. *** p<0.01, ** p<0.05, * p<0.1

Overall, the underlying cause of the increase in sorting of workers by skill levels across cities was due to an increases in sorting of productivity by skill across cities. As presented earlier, Figure 1.2 shows the widening of the distribution of college employment ratios across cities from 1980 to 2000. Figure 1.7 provides a similar histogram of the ratio of college productivity to non-college productivity across cities for each decade. The productivity ratio in each city is defined as:

$$\text{Productivity Ratio}=\frac{\exp\left(\text{Prod}_{jt}^{H}\right)}{\exp\left(\text{Prod}_{jt}^{L}\right)},$$

where Prod_{jt}^{H}, Prod_{jt}^{L} are the local productivities of high and low skill workers, as defined in equations $(1.25),(1.26)$.[42] As seen in Figure 1.7, the productivity ratio distribution across cities exhibits the same widening over decades as seen in the college share distribution across cities in Figure 1.2.

In Appendix A.5, I analyze the role of housing supply elasticity heterogeneity in explaining the increase in geographic skill sorting. As the nation's population grows, the cities with less elastic housing supply experience relatively higher increases in housing rents. Since low skill workers are more price sensitive, the rent increases in inelastic cities lead to more out migration of low skill than high skill workers. This, in turn, leads to larger increases in the college employment ratios of housing inelastic cities. Increases in a city's college employment ratio raise all workers' wages, as well as increase the city's amenities. These effects will mitigate the out-migration driven by the initial increase in the city's rent, keeping the rent high in these housing inelastic cities. Thus, nationwide population growth over time should lead to larger increases in wages, rents, and college employment ratios in housing inelastic cities. While I find evidence for these effects, they only explain a small amount of the changes in cities' college employment ratios. The dominant cause is local productivity change.

[42] The productivities are exponentiated since they are measured in log wage units. This makes the ratio of high and low skill productivity comparable to the college employment ratio because both now are measured in levels, not logs.

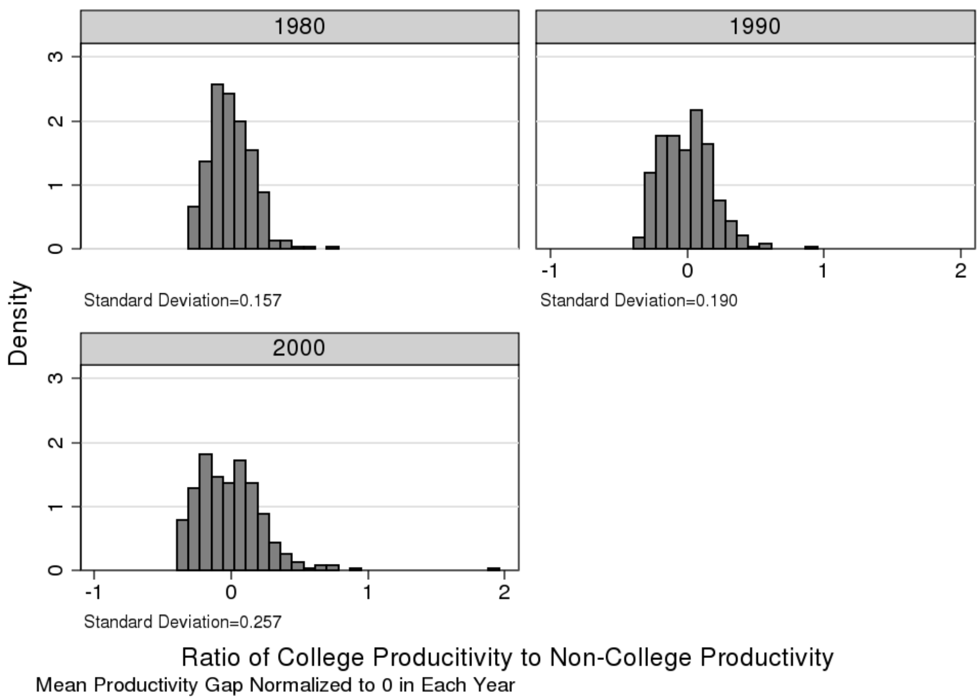


Figure 1.7: Distribution of College Productivity Gap Across MSAs: 1980-2000

1.7.4 Summary of Skill Sorting Mechanisms

To summarize, I find the underlying cause of the relative migration choices of high and low skill workers to be changes in high and low skill workers' relative productivity across cities. The structural model estimates show that a migration response to local productivity changes which increases a city's college employment ratio further increases the wages of all workers in the city and increases the local amenities. Wages rise due the combined effect of endogenous productivity changes and movement along firm's labor demand curves. Amenities such as pollutions levels, crime rates, and the quality of the local goods market improve in response to increases in the college employment ratio.

The combination of desirable wage and amenity growth for all workers led to large amounts of in-migration of both high and low skill workers. College workers were particularly attracted by desirable amenities, while low skill workers were particularly attracted by desirable wages. The increase in population in high college share cities led to large rent increases. Since low skill workers are more price sensitive, the increases in rent disproportionately discouraged low skill workers from living in these high wage, high amenity cities. Thus, in equilibrium, college workers sort into high wage, high rent, high amenity cities. Lower skill workers are not willing to pay the "price" of a lower local real wage to live in high amenity cities, leading them to chose more affordable, low amenity cities.

1.8 Welfare Implications & Well-Being Inequality

It is well documented that the nationwide wage gap between college workers and high school graduates has increased significantly from 1980 to 2000. Table 1.9 shows that the nationwide college wage gap has increased by 0.20 log points.[43] However, increases in wage inequality do not necessarily reflect increases in well-being inequality. College workers increasingly chose

[43] I focus on the college graduate-high school graduate wage gap because most of the literature has used this as a key wage inequality statistic. My model assumes all non-college workers face the same wage diffentials across cities. To make the welfare analysis comparable to the college-high school wage gap, I adjust the non-college workers' wages nationwide to represent the wages of a high school graduate, instead of the typical non-college worker. This does not impact the relative wages across cities.

to live in cities with higher wages, high rents, and more desirable amenities than non-college workers. The additional welfare effects of local rents and amenities could either add to or offset the welfare effects of wage changes.

First, I look only at wage and rent changes. I measure changes in the college "local real wage gap", where I define a worker's local real wage as his utility from wages and rent, measured in log wage units:

$$\text{Local Real Wage}_{ijt} = w_{jt}^{edu} - \frac{\beta^r z_i}{\beta^w z_i} r_{jt}.$$

$\frac{\beta^r z_i}{\beta^w z_i}$ measure worker i's expenditure share on locally priced goods. Table 1.9 reports changes in the college local real wage gap from 1980-2000. Similar to the findings of Moretti (2011b), I find the local real wage gap has increased 0.15 log points, 24% less than the increase in the college wage gap. However, this is not a full welfare metric.

Part of the reason college workers chose to pay such high housing rents was because they gained utility from the areas' amenities. To measure how changes in cities' wages, rents, and amenities each contributed to well-being inequality, I conduct a welfare decomposition. First, I measure each worker's expected utility change from 1980 to 2000 if the only cities' wages had changed, but local rents and amenities had stayed fixed. A worker $i's$ expected utility in 1980 is measured by the expected utility he would receive from living in his first choice city:

$$\begin{aligned} E\left(U_{i1980}\right) &= E\left(\max_j V_{ij1980}\right) \\ V_{ij1980} &= \beta^w z_i w_{j1980}^{edu} - \beta^r z_i r_{j1980} + \xi_{j1980}^z + \beta^{\text{col}} z_i \frac{H_{j1980}}{L_{j1980}} \\ &\quad + \beta^{st} z_i st_i x_j^{st} + \beta^{\text{div}} z_i \operatorname{div}_i x_j^{\text{div}} + \varepsilon_{ij1980}. \end{aligned}$$

Table 1.9: Changes in Inequality due to Wages, Rents, & Endogenous Amenities

Observed Changes in Wages & Local Real Wages: 1980-2000			
	(1)	(2)	(3)
Year	College-High School Grad Wage Gap	College-High School Grad Rent Gap	Local Real Wage Gap
1980	0.417	0.045	0.386
	[0.0011]	[0.0003]	[0.0011]
1990	0.561	0.130	0.474
	[0.0008]	[0.0006]	[0.0009]
2000	0.616	0.116	0.538
	[0.0007]	[0.0003]	[0.0008]
Change: '80-'00	0.199	0.071	0.152

Decomposition of Well-Being Inequality:Wages, Rents, & Endogenous Amenities: 1980-2000				
	(1)	(2)	(3)	(4)
Year	Well-being Gap Due to Wages	Well-being Gap Due to Wages & Rent	Well-being Gap Due to Wage, Rent, + Endog. Amenities from Re-Sorting Across Cities	Well-being Gap Due to Wage, Rent, Endog. Amenites from National Growth in College Ratio + Re-Sorting
1980	0.417	0.417	0.417	0.417
	-	-	-	-
1990	0.529	0.510	0.543	0.643
	[0.0011]	[0.0125]	[0.0147]	[0.0218]
2000	0.609	0.592	0.655	0.877
	[0.0015]	[0.0189]	[0.0216]	[0.0396]
Change: '80-'00	0.193	0.175	0.239	0.461
	[0.0015]	[0.0189]	[0.0216]	[0.0396]
Wald Test: Well-being Impact vs. Well-being Impact of Wages	-	-1.018	2.277**	6.555***

Notes: Wage gap measured the log wage difference between college and high school graduates. Local real wage gap measures the wages net of local rents gap. Well-being gap is measured by the difference in a college and high school graduate's willingness to pay to live in his first choice city from the choices availble in 2000 versus his first choice in 1980. For example, the well-being gap due to wage changes only accounts for the welfare impact of wage changes from 1980 to 2000, while the well-being due to wages and rents accounts for both the impacts of wages and rents. The well-being gap is normalized to the college wage gap in 1980. Standard errors for welfare estimates use the delta method.

Since I do not observe each worker's idiosyncratic taste for each city, I must integrate out over the error distribution to calculate his expected utility from the city he chooses to live in. Since the error terms are distributed Type I extreme value, a worker's expected utility from his top choice city is:

$$E\left(U_{i1980}\right)=\ln\left(\sum_j \exp\left(\begin{array}{c}\beta^w z_i w^{edu}_{j1980}-\beta^r z_i r_{j1980}+\xi^z_{j1980}+\beta^{\text{col}} z_i \frac{H_{j1980}}{L_{j1980}}\\ +\beta^{st} z_i st_i x^{st}_j+\beta^{\text{div}} z_i \,\text{div}_i\, x^{\text{div}}_j\end{array}\right)\right).$$

Similarly, worker i's expected utility if wages adjust to the levels observed in 2000,$E\left(\hat{U}^w_{i2000}\right)$, is measured by:

$$E\left(\hat{U}^w_{i2000}\right)=\ln\left(\sum_j \exp\left(\begin{array}{c}\beta^w z_i w^{edu}_{j2000}-\beta^r z_i r_{j1980}+\xi^z_{j1980}+\beta^{\text{col}} z_i \frac{H_{j1980}}{L_{j1980}}\\ +\beta^{st} z_i st_i x^{st}_j+\beta^{\text{div}} z_i \,\text{div}_i\, x^{\text{div}}_j\end{array}\right)\right).$$

$E\left(\hat{U}^w_{i2000}\right)$ measures the utility worker i receives from living in the city he finds most desirable. Combining these, the expected utility impact due to cities' wage changes from 1980 to 2000 is:

$$\frac{E\left(\hat{U}^w_{i2000}\right)-E\left(\hat{U}^w_{i1980}\right)}{\beta^w z_i}.$$

The change in utility is divided by worker i's marginal utility of wages, so that utility is measured in log wage units. Intuitively, the expected utility change measures each workers willingness to pay (in log wages) to live in his first choice counterfactual city instead of his first choice city from the set available in 1980. I compute this for each worker in 2000 and compare the average utility impact for college workers to the that of non-college workers.

Table 1.9 reports these estimates. From 1980-2000 changes in cities' wages led to an increase in the college well-being gap equivalent to a nationwide increase of 0.193 log points in the college wage gap, which is very close to observed increase of 0.199 in the college wage gap. Even if local amenities and rents had not changed, there still would have been a substantial increase in well-being inequality between college and non-college workers due to local wage changes.

To account for the additional effect of local rent changes, I perform a similar calculation that allows local wages and rents to adjust to the level observed in 2000. The expected utility of worker i if wages and rent adjust to the level observed in 2000 $E\left(\hat{U}_{i2000}^{wr}\right)$ is:

$$E\left(\hat{U}_{i2000}^{wr}\right) = \ln\left(\sum_j \exp\left(\begin{array}{c} \beta^w z_i w_{j2000}^{edu} - \beta^r z_i r_{j2000} + \xi_{j1980}^z + \beta^{\text{col}} z_i \frac{H_{j1980}}{L_{j1980}} \\ +\beta^{st} z_i st_i x_j^{st} + \beta^{\text{div}} z_i \,\text{div}_i \, x_j^{\text{div}} \end{array}\right)\right).$$

The change in well-being inequality between college and and high school graduates due to wage and rent changes from 1980 to 2000 is equivalent to a nationwide increase of 0.175 log points in the college wage gap. The welfare impacts of wages and rents lead to a smaller increase in well-being inequality because the cities which offered the most desirable wages for college workers also had the highest rents, offsetting some of the wage benefits.

To measure the additional contribution of amenity changes to well-being inequality, I can only quantify the welfare impacts of endogenous amenity changes due to changes in cities' college employment ratios. Since the model infers unobserved exogenous amenity changes by measuring which cities have larger population growth than would be expected from the local wage and rent changes, the model only identifies *relative* amenity changes between cities across years.

The model cannot identify the overall magnitude of amenity changes across decades. To see this consider a simple example of 2 cities: New York and Chicago. Assume New York and Chicago are equally appealing in year 1, and have equal populations. In year 2, there is large migration from New York to Chicago, which cannot be explained by wage and rent changes. One can conclude that the amenities of Chicago must have improved, *relative* to the amenities of New York. If the amenities of New York stayed fixed, while the amenities of Chicago improved, workers were able increase their utility, since New York is equally desirable in years 1 and 2, but Chicago improved. In contrast, if the amenities of New York declined, but Chicago's amenities stayed fixed, workers would be worse off in year 2 than year 1, because both cities are weakly less desirable in year 2 than year 1. Yet these two scenarios produced *identical* migration patterns, which makes inferring the welfare effects of unobserved amenity changes over time impossible.

The welfare effects of endogenous amenity changes over time, however, can be measured. Since a city's college employment ratio represents an index of the city's endogenous amenity level, an increase in a city's college employment ratio over time means that the endogenous amenities must have improved from one year to the next. This welfare effect can be measured directly.

There are two main reasons the endogenous amenities of cities have changed over time. First, there has been a nationwide increase in the share of the population with a college degree. This led to increases in the college shares of almost all cities from 1980 to 2000. Second, there has been a re-sorting of college and non-college workers across cities, which, coupled with the nationwide college share increase, led to increases in some cities' college shares more than others.

First, I measure the impact of amenity changes on well-being inequality driven only by the re-sorting of college and non-college workers across cities. I hold the nationwide college share fixed at the 1980 level. The expected utility of worker i if wages, rents, and endogenous amenities due to resorting adjust to the level observed in 2000 $E\left(\hat{U}_{i2000}^{wr}\right)$, is measured by:

$$
\begin{aligned}
E\left(\hat{U}_{i2000}^{wr}\right) &= \ln\left(\sum_j \exp\left(\begin{array}{c} \beta^w z_i w_{j2000}^{edu} - \beta^r z_i r_{j2000} + \xi_{j1980}^z + \beta^{\text{col}} z_i \frac{\hat{H}_{j2000}}{\hat{L}_{j2000}} \\ +\beta^{st} z_i st_i x_j^{st} + \beta^{\text{div}} z_i \,\text{div}_i\, x_j^{\text{div}} \end{array}\right)\right) \\
\hat{H}_{j2000} &= \frac{H_{j2000}}{H_{2000}} H_{1980}, \hat{L}_{j2000} = \frac{L_{j2000}}{L_{2000}} L_{1980}.
\end{aligned}
$$

$\hat{H}_{j2000}$ measures the share of all high skill workers living in city j in year 2000, scaled by the national population size of high skill workers in 1980:H_{1980}. $\hat{L}_{j2000}$ is similarly defined, for low skill workers.

The change in well-being inequality between college and high school graduates due to wage, rent, and endogenous amenity changes due to workers re-sorting across cities from 1980 to 2000 is equivalent to a nationwide increase of 0.238 log points in the college wage gap. This change in well-being inequality is 20% *larger* than the observed increase in the actual college wage gap from 1980 to 2000. A Wald test of whether this increase in well-being inequality is statistically indistinguishable from the well-being increases due only to wage changes is rejected.

The additional nationwide growth in the country's college share led to large amenity changes across almost all US cities. Adding on the additional effect of the change in endogenous amenities due to the nationwide increase in all cities' college shares leads to an overall increase in well-being inequality equivalent to 0.461 log point increase in the college wage gap. This figure, however, should be interpreted with caution. There are surely many other nationwide changes in the US which differentially effected the well-being of college and non-college workers. For example, nationwide improvements in health care, life expectancy, air-conditioning, television, and the internet likely influenced the well-being of all workers nationwide. Since the model can only capture the welfare effects of college share changes and not the many other nationwide change, one should not interpret the welfare effects of the nationwide increase in college graduates as an accurate measure of changes in overall well-being inequality. It is difficult to gauge what aspects of well-being inequality changes are measured in the nationwide increase in cities' endogenous amenities.

For these reasons, I place more confidence in the estimated changes in well-being inequality due to wage, rent, and endogenous amenity changes driven by workers re-sorting across cities. The combined welfare effects of changes in wages, rents, and endogenous amenities driven only by the re-sorting of workers across locations have led to at least a 20% larger increase in well-being inequality than is apparent in the changes in the college wage gap alone.

1.9 Conclusion

The divergence in the location choices of high and low skill workers from 1980 to 2000 was fundamentally caused by a divergence in high and low skill productivity across space. By estimating a structural spatial equilibrium model of local labor demand, housing supply, and labor supply to cities, I quantify the ways through which local productivity changes led to a re-sorting of workers across cities. The estimates show that cities which became disproportionately productive for high skill workers attracted a larger share of skilled workers. The rise in these cities' college shares caused increases in local productivity, boosting all workers' wages, and improved the local amenities. The combination of desirable wage and amenity growth caused large amounts of in-migration, driving up local rents. However, low skill workers were less willing to pay the "price" of a lower real wage to live in high amenity cities, leading them to prefer more affordable, low amenity locations.

The net welfare impacts of the changes in cities' wages, rents, and endogenous amenities led to an increase in well-being inequality between college and high school graduates of at least 20% more than the increase in the college wage gap alone. The additional utility college workers gained from of being able to enjoy more desirable amenities, despite the high housing local prices, increased college workers' well-being, relative to high school graduates.

While the model quantifies high and low skill productivity levels across cities, it does not fully analyze the underlying determinants of these productivity differences across locations. A point of future work is to study the causes of the geographic decoupling of high skill productivity from low skill productivity and the resulting re-sorting of firms across US cities.

Chapter 2: Housing Supply Elasticity and Rent Extraction by State and Local Governments

2.1 Introduction

Can government workers extract rent from private sector workers by charging high tax rates and paying themselves high wages? The determinants and justification of government workers' compensation levels has taken on considerable heat in the past few years, as many states and localities face budgetary stress. Since state and local governments set taxes and government employee wages, government employees could earn rents by charging high taxes and receiving high wages. There has long been debate over whether the government acts as a benevolent social planner for its citizens or uses its market power to benefit its workers and political interest groups. (See Gregory and Borland (1999) for a review of this literature.) In particular, the high unionization rate in the public sector may allow union bargaining to influence the political process and the decisions of elected officials (Freeman (1986)). In this paper, I analyze whether government workers receive higher wages than similar private sector workers in areas where state and local governments have stronger abilities to exercise market power.

This paper develops a model where state and local governments set taxes and the level of government services to maximize government "profits", which can then be paid to employees as excessive wages. I use a Rosen (1979) Roback (1982) spatial equilibrium model where workers maximize their utility by living in the city which offers them the most utility based on the city's wage, rental rate of housing, tax rate, government services, and other amenities. Thus, governments must compete for residents to tax, and workers can "vote with their feet" by migrating away from excessively rent extractive governments.

I show that if state and local governments are using their market power to over pay their employees, their abilities to extract rents from their citizens is determined by the equilibrium migration elasticity of private sector residents with respect to local tax rates. Governments must trade off the benefits of a higher tax with the cost that a higher tax will cause workers to migrate away, leaving the government with a smaller population to tax. This is analogous to the standard result found in analysis of imperfect competition between product producers where a firm's optimal price markup over cost is equal to the inverse elasticity of consumer demand with respect to price for the firm's product.

Unlike firm competition for consumer demand, I show that a government's market power to charge wasteful taxes remains even when there are a large number of governments competing for residents and every government is small.[44] The spatial equilibrium model shows that when a government raises taxes, workers will migrate away to other jurisdictions. However, this out-migration decreases the level of labor supply and housing demand in the area. Assuming labor demand curves slope down and housing supply curves slope up, this decrease in population raises wages and decreases housing rents. Thus, some of the disutility of a tax increase will be offset by an increase in the desirability of local wages and rents, which limits the amount of out-migration caused by the tax increase. Since the local housing and labor markets will respond to government imposed taxes through migration, the government will always have market power.

An area's elasticity of housing supply will determine how local housing rents respond to population changes in an area. Governments presiding over areas with inelastic housing supplies will have more market power than governments in housing elastic areas. A tax hike by a government in an area with inelastic housing supply leads to a small amount of out-migration because housing prices sharply fall due to the decrease in housing demand driven by the tax hike. The housing cost decline offsets the negative utility impact of a tax increase with a only small amount of out-migration in the housing inelastic area. Thus, governments in housing inelastic areas can charge higher taxes without shrinking their tax base since housing price changes limit the migration response.

[44] This result is closely related to Epple and Zelenitz (1981), which shows that worker migration between government juristidictions is not enough to entirely compete away a government's market power.

If state and local governments exercise more market power in areas with inelastic housing supplies, the wage gap between public and private sectors workers should be larger in these areas. I test the model's prediction by measuring variation in public-private sector wage gaps across areas with different housing supply elasticities. I measure workers' wages using data from the 1995-2011 Current Population Survey Merged Outgoing Rotation Groups (CPS-MORG). I proxy for a metropolitan areas's housing supply elasticity using data from Saiz (2010) on the share of land within 50km of a city's center unavailable for real-estate development due to geographic constraints, such as the presence of swamps, steep grades, or bodies of water. With less available land around to build on, the city must expand farther away from the central business area to accommodate a given amount of population, driving up average housing costs.[45] I also use the Wharton Land Use Regulation Index from Gyourko, Saiz, and Summers (2008) as component of housing supply elasticity. Since the decision to regulate real-estate development is endogenous and possibly correlated with unobserved characteristics which could impact government workers wages, I focus on the Saiz (2010) measure of geographic constraints on real-estate development as an exogenous source of variation in housing supply elasticity. These data are the metropolitan area level. To measure states' housing supply elasticities I use an average of these measures across each state's MSAs, weighted by the MSAs' populations.

I find that the public-private sector wage gap is higher in states and metropolitan areas with less elastic housing supplies. This result holds when analyzing variation in state government-private sector wage gaps across states and in local government-private sector wage gaps across MSAs. This finding is robust to including a host of controls for workers' demographics and characteristics, including dummies for three digit occupation codes. Ad-

[45] A full micro-foundation of this mechanism can be derived from the Alonso-Muth-Mills model (Brueckner (1987)) where housing expands around a city's central business district and workers must commute from their house to the city center to work. Within-city house prices are set such that workers are indifferent between having a shorter versus longer commute to work. Average housing prices rise as the population grows since the houses on the edge of the city must offer the same utility as the houses closer in. As the city population expands, the edge of the city becomes farther away from the center, making the commuting costs of workers living on the edge higher than those in a smaller city. Since the edge of the city must offer the same utility value as the center of the city, housing prices rise in the interior parts of the city.

ditionally, the local government-private sector wage gap is found to be higher in housing inelastic MSAs, even when only comparing MSAs within the same state.

As falsification tests, I show that housing supply elasticity has no impact on the federal government worker-private sector wage gap. Since federal workers' compensation is not derived from government revenues of their place of residence, the market power of the state and local government should have no impact on their wages. Additionally, I show that variation in the state government worker-private sector wage gap does not vary across MSAs, within a state. The public-private wage gaps only vary with housing supply elasticities when the housing supply elasticity variation impacts the government's market power. I also show that the effect is larger for government workers who are union members, suggesting unions allow government workers to bargain for a larger share of government rents.

The CPS-MORG only reports data on workers' earnings, and does not include data on the value of workers' benefts. Gittleman and Pierce (2012) show that goverment employees receive more generous benefits than similar private sector workers, on average. I use data from on average government pension payouts per beneficiary across states from the Census' 2007-2010 Annual Surveys of Public Employee Retirement Systems as a measure of state government workers' benefits. While I do not have a data source for similar private sector workers' retirement benefits, I show that average annual state government pension payouts per beneficiary are higher in states with less elastic housing supplies. This suggests that the wage gap estimates from the CPS understate the full impact of housing supply elasticity on government worker compensation.

Previous work has also found evidence suggesting government jobs are more desirable than similar private sector jobs. Gittleman and Pierce (2012) show that public sector employees are more generously compensated than similarly qualified private sector employees. In particular, they find that government worker wages tend to be slightly lower than similar private sector workers. However, the value of government workers' benefits strongly outweigh those of the private sector, leading to public sector employees to be better compensated overall. Krueger (1988) finds that there are more job applications for each government job than for each private sector job, suggesting that government jobs are more desirable to workers, on average. Additionally, average job quit rates reported from the 2002-2006 Job Openings and Labor Turnover Surveys show that average annual quit rate is 28% for private sector workers, but only 8% for public sector employees. These fact taken together suggest that government jobs are better compensated than private sector jobs, and that there appears to be excess labor supply for these jobs, which is consistent with government workers receiving rents. While this evidence shows that government jobs appear desirable to workers, it is not clear that this desirability is due to rent-seeking behavior of governments exercising market power. My paper shows that an increase in governments' abilities to extract rent directly leads to better paid government employees.

The public sector workforce is also highly unionized, enabling government employees to bargain for government rents. Gyourko and Tracy (1991) use a spatial equilibrium model to show that if the cost of government taxes to citizens are not completely offset by benefits of government services, they will be capitalized into housing prices. Similarly, if high levels of public sector unionization lead to more government rent extraction, the public sector unionization rate will proxy for government waste and also be capitalized into housing prices. While Gyourko and Tracy (1991) find evidence for both of these effects, it is unclear what drives the variation in taxes and unionization rates across localities. This paper uses housing supply elasticity as a source of exogenous variation in government market power to assess whether government take advantage of their power to over pay employees.

Brueckner and Neumark (2011) considers whether government can extract more rent from local residents if the government presides over an area with more desirable amenities. They use a similar setup to this paper where profit maximizing governments compete for residents by setting local tax rates. They allow local governments to play a game in tax-competition where the number of competing governments is small. My model differs from theirs by allowing each government to be small when deriving tax rates chosen by governments. They find evidence that amenity differences are positively associated with public-private wage gaps. However, it is possible that some of the amenity measures, such as coastal proximity and population density, are correlated with housing supply elasticity differences.

The paper proceeds as follows. Section 2 layouts of the model. Section 3 presents empirical evidence, and Section 4 concludes.

2.2 Model

The model detailed below uses a Rosen (1979) Roback (1982) spatial equilibrium to analyzes how local governments set taxes and compete for residents. In the model, I assume that governments use a head tax to collect revenue, however in reality, most state and local governments use property and income tax instruments. In Appendix B I derive results for the case of a government income or property tax and show the same results. I also abstract away from the political election process in each area. While politics could surely influence the extent of government rent seeking, my goal is to analyze contributors to governments' abilities to excercise market power if they had a rent seeking motivation.

The nationwide economy is made up of many cities. There are N cities, where N is large. Cities are differentiated by their endowed amenity levels A_j, which impact how desirable workers find the city, and their endowed productivity levels θ_j, which impact how productive firms are in the city. Workers are free to migrate to any city within the country. Each city has a local labor and housing market, which determine local wages and rents. The local government provides government services and collects taxes.

2.2.1 Government

The local government of city j charges a head tax τ_j to workers who choose to reside within the city. The local government also produces government services, which cost s_j for each worker in the city. The government revenue and cost are

$$\begin{aligned} \text{Revenue}_j &= \tau_j N_j \\ \text{Cost}_j &= s_j N_j. \end{aligned}$$

N_j measure the population of city j. The local government is not benevolent and maximizes profits. These profits could be spent on inefficient production of s_j (thus, making the government benevolent, but naive). They could also be directly pocketed by government workers, such as through union negotiations. The local government maximizes:

$$\max_{\tau_j, s_j} \tau_j N_j - s_j N_j$$

2.2.2 Workers

All workers are homogeneous. Workers living in city j inelastically supply one unit of labor, and earn wage w_j. Each worker must rent a house to live in the city at rental rate r_j and pay the local tax τ_j. Workers value the local amenities as measure by A_j. The desirability of government services s_j is represented by $g(s_j)$. Thus, workers' utility from living in city j is:

$$U_j = w_j - r_j + A_j + g(s_j) - \tau_j.$$

Workers maximize their utility by living in the city which they find the most desirable.

2.2.3 Firms

All firms are homogenous and produce a tradeable output Y. Cities exogenously differ in their productivity as measured by θ_j. Local government services impact firms productivity, as measured by $b(s_j)$. The production function is:

$$Y_j = \theta_j N_j + b(s_j) N_j + F(N_j),$$

where $F'(N_j) > 0$ and $F''(N_j) < 0$ in labor.

The labor market is perfectly competitive, so wages equal the marginal product of labor:

$$w_j = \theta_j + b(s_j) + F'(N_j).$$

2.2.4 Housing

Housing is produced using construction materials and land. All houses are identical. Houses are sold at the marginal cost of production to absentee landlords, who rents housing to the residents. The asset market is in long-run steady state equilibrium, making housing price equal the present discounted value of rents. Housing supply elasticities differ across cities. Differences in housing supply elasticity are due to topography and land use regulation, which makes the marginal cost of building an additional house more responsive to population changes (Saiz (2010)). The housing supply curve is:

$$\begin{aligned} r_j &= a_j + \gamma_j \log(N_j), \\ \gamma_j &= \gamma x_j^{\text{house}} \end{aligned}$$

where x_j^{house} is a vector of city characteristics which impact the elasticity of housing supply.

2.2.5 Equilibrium in Labor and Housing

Since all workers are identical, all cities with positive population must offer equal utility to workers. In equilibrium, all workers must be indifferent between all cities. Thus:

$$U_j = w_j - r_j + A_j + g(s_j) - \tau_j = \bar{U}.$$

Plugging in labor demand and housing supply gives:

$$\theta_j + b(s_j) + F'(N_j) - a_j - \gamma_j \log N_j + A_j + g(s_j) - \tau_j = \bar{U}. \tag{2.1}$$

Equation (2.1) determines the equilibrium distribution of workers across cities.

2.2.6 Government Tax Competition

Local governments set city tax rates and the level of government services to maximize profits, taking into account the endogenous response of workers and firms in equilibrium, equation (2.1). Each city is assumed to be small, meaning out-migration of workers to other cities does not impact other cities' equilibrium wages and rents. If there were a small number of cities, each city would have even more market power than in this limiting case. The results below can be thought of as a lower bound on the market power of local governments competing for residents. They maximize:

$$\max_{s_j,\tau_j} \tau_j N_j - s_j N_j.$$

The first order conditions are:

$$\begin{aligned} 0 &= \tau_j \frac{\partial N_j}{\partial s_j} - N_j - s_j \frac{\partial N_j}{\partial s_j} \\ 0 &= \tau_j \frac{\partial N_j}{\partial \tau_j} + N_j - s_j \frac{\partial N_j}{\partial \tau_j}. \end{aligned} \tag{2.2}$$

Differentiating equation (2.1) to solve for $\frac{\partial N_j}{\partial s_j}$ and $\frac{\partial N_j}{\partial \tau_j}$ gives:

$$\begin{aligned} \frac{\partial N_j}{\partial s_j} &= \frac{b'(s_j) + g'(s_j)}{\left(\frac{\gamma_j}{N_j} - F''(N_j)\right)} > 0 \\ \frac{\partial N_j}{\partial \tau_j} &= \frac{-1}{\left(\frac{\gamma_j}{N_j} - F''(N_j)\right)} < 0. \end{aligned} \tag{2.3}$$

Population increases with government services and decreases in taxes. Plugging these into (2.2) gives:

$$\begin{aligned}
0 &= (\tau_j - s_j)\left(\frac{b'(s_j) + g'(s_j)}{\left(\frac{\gamma_j}{N_j} - F''(N_j)\right)}\right) - N_j \\
\tau_j &= N_j\left(\frac{\gamma_j}{N_j} - F''(N_j)\right) + s_j.
\end{aligned}$$

Combining the first order conditions shows that government services are provided such that the marginal benefit $(b'(s_j) + g'(s_j))$ equals marginal cost (1) :

$$b'\left(s_j^*\right) + g'\left(s_j^*\right) = 1.$$

This is the socially optimal level of government service.

The equilibrium tax rate is:

$$\tau_j^* = \gamma_j - N_j F''(N_j) + s_j^*. \tag{2.4}$$

The elasticity of city population with respect to the tax rate $\left(\varepsilon_j^{migrate}\right)$ can be written as:

$$\varepsilon_j^{migrate} = \frac{\partial N_j}{\partial \tau_j}\frac{\tau_j}{N_j}.$$

Plugging in equation (2.3) for $\frac{\partial N_j}{\partial \tau_j}$ and rearranging gives:

$$\left(\frac{\gamma_j}{N_j} - F'(N_j)\right) N_j = \frac{-\tau_j}{\varepsilon_j^{migrate}}.$$

Substituting this expression into the equation (2.4) shows that the tax markup can be written as:

$$\frac{\tau_j^* - s_j^*}{\tau_j^*} = \frac{-1}{\varepsilon_j^{migrate}}.$$

The tax markup above cost is equal to the inverse elasticity of city population with respect to the tax rate. While workers are perfectly mobile between cities, worker migration causes shifts along the local labor demand and housing supply curves. An increase in local taxes would cause workers to migrate to other cities. A decrease in population will increase local wages, since I have assumed a downward sloping labor demand curve. The decrease in population will also cause rents to fall, by moving along the housing supply curve. This increase in wages and decrease in rents will increase the desirability of the city to workers, limiting the migration response to the tax increase. The government takes into account the equilibrium wage and rent response to a tax hike when setting taxes to profit maximize. Thus, if migration leads to large changes in local wages and rent, a tax increase will not lead to large amounts of out-migration, since workers will be compensated for the tax with more desirable wages and rents.

To analyze the effect of housing supply elasticity on governments' ability to extract rent from taxes, I differentiate the tax markup with respect to the slope of the inverse housing supply curve, γ_j.

$$\frac{\partial}{\partial \gamma_j}\left(\tau_j^* - s_j^*\right) = 1 - \frac{\partial}{\partial N_j}\left(N_j F''\left(N_j\right)\right)\frac{\partial N_j}{\partial \gamma_j}. \tag{2.5}$$

The first term (1) in equation (2.5) represents the increased rent response to migration induced by a tax hike in a city with an inelastic housing supply. The equilibrium condition, equation (2.1), shows that out-migration will continue until the negative utility impact of the tak hike has been completely offset by changes in the city's wage and rent. In a city with a less elastic housing supply, a smaller amount of migration is needed to push housing rents down to offset the negative utility impact of the tax hike. The second term in equation (2.5) represents the change in the elasticity of labor demand due to being at a different point on the labor demand curve $\left(\frac{\partial}{\partial N_j}\left(N_j F''\left(N_j\right)\right)\right)$. Since a city with a less elastic housing supply has a smaller equilibrium population, the slope of the labor demand curve in a smaller city could differ from the slope of the labor demand curve in a larger city. I will assume

$\frac{\partial}{\partial N_j}\left(N_j F''\left(N_j\right)\right) = 0$, which is equivalent to assuming $\exp\left(F'\left(N_j\right)\right)$ has a constant elasticity with respect to N_j. Under this assumption, the derivative of the tax markup with respect to the slope of the inverse housing curve is:

$$\frac{\partial}{\partial \gamma_j}\left(\tau_j^* - s_j^*\right) = 1 > 0.$$

The government can extract more rent through higher taxes in a city with a less elastic housing supply.

Note that this result assumes there are a large number of cities. When there are a small number of cities, the incentives for rent extraction will be even higher. Outward migration from a city in response to a tax increase will lead to increases in other cities' rents and decreases in their wages, leading to less outward migration in response to tax increases. I have assumed this effect away by not allowing the equilibrium utility level across all cities to fall in response to a given city's tax increase. Cities can extract rent even in an environment where there are a large number of competitors because household demand for city residence can never be infinite in equilibrium.

Additionally, this model assumes cities charge a head tax, while in reality most cities and states tax their population through income taxes and property taxes. The amount of rent extraction depends on the elasticity of tax revenue with respect to the tax rate. Thus, an income tax will depend both on the wage response to the tax rate, as well as the migration response. Appendix B shows that when using an income tax, governments can still excercise more market power in housing inelastic areas.

In the case of a property tax, government revenue will depend on local the rental rate and the size of the tax base. An increase in the property tax rate can decrease government revenue both by incentivizing workers to migrate away, shrinking the tax base, and decreasing housing rents, lowering tax revenue from each household. However, I show in appendix B that if local labor demand is perfectly elastic, the housing supply elasticity will not impact the size of the rental rate decrease in response to a given tax hike. To see this, recall the equilibrium condition, equation (2.1). For workers to derive utility $\bar{U}$ from a local area, the utility impact of a tax increase must be perfectly offset by a rent decrease, if labor demand is perfectly elastic. Thus, the equilibrium rental rate response to a given tax increase does not depend on the local housing supply elasticity. Indeed, the housing supply elasticity determines the migration response required to change housing rents in order to offset the utility impact of the tax increase. Thus, a less elastic housing supply decreases the elasticity of government revenue with respect to the tax rate, giving the government more market power when using a property tax instrument. See Appendix B for the full derivation of this result.

Regardless of the tax instrument, governments of cities with less elastic housing supplies are able to extract more rent from their residents. In the next section, I empirically test this prediction.

2.3 Empirical Evidence

The model predicts that local governments in areas with less elastic housing supplies will be able to extract more rent from their residents. While this extra money could be spent in a number of ways, it is likely that some of it gets absorbed into public sector workers' wages. Since most public sector workers are unionized, they will be able to bargain to gain some of these rents as wages. Thus, the wage gap between government employees and similarly qualified private sector workers should be higher in areas with less elastic housing supplies. The effect should hold across metropolitan areas for local government-private sector wage gaps and across states for state government-private sector wage gaps.

To test this, I estimate how states' and MSAs' public/private sector wage gaps vary with characteristics which impact local housing supply elasticities. Saiz (2010) shows that the topological characteristics of land around an MSA's center impact whether the land can used for real-estate development. Cities located next to wetlands, bodies of waters, swamps, or extreme hilliness have limits on how many building can be built close to the city center, which impacts the elasticity of housing supply to the area. Saiz (2010) uses satellite data to measure the share of land within 50km of an MSA's center which cannot be developed due to these topological constraints. A rent-seeking government is able to charge higher taxes in areas with less land available for development. Thus, the public-private sector wage gap should be higher in these areas.

A city's housing supply elasticity is also influenced by the amount of land-use regulation in the area. The 2005 Wharton Land Use Regulation Survey Gyourko, Saiz, and Summers (2008) collected survey data on a number of land-use regulations and practices, which were aggregated into the Wharton Land Use Regulation Index (WLURI). Saiz (2010) aggregates this municipality measure into a MSA-level index.

I z-score the MSA level data from the WLURI and the land unavailability measure and will use both as measures of cities' housing supply elasticities. I also aggregate these measures to a state-level index, where I weight each MSA measure by the state population in each MSA. The state-level housing supply elasticity measure is a noisy measure of the overall housing supply elasticity for the state, since the data is only based off of the MSAs covered by Saiz's sample. Table 2.1 reports summary statistics on these measures. The data covers 48 states (there is no data for Hawaii or Alaska) and 228 MSAs.

Table 2.1: Summary Statistics

A. CPS Data 1995-2011					
	Mean	Standard Dev.	Min	Max	N
Local Government Worker Ln Weekly Earnings	6.723	0.519	4.854	8.695	112639
State Government Worker Ln Weekly Earnings	6.725	0.509	4.857	8.695	62994
Federal Government Worker Ln Weekly Earnings	6.964	0.496	4.855	8.695	37008
Private Sector Worker Ln Weekly Earnings	6.686	0.613	4.852	8.695	968349

B. Housing Supply Elasticity Measures					
	Mean	Standard Dev.	Min	Max	N
State Aggregated Land Unavailability: Z-Score	0.000	1.000	-1.427	2.982	48
State Aggregated Wharton Land Use Regulation Index: Z-Score	0.000	1.000	-1.640	2.348	48
MSA Land Unavailability: Z-Score	0.000	1.000	-1.205	2.824	228
MSA Wharton Land Use Regulation Index: Z-Score	0.000	1.000	-1.746	3.938	228

C. State Government Pension Payouts: 2007-2010					
	Mean	Standard Dev.	Min	Max	N
Log Thousand Dollars of Annual Payout per beneficiary from State Pension	2.9897	0.24996	2.5289	3.629	192

Notes: Wages are measured as weekly wages deflated by the CPI-U and reported in constant 2011 dollars for 25-55 year old workers working at least 35 hours per week. Workers with imputed weekly earnings are dropped from the analysis. Top coded weekly earnings as set to 1.5 times the top coded value and weekly earnings below $128 (in real 2011 dollars) are dropped from the analysis. Sector of worker (local/state/federal/private) is measured by reported class of worker. MSA land unavailability measures the share of land within 50km of an MSA's center which cannot be developed due to these topological constraints from Saiz (2010). This measure is then Z-scored. The Wharton Land Use Regulation index aggregates survey data on a large set of local land use practices by local municipalities. This is aggregated to the MSA level and then Z-scored. State aggregated housing supply elasticity meaures use a population weighted average of MSA level data. State and MSA level housing supply elasticity measures are z-scored. Government pension data come from 2007-2010 Annual Surveys of Public-Employee Retirement Systems.

2.3.1 Wage Gap Regressions

To measure public-private sector wage gaps across MSAs and states, I use data from the Current Population Survey Merged Outing Rotation groups from 1995-2011.[46] The CPS-MORG is a household survey which collects data on a large number of outcomes including workers' weekly earnings, hours worked, public/private sector of employment, union status, and a host of demographics. I restrict the sample to 25 to 55 year old workers with positive labor income, working at least 35 hours per week, to have a standardized measure of weekly earnings. The CPS's usual weekly earnings question does not include the self-employment income so all analysis excludes the self-employed. I also restrict analysis to workers whose wages are not imputed to avoid any bias due to the CPS's wage imputation algorithm (Bollinger and Hirsch (2006)). I measure earnings using workers' log usual weekly earnings, deflated by the CPI-U and measured in real 2011 dollars. Top coded weekly earnings are multiplied by 1.5 and weekly earnings below $128 are dropped from the analysis.[47] All analysis is weighted by the CPS earnings weights.

Table 2.1 reports summary statistics of workers' log weekly earnings each for workers employed in the private sector, local government, state government, and federal government.[48] Consistent with previous works, such as Gittleman and Pierce (2012), the raw earnings are higher for all three classes of government workers than for private sector workers. However, these raw earnings differences do not account for differences in the characteristics of workers between the public and private sector. To test the model's predictions, I will control for worker characteristics when evaluating differences in the public private sector wage gap.

[46] Since there was a significant change in the CPS's earnings questions in 1994, I restrict analysis to 1994-2011. I also focus my analysis on workers whose wages are not imputed in the CPS. Since sector, occupation, and union status and not used in the CPS's imputation algorithm, analyzing government wage gaps and union wage gaps using imputed wages can be problematic (Bollinger and Hirsch (2006)). Thus, I focus only on the non-imputed wage sample. The data flagging which wages were imputed are missing in the 1994 data, so I drop this year, leaving me with a 1995-2011 sample.

[47] I follow Autor, Katz, and Kearney (2008)'s top and bottom coding procedures. Autor, Katz, and Kearney (2008) drops all reported hourly wages below $2.80 in real 2000 dollars. This translates to $128 per week in real 2011 dollars, assuming a 35 hour work week. They also scale top coded wages by 1.5.

[48] A worker's sector is measured by the CPS variable reporting a worker's class.

Additionally, the CPS only collects data on workers' earnings, but not compensation paid to workers in the form of benefits. Gittleman and Pierce (2012) show using the BLS' restricted-use Employer Cost of Employee Compensation microdata that government employees receive significantly more generous benefits than similar workers in the private sector. I will return to the question of benefits compensation, but first focus on public-private sector wage gaps.

To test the model's predictions, I estimate the following regression:

$$\ln w_{ijt} = \delta_j + \alpha_t + \beta^{gov} gov_{it} + \beta^{elast} z_j^{elast} * gov_{it} + \beta X_{it} + \varepsilon_{ijt}. \tag{2.6}$$

As controls, I include location fixed-effects δ_j, year fixed effects, α_t, and a set of worker demographics which include 15 dummies for education categories, gender, race, Hispanic origin, a quartic in age, and a rural dummy. gov_i is a dummy for whether the worker is government worker, z_j^{elast} measures land use regulation and topography. Standard errors are clustered by state when using state-level measures of housing supply elasticity and clustered by MSA when using MSA variation in housing supply elasticity.

The nationwide average public-private wage gap is measured by β^{gov}. The model predicts that public-private wage gap should be higher in areas with less elastic housing supplies:

$$\beta^{elast} > 0.$$

The prediction should hold both for areas with less land available for development and for areas with stricter land-use regulations. Since land-use regulations are chosen by local municipalities, it is possible that the decision to regulate land-use could be correlated with other characteristics of the area which could impact workers' wages. Since the topological constraints around a city are pre-determined, they are likely a measure of exogenous differences in housing supply elasticities across areas. I perform all analysis using both measures, but I also drop the regulation index to focus directly on the impact of land availability, which is likely a cleaner estimate of the impact of housing supply elasticity on public-private sector wage gaps.

I test this prediction first using a sample including private sector workers and state government workers. Column 1 of Table 2.2 shows that the nationwide average wage gap between state government employees and private sector workers is -0.112 log points. Consistent with Gittleman and Pierce (2012), after controlling for worker demographics, government workers' earnings are lower than similar private sector workers, on average. However, the state worker-private sector wage gap increases by 0.017 log points in states with a 1 standard deviation increase in land unavailability. This effect is significant at the 10% level. The wage gap is 0.026 log points higher in states with a 1 standard deviation increase land-use regulation. This effect is significant at the 5% level. In a regression which drops the land-use index, a 1 standard deviation increase in a state's land availability increases the wage gap by 0.026 log points. This effect is significant at the 1% level. Figure 2.1 plots states' land unavailability against states' state government worker-private sector wage gaps, after wages have been residualized against the set of controls included in equation (2.6) . Figure 1 shows the state government-private sector wage gaps are higher in states including California, Vermont, Florida, and Connecticut, but must lower in states such as Iowa, Texas, Montana, and Kentucky which lines up with these states' land unavailability. Note that states such as Utah are significant outliers. However, Utah's land availability was measured only based on Salt Lake City, which has a large share of land unavailable for development due proximity to the Great Salt Lake. This is likely a poor measure of the overall state housing supply elasticity. Despite the short comings of the state-level data, I find that states with less elastic housing supplies have significantly better paid state government employees, as compared to the private sector employees residing in the state.

Table 2.2: Ln Wage vs. State & Local Public Sector-Housing Supply Elasticity Interactions

	State Gov-Private Sector Wage Gaps		Local Gov-Private Sector Wage Gaps			
	1	2	3	4	5	6
Government Worker	-0.112***	-0.113***	-0.0798***	-0.0707***		
	[0.00909]	[0.00931]	[0.00721]	[0.00758]		
Gov* Land Unavailability	0.0170*	0.0262***	0.0287***	0.0366***	0.00856	0.0108*
	[0.0101]	[0.00976]	[0.00781]	[0.00855]	[0.00637]	[0.00657]
Gov* Land Use Regulation	0.0263**		0.0348***		0.0220**	
	[0.0125]		[0.00803]		[0.00920]	
Constant	2.493***	2.487***	2.382***	2.378***	2.396***	2.399***
	[0.289]	[0.290]	[0.397]	[0.396]	[0.397]	[0.397]
State x Gov Worker FE:					X	X
State Elasticity Measures:	X	X				
MSA Elasticity Measures:			X	X	X	X
State Gov Workers Sample:	X	X				
Local Gov Workers Sample:			X	X	X	X
Observations	973,792	973,792	586,696	586,696	586,696	586,696
R-squared	0.384	0.384	0.389	0.389	0.39	0.39

Note: Standard errors clustered by state for state worker regressions. Standard errors clustered by MSA for local government worker regressions. Weekly wage data from 1995-2011 CPS MORG. Wage data is restricted to 25-55 year old workers working at least 35 hours per week. State government worker sample includes private sector and state government workers. Local government worker sample includes private sector and local government workers living in MSAs. Land unavailability measures a z-score of the share of land within 50km of an MSA's center which cannot be developed due to topological constraints. Land use regulation is an index of aggregated survey data on a large set of local land use practices by local municipalities. This is aggregated to the MSA level and then Z-scored. State level measures are z-scores of average MSA-level measures within the state, weighted by MSA population. Controls include 15 dummies for education categories, gender, race and hispanic origin, a quartic in age, a rural dummy, and year dummies. All regressions weighted using CPS MORG earnings weights.

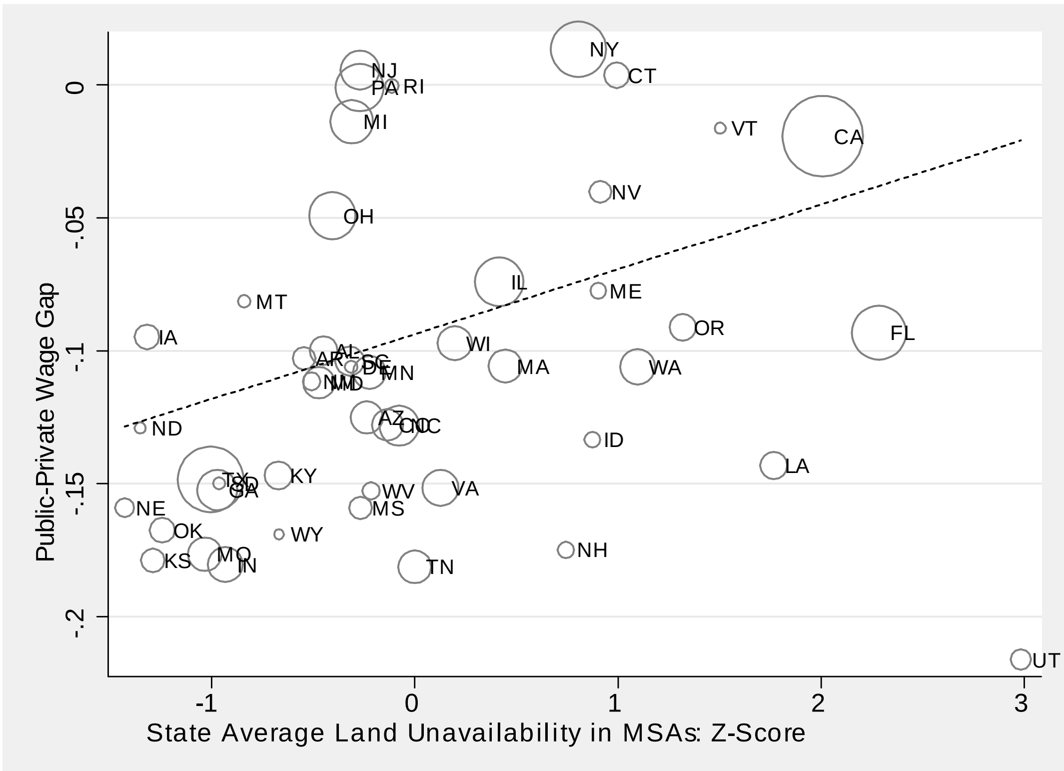


Figure 2.1: State Goverment-Private Sector Wages Gaps vs. State Land Unavailability Z-Score. State wage gaps calculated after residualizing wages against 15 dummies for education categories, gender, race, hispanic origin, a quartic in age, a rural dummy, and year dummies. Regression weighted using CPS earnings weights.

Performing the same analysis on local government employees, I compare the wage gaps between local government workers and private sector workers across 229 MSAs. The controls in this setup now include MSA fixed effects and the housing supply measures are now at the MSA level. Column 3 of Table 2.2 shows that the nationwide local government worker-private sector wage gap is -0.080 log points. A 1 standard deviation increase in land unavailability increases the wage gap by 0.029 log points and a 1 standard deviation increase in land-use regulation increases the wage gap by 0.0348 log points. Both of these effects are significant at the 1% level. Dropping the land-use regulation, I find the coefficient on land unavailability increases to 0.037 log points, and is significant at the 1% level. Figure 2.2 plots MSAs' land unavailability against MSAs' local government worker-private sector wage gaps, after wages have been residualized against the set of controls included in equation (2.6) . The plot shows high local government wages gaps in land unavailable cities including Los Angeles, New York, Cleveland, Chicago, and Portland and low government wage gaps in cities with lots of land to develop including Atlanta, Houston, Minneapolis, and Phoenix. Housing supply elasticity explains a significant amount of the cross-section variation in public-private wage gaps.

To test whether the local housing supply elasticity measures impact local government worker-private sector wage gaps within states, across MSAs, I add controls for state differences in the local government worker-private sector wage gaps. I now estimate:

$$\ln w_{ijkt} = \delta_j + \alpha_t + \beta^{gov} gov_{it} + \beta_k^{gov} gov_{it} + \beta^{elast} z_j^{elast} * gov_{it} + \beta X_{it} + \varepsilon_{ijt},$$

where j represents an MSA and k represents a state. Columns 5 and 6 of Table 2.2 show that the impact of land unavailability on the local government-private sector wage gap falls slightly, but remains statistically siginficant when land-use regulations are not included in the regression. Land unavailability consistently has a positive impact the public-private sector wage gap both for local and state government workers, as predicted by the model.

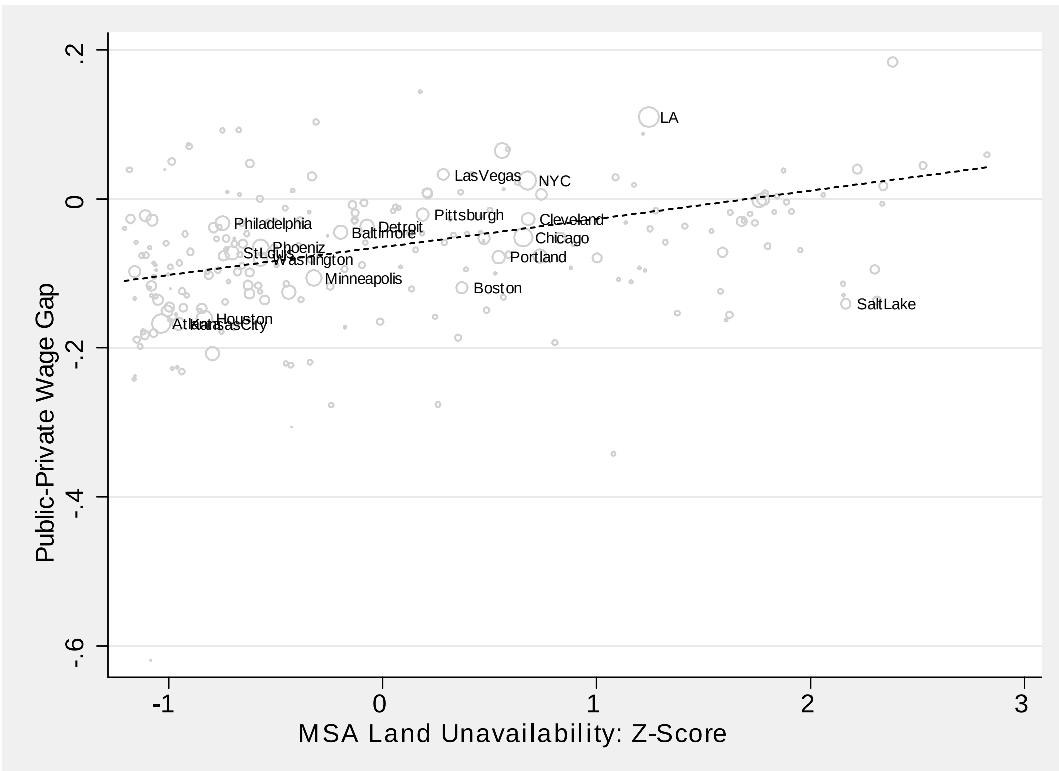


Figure 2.2: Local Goverment-Private Sector Wages Gaps vs. MSA Land Unavailability Z-Score. MSA wage gaps calculated after residualizing wages against 15 dummies for education categories, gender, race, Hispanic origin, a quartic in age, a rural dummy, and year dummies. Regression weighted using CPS earnings weights.

Table 2.3 repeats the analysis adding in dummies for each three-digit occupation code to attempt to further control for differences in workers' skills in the public and private sectors. The point estimate measuring the impact of land unavailability of the state government-private sector wage gap remains positive, but the standard errors increase, making the effect only statistically significant when land-use regulations are dropped from the regression. However, the large standard errors shows that one cannot rule out a point estimate equal to the magnitude found when 3-digit occupation code dummies were not included in the regression.

The estimates for local-government worker wage gaps are positive and statistically significant at the 1% level. Column 4 of Table 2.3 shows that even when including the full set of 3 digit occupation dummies, a 1 standard deviation increase in land unavailability increases the local public-private wage gap by 0.028 log points. Note that with the inclusion of occupation dummies, the nationwide local government worker-private sector wage gap is now positive at 0.028 log points.

One possible way government workers can raise their wages is through union wage bargaining. I repeat the analysis adding in additional housing supply elasticity interactions terms for whether the government worker is part of a labor union. I control for the direct effect of being in a union, and its interaction with the housing supply elasticity characteristics. This controls for differences in union bargaining power across states and MSAs for all unions, public and private. Table 2.4 shows that a standard deviation increase in land unavailability raises state government-private sector wage gaps by 0.0127 log points for non-union members and an addition 0.0214 log points for unionized government workers. State government worker unions appear to be able to bargain for better wages in inelastic areas, relative to non-unionized government workers. The point estimates in Columns 3 and 4 of Table 2.4 are similar for local government-private sector wage gaps. However, the additional impact of goverment labor unions is positive, but not statistically significant. These point estimates suggest government workers part of a labor union might be able to use their market power to negotiate for larger rents beyond the offerings of the private sector.

Table 2.3: Ln Wage vs. State & Local Public Sector-Housing Supply Elasticity Interactions with 3 Digit Occupation Dummy controls

	State Gov-Private Sector Wage Gaps		Local Gov-Private Sector Wage Gaps	
	1	2	3	4
Government Worker	-0.0320***	-0.0331***	0.0181***	0.0275***
	[0.00830]	[0.00837]	[0.00538]	[0.00545]
Gov* Land Unavailability	0.0107	0.0188*	0.0197***	0.0275***
	[0.0111]	[0.0109]	[0.00675]	[0.00743]
Gov* Land Use Regulation	0.0234**		0.0350***	
	[0.0112]		[0.00716]	
Constant	4.263***	4.258***	4.211***	4.208***
	[0.274]	[0.274]	[0.386]	[0.386]
State Elasticity Measures:	X	X		
MSA Elasticity Measures:			X	X
State Gov Workers Sample:	X	X		
Local Gov Workers Sample:			X	X
Observations	973,792	973,792	586,696	586,696
R-squared	0.524	0.524	0.526	0.526

Note: Standard errors clustered by state for state worker regressions. Standard errors clustered by MSA for local government worker regressions. Weekly wage data from 1995-2011 CPS MORG. Wage data is restricted to 25-55 year old workers working at least 35 hours per week. State government worker sample includes private sector and state government workers. Local government work sampe includes private sector and local government workers living in MSAs. Land unavailability measures a z-score of the share of land within 50km of an MSA's center which cannot be developed due to topological constraints. Land use regulation is an index of aggregated survey data on a large set of local land use practices by local municipalities. This is aggregated to the MSA level and then Z-scored. State level measures are z-scores of average MSA-level measures within the state, weighted by MSA population. Controls include 1057 occupation dummies, 15 dummies for education categories, gender, race, hispanic origin, a quartic in age, a rural dummy, and year dummies. Three digit occupation code definitions change in 2000, so I include the full set of combined 3 digit occupation dummies. I treat occupation codes used in years 2000-2011 as distinct occupations from those in previous years, giving a total of 1057 occupation codes. All regressions weighted using CPS earnings weights. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.4: Ln Wage vs. State & Local Public Sector-Housing Supply Elasticity Interactions: Union Government Workers

	State Gov-Private Sector Wage Gaps		Local Gov-Private Sector Wage Gaps	
	1	2	3	4
Government Worker	-0.140***	-0.142***	-0.123***	-0.119***
	[0.00613]	[0.00575]	[0.00695]	[0.00718]
Gov* Land Unavailability	0.0101	0.0127**	0.0269***	0.0316***
	[0.00705]	[0.00624]	[0.00739]	[0.00801]
Gov* Land Use Regulation	0.00782		0.0238***	
	[0.00889]		[0.00801]	
Gov* Land Unavailability*Union	0.00372	0.0214*	-0.00373	0.00753
	[0.0120]	[0.0128]	[0.00845]	[0.00934]
Gov* Land Use Regulation*Union	0.0492**		0.0356***	
	[0.0191]		[0.0106]	
Constant	2.532***	2.540***	2.456***	2.442***
	[0.289]	[0.288]	[0.395]	[0.396]
State Elasticity Measures:	X	X		
MSA Elasticity Measures:			X	X
State Gov Workers Sample:	X	X		
Local Gov Workers Sample:			X	X
	973,792	973,792	586,696	586,696
	0.39	0.389	0.395	0.394

Note:Standard errors clustered by state for state worker regressions. Standard errors clustered by MSA for local government worker regressions. Weekly wage data from 1995-2011 CPS MORG. Wage data is restricted to 25-55 year old workers working at least 35 hours per week. State government worker sample includes private sector and state government workers. Local government work sampe includes private sector and local government workers living in MSAs. Land unavailability measures a z-score of the share of land within 50km of an MSA's center which cannot be developed due to topological constraints. Land use regulation is an index of aggregated survey data on a large set of local land use practices by local municipalities. This is aggregated to the MSA level and then Z-scored. State level measures are z-scores of average MSA-level measures within the state, weighted by MSA population. Controls include 15 dummies for education categories, gender, race and hispanic origin, a quartic in age, a rural dummy, and year dummies. Controls also include a labor union dummy, labor union dummy interacted with government dummy, and the labor union dummy interacted with the housing supply elasticity measures. Union is defined as a member of a labor union. All regressions weighted using CPS earnings weights. *** p<0.01, ** p<0.05, * p<0.1

Table 2.5 repeats the analysis separately for workers with and without a 4 year college education. I find a positive and statistically significant effect both for college and non-college educated workers. The impact of land unavailability is stronger for low skill workers than for those with a college education. Overall, housing supply elasticity appears to impact the public-private wage gap.

2.3.2 Falsification Tests

The evidence presented thus far suggests that governments are exercising their market power by extracting rents and paying government employees higher wages than are paid by local private sector employers. Variation in housing supply elasticities across areas impacts the extent to which governments can exercise market power. A falsification test of these predictions is to analyze whether the federal government-private sector wage gaps across cities and states exhibit similar properties. Since federal workers are not paid by the state or local government which presides over their location of residence, housing supply elasticity should have no impact on federal workers' wages.

Columns 1 through 4 of Table 2.6 estimate the same state and local wage gap regressions, but use federal workers instead of state and local workers. The point estimate of the impact of land unavailability of the federal worker-private sector wage gap is consistently negative. As predicted by the model, the federal worker-private sector wage gaps are not inflated by the housing supply elasticity of these workers' cities or states of residence.

Table 2.5: Ln Wage vs. State & Local Public Sector-Housing Supply Elasticity Interactions: Subsamples by College Education of Workers

	College Sample				Non-College Sample			
	State Gov-Private Sector Wage Gaps		Local Gov-Private Sector Wage Gaps		State Gov-Private Sector Wage Gaps		Local Gov-Private Sector Wage Gaps	
	1	2	3	4	5	6	7	8
Government Worker	-0.182***	-0.182***	-0.123***	-0.119***	-0.0183	-0.0201	-0.0361***	-0.0221**
	[0.00837]	[0.00832]	[0.00852]	[0.00866]	[0.0130]	[0.0143]	[0.00845]	[0.00949]
Gov* Land Unavailability	0.0113*	0.00889*	0.0231***	0.0271***	0.0243*	0.0424***	0.0359***	0.0481***
	[0.00644]	[0.00518]	[0.00880]	[0.00894]	[0.0142]	[0.0148]	[0.00881]	[0.0102]
Gov* Land Use Regulation	-0.00686		0.0169*		0.0520***		0.0563***	
	[0.00981]		[0.00994]		[0.0156]		[0.00959]	
Constant	2.767***	2.773***	2.788***	2.791***	2.472***	2.471***	2.393***	2.379***
	[0.768]	[0.769]	[0.840]	[0.840]	[0.325]	[0.325]	[0.434]	[0.434]
State Elasticity Measures:	X	X			X	X		
MSA Elasticity Measures:			X	X			X	X
State Gov Workers Sample:	X	X			X	X		
Local Gov Workers Sample:			X	X			X	X
Observations	307,951	307,951	201,691	201,691	665,841	665,841	385,005	385,005
R-squared	0.227	0.227	0.229	0.229	0.26	0.26	0.278	0.277

Note: Standard errors clustered by state for state worker regressions. Standard errors clustered by MSA for local government worker regressions. Weekly wage data from 1995-2011 CPS MORG. Wage data is restricted to 25-55 year old workers working at least 35 hours per week. State government worker sample includes private sector and state government workers. Local government worker sample includes private sector and local government workers living in MSAs. Land unavailability measures a z-score of the share of land within 50km of an MSA's center which cannot be developed due to topological constraints. Land use regulation is an index of aggregated survey data on a large set of local land use practices by local municipalities. This is aggregated to the MSA level and then Z-scored. State level measures are z-scores of average MSA-level measures within the state, weighted by MSA population. Controls include 15 dummies for education categories, gender, race and hispanic origin, a quartic in age, a rural dummy, and year dummies. All regressions weighted using CPS MORG earnings weights. *** p<0.01, ** p<0.05, * p<0.1

Table 2.6: State & Federal Government Workers Falsification Tests

	Federal Gov-Private Sector Wage Gaps				State Gov-Private Sector Wage Gaps	
	1	2	3	4	5	6
Government Worker	0.162***	0.161***	0.158***	0.154***		
	[0.0130]	[0.0130]	[0.00701]	[0.00748]		
Gov* Land Unavailability	-0.0172	-0.0216**	-0.00698	-0.00972	-0.0122	-0.0117
	[0.0112]	[0.0103]	[0.00805]	[0.00776]	[0.00786]	[0.00793]
Gov* Land Use Regulation	-0.0151		-0.0158		0.00625	
	[0.0122]		[0.0114]		[0.0100]	
Constant	2.338***	2.337***	2.422***	2.418***	2.701***	2.701***
	[0.291]	[0.291]	[0.413]	[0.412]	[0.413]	[0.413]
State FE	X	X				
MSA FE			X	X	X	X
State x Gov Worker FE:					X	X
State Elasticity Measures:	X	X				
MSA Elasticity Measures:			X	X	X	X
Federal Gov Worker Sample:	X	X	X	X		
State Gov Workers Sample:					X	X
	948,785	948,785	549,857	549,857	560,367	560,367
	0.383	0.383	0.394	0.394	0.39	0.39

Note:Standard errors clustered by state for state worker regressions. Standard errors clustered by MSA for local government worker regressions. Weekly wage data from 1995-2011 CPS MORG. Wage data is restricted to 25-55 year old workers working at least 35 hours per week. State government worker sample includes private sector and state government workers. Federal government work sampe includes private sector and federal government workers. Land unavailability measures a z-score of the share of land within 50km of an MSA's center which cannot be developed due to topological constraints. Land use regulation is an index of aggregated survey data on a large set of local land use practices by local municipalities. This is aggregated to the MSA level and then Z-scored. State level measures are z-scores of average MSA-level measures within the state, weighted by MSA population. Controls include 15 dummies for education categories, gender, race, hispanic origin, a quartic in age, a rural dummy, and year dummies. All regressions weighted using CPS earnings weights. *** p<0.01, ** p<0.05, * p<0.1

As an additional falsification test, I compare the wage gaps between state government and private sector workers across MSAs within states. Since the revenues used to pay state government workers are collected from all areas within a state, the MSA of residence of a state governments should not impact their pay, relative to private sector workers living in the same MSA. I add state fixed effects interacted with whether the worker is employed by the state government as controls:

$$\ln w_{ijkt} = \delta_j + \alpha_t + \beta^{gov} gov_{it} + \beta_k^{gov} gov_{it} + \beta^{elast} z_j^{elast} * gov_{it} + \beta X_{it} + \varepsilon_{ijt}.$$

This setup estimates the relation between state government-private sector wage gaps and local housing supply elasticities within states, across MSAs. Columns 5 and 6 of Table 2.6 show that the impact of land unavailability on state government-private sector wages gaps is not statically significant and that the point estimates are negative.

While state level variation in housing supply elasticity impacts state worker-private sector wage gaps, variation across MSAs within a state have no impact on the state worker-private sector wage gap, exactly as predicted by the model. Further, federal worker-private sector wages gaps are unaffected by state level or MSA level variation in housing supply elasticities, as also predicted by the model. However, local government worker-private sector wage gaps vary across MSAs both within and across states. Additionally, the impact of housing supply elasticities of these the public-private sector wage gaps is larger for unionized government workers. This evidence suggests that governments are exercising market power and over paying their employees, relative to the private sector.

The empirical evidence shows that housing supply elasticity impacts the average wage gap between public and private sector workers. A possible alternative explanation for this result other than rent-seeking and market power is that housing supply elasticity influences the type of workers state and local governments choose to employ. The wage gap between public and private sector workers could represent unobserved skill differences between workers employed in the public and private sectors. If this were true, the regressions previously presented which controlled for 3-digit occupation codes should have had much smaller point estimates than those which did not control for occupation, since there is likely less variation in worker skill within occupation than between.

As an additional test of this alternative hypothesis, I assess whether public-private sector workers years of education gaps vary with state and local housing supply elasticities. Table 2.7 preforms the standard analysis used to analyze state and local wage gaps, but replaces the left hand side variable with a worker's years of education. If government workers are higher skilled that private sector workers in housing inelastic areas, then this should hold both for observed skills (education) and unobserved skills (which cannot be tested). Table 2.7 shows that impact of land unavailability on public-private sector education gaps is not statistically significant. This holds in the state government workers sample and local government workers sample. This result is also robust to dropping worker demographics as controls in the regressions. Overall, differences in public and private sector workers' years of schooling to not appear to relate to state and local housing supply elasticities. Columns 5 through 8 of Table 2.7 reports additional robustness by re-doing the same analysis with the left-hand side variable equal to a dummy of whether the worker has a four year college degree. These results further show housing supply elasticity does not positively impact public-private sector worker skill differences. Government workers' wages appear to reflect the market power of state and local governments.

Table 2.7: Workers' Education vs. State Public Sector-Housing Supply Elasticity Interactions

	Education Measure: Years of Schooling				Education Measure: 4 year College Degree (0/1)			
	State Gov-Private Sector Wage Gaps		Local Gov-Private Sector Wage Gaps		State Gov-Private Sector Wage Gaps		Local Gov-Private Sector Wage Gaps	
	1	2	3	4	5	6	7	8
Government Worker	1.719***	1.836***	1.390***	1.463***	0.266***	0.264***	0.238***	0.230***
	[0.0634]	[0.0780]	[0.0356]	[0.0655]	[0.0119]	[0.0121]	[0.00679]	[0.00732]
Gov* Land Unavailability	0.0383	0.124	0.00751	0.0794	-0.00263	0.0048	-0.0095	-0.00127
	[0.0577]	[0.0875]	[0.0421]	[0.0823]	[0.00827]	[0.00839]	[0.00666]	[0.00732]
Gov* Land Use Regulation	-0.235***	-0.249***	-0.0299	0.0259	-0.0535***	-0.0530***	-0.0191**	-0.00755
	[0.0641]	[0.0714]	[0.0425]	[0.0538]	[0.0100]	[0.0101]	[0.00758]	[0.00653]
Constant	13.52***	12.94***	13.91***	13.44***	0.354***	0.178***	0.424***	0.269***
	[0.185]	[0.0174]	[0.172]	[0.0203]	[0.0327]	[0.00288]	[0.0385]	[0.00585]
Demographic Controls:	X		X		X		X	
MSA FE			X	X	X	X		
State FE	X	X					X	X
State Gov Workers Sample:	X	X			X	X		
Local Gov Workers Sample:			X	X			X	X
Observations	973,792	1,272,258	586,696	586,696	973,792	1,272,258	586,696	586,696
R-Squared	0.181	0.039	0.186	0.042	0.091	0.03	0.094	0.029

Note: Standard errors clustered by state for state worker regressions. Standard errors clustered by MSA for local government worker regressions. Sample is from 1995-2011 CPS MORG. Sample is restricted to 25-55 year old workers working at least 35 hours per week. State government worker sample includes private sector and state government workers. Local government worker sample includes private sector and local government workers living in MSAs. Land unavailability measures a z-score of the share of land within 50km of an MSA's center which cannot be developed due to topological constraints. Land use regulation is an index of aggregated survey data on a large set of local land use practices by local municipalities. This is aggregated to the MSA level and then Z-scored. State level measures are z-scores of average MSA-level measures within the state, weighted by MSA population. Controls include gender, race and hispanic origin, a quartic in age, a rural dummy, and year dummies. All regressions weighted using CPS MORG earnings weights. *** p<0.01, ** p<0.05, * p<0.1

2.3.3 Benefits

Gittleman and Pierce (2012) show that government workers' benefits are more generous than private sector workers' benefits. If the market power of state and local governments allows government workers to earn more desirable wages than similar private sector workers, this should also be true for public-private differences in the generosity of benefits.

As a measure of government workers' pension benefits, I use data from the Census' 2007-2010 Annual Surveys of Public Employee Retirement Systems. This data is collected annually from states governments' pension plans on the aggregate amount of retirement benefits paid out during the year, as well as the total number of beneficiaries who received a transfer that year. Taking the ratio of these, gives the average pension payout per beneficiary. Table 2.1 reports summary statistics on this data. Unfortunately, there is not a similar data set for retirement payouts to private sector workers.

An indirect test of whether benefits augment or offset wage gap differences is to assess whether the state worker-private sector wage gap negatively varies with pension payouts per retiree. If the public-private wage gap is high when public pension benefits are low, than changes in wage gaps across states might be offset by changes in benefits across states. However, Table 2.8 shows a regression of state government pensions payouts per retiree is strongly positively correlated with the public-private sector wage gap. This suggests that increases in the public-private wage gap are positively associated with increases in the public-private benefits gap. The wage gap estimates are likely a lower bound of impact of government market power of government employees compensation since they do not account for the impacts on benefits.

Table 2.8: State Housing Supply Elasticity, State Pension Payout per State beneficiary & State Public-Private Government Wage Gaps

	1	2	3
Land Unavailability	0.0317	0.0674**	
	[0.0311]	[0.0329]	
Land Use Regulation	0.0925**		
	[0.0366]		
Public-Private State Wage Gap			2.830***
			[0.651]
Constant	2.932***	2.931***	3.130***
	[0.0318]	[0.0342]	[0.0534]
Observations	192	192	192
R-squared	0.219	0.099	0.287

Standard errors clustered by state. Data from 2007-2010 Annual Surveys of Public-Employee Retirement Systems. Annual payout per beneficiary is measured in log thousands of dollars.

If private sector benefits do not vary with states' housing supply elasticities, than a regression of state pension payouts per beneficiary on states' housing supply elasticities measures the impact of housing supply elasticity on government retirement benefits. Table 2.9 reports these regressions. I find a 1 standard deviation increases in a states' land unavailability increases annual retirement payouts per retired state government employee by 0.0674 log points. Government workers appear to receive better compensation in both wages and retirement benefits in areas where the government can exercise more market power.

2.4 Conclusion

By using housing supply elasticity as exogenous variation in governments' abilities to exercise market power, I show that the public-private sector wage gap is higher in areas where the government can extract more rent from residents. Further, this effect is stronger for unionized government workers, suggesting that public sector unions might influence governments to engage in rent seeking behavior. While I cannot gauge to what extent government workers are overcompensated overall, government market power appears to play a role in government worker compensation.

The spatial equilibrium model shows that the scope of governments' market power does not disappear when there is competition between a large number of governments or when each government is small. The local labor and housing market will respond to the tax policy choices of the state and local government, mitigating the disciplining effects of workers' voting with their feet through migration.

It is possible that the unmodeled political system where multiple candidates run for election and campaign for less wasteful government policies could compete away some of this government market power. However, the empirical evidence of this paper suggests that these rents have not been fully competed away.

These results also speak to the welfare effects of land-use regulation policy. While the decision to regulate real-estate development and population expansion has many costs and benefits not studied in this paper, my results show that decreasing a city's housing supply elasticity through regulation gives the local government more market power. Thus, the rise in land-use regulations since the 1970s may have had an unintended consequence of increasing rent seeking by governments and leading to overpaid government workers. State and local governments appear to take advantage of their market power and some of these rents are shared with government employees.

Chapter 3
Clustering, Spatial Correlations and Randomization Inference

3.1 Introduction

Many economic studies that analyze the causal effects of interventions on economic behavior study interventions or treatments that are constant within clusters whereas the outcomes vary at a more disaggregate level. In a typical example, and the one we focus on in this paper, outcomes are measured at the individual level, whereas interventions vary only at the state (cluster) level. Often, the effect of interventions is estimated using least squares regression. Since the mid-eighties Liang and Zeger (1986), Moulton (1986) empirical researchers in social sciences have generally been aware of the implications of within-cluster correlations in outcomes for the precision of such estimates. The typical approach is to allow for correlation between outcomes in the same state in the specification of the error covariance matrix. However, there may well be more complex correlation patterns in the data. Correlation in outcomes between individuals may extend beyond state boundaries, it may vary in magnitude between states, and it may be stronger in more narrowly defined geographical areas.

In this paper we investigate the implications, for the repeated sampling variation of least squares estimators based on individual-level data, of the presence of correlation structures beyond those that are constant within and identical across states, and vanish between states. First, we address the empirical question whether in census data on earnings with states as clusters such correlation patterns are present. We estimate general spatial correlations for the logarithm of earnings, and find that, indeed, such correlations are present, with substantial correlations within groups of nearby states, and correlations within smaller geographic units (specifically pumas, public use microdata areas) considerably larger than within states. Second, we address whether accounting for such correlations is important for the properties of confidence intervals for the effects of state-level regulations or interventions. We report theoretical results, and demonstrate their relevance using illustrations based on earnings data and state regulations, as well as Monte Carlo evidence. The theoretical results show that if covariate values are as good as randomly assigned to clusters, implying there is no spatial correlation in the covariates beyond the clusters, variance estimators that incorporate only cluster-level outcome correlations remain valid despite the misspecification of the error-covariance matrix. Whether this theoretical result is useful in practice depends on the magnitude of the spatial correlations in the covariates. We provide some illustrations that show that, given the spatial correlation patterns we find in the individual-level variables, spatial correlations in state level regulations can have a substantial impact on the precision of estimates of the effects of interventions.

The paper draws on three strands of literature that have largely evolved separately. First, it is related to the literature on clustering, where a primary focus is on adjustments to standard errors to take into account clustering of explanatory variables. See, e.g., Liang and Zeger (1986), Moulton (1986), Bertrand, Duflo, and Mullainathan (2004), Hansen (2007) and the textbook discussions in Angrist and Pischke (2009), Diggle, Heagerty, Liang, and Zeger (2002), and Wooldridge (2002). Second, the current paper draws on the literature on spatial statistics. Here a major focus is on the specification and estimation of the covariance structure of spatially linked data. For textbook discussions see Schabenberger and Gotway (2004) and and Gelfand, Diggle, Guttorp, and Fuentes (2010). In interesting recent work Bester, Conley, and Hansen (2011) and Ibragimov and Muller (2010) link some of the inferential issues in the spatial and clustering literatures. Finally, we use results from the literature on randomization inference going back to Fisher (1925) and Neyman (1990). For a recent textbook discussion see Rosenbaum (2002). Although the calculation of Fisher exact p-values based on randomization inference is frequently used in the spatial statistics literature Schabenberger and Gotway (2004), and sometimes in the clustering literature Bertrand, Duflo, and Mullainathan (2004), Abadie, Diamond, and Hainmueller (2010), Neyman's approach to constructing confidence intervals using the randomization distribution is rarely used in these settings. We will argue that the randomization perspective provides useful insights into the interpretation and properties of confidence intervals in the context of spatially linked data.

The paper is organized as follows. In Section 3.2 we introduce the basic set-up. Next, in Section 3.3, using census data on earnings, we establish the presence of spatial correlation patterns beyond the constant-within-state correlations typically allowed for in empirical work. In Section 3.4 we discuss randomization-based methods for inference, first focusing on the case with randomization at the individual level. Section 3.5 extends the results to cluster-level randomization. In Section 3.6, we present the main theoretical results. We show that if cluster-level covariates are randomly assigned to the clusters, the standard variance estimator based on within-cluster correlations can be robust to misspecification of the error-covariance matrix. Next, in Section 3.7 we show, using Mantel-type tests, that a number of regulations exhibit substantial regional correlations, suggesting that ignoring the error correlation structure may lead to invalid confidence intervals. Section 3.8 reports the results of a small simulation study. Section 3.9 concludes. Proofs are collected in the Appendix.

3.2 Framework

Consider a setting where we have information on N units, say individuals in the United States, indexed by $i = 1, \ldots, N$. Associated with each unit is a location Z_i, measuring latitude and longitude for individual i. Associated with a location z are a unique puma $P(z)$ (public use microdata area, a census bureau defined area with at least 100,000 individuals), a state $S(z)$, and a division $D(z)$ (also a census bureau defined concept, with nine divisions in the United States). In our application the sample is divided into 9 divisions, which are then divided into a total of 49 states (we leave out individuals from Hawaii and Alaska, and include the District of Columbia as a separate state), which are then divided into 2,057 pumas. For individual i, with location Z_i, let P_i, S_i, and D_i, denote the puma, state, and division associated with the location Z_i. The distance $d(z, z')$ between two locations z and z' is defined as the shortest distance, in miles, on the earth's surface connecting the two points. To be precise, let $z = (z_{\text{lat}}, z_{\text{long}})$ be the latitude and longitude of a location. Then the formula for the distance in miles between two locations z and z' we use is

$$d(z, z') = 3,959 \times \arccos(\cos(z_{\text{long}} - z'_{\text{long}}) \cdot \cos(z_{\text{lat}}) \cdot \cos(z'_{\text{lat}}) + \sin(z_{\text{lat}}) \cdot \sin(z'_{\text{lat}})).$$

In this paper, we focus primarily on estimating the slope coefficient β in a linear regression of some outcome Y_i (e.g., the logarithm of individual level earnings for working men) on a binary intervention or treatment W_i (e.g., a state-level regulation), of the form

$$Y_i = \alpha + \beta \cdot W_i + \varepsilon_i. \tag{3.1}$$

A key issue is that the explanatory variable W_i may be constant withing clusters of individuals. In our application W_i varies at the state level.

Let ε denote the N-vector with typical element ε_i, and let $\mathbf{Y}$, $\mathbf{W}$, $\mathbf{P}$, $\mathbf{S}$, and $\mathbf{D}$, denote the N-vectors with typical elements Y_i, W_i, P_i, S_i, and D_i. Let ι_N denote the N-vector of ones, let $X_i = (1, W_i)$, and let $\mathbf{X}$ and $\mathbf{Z}$ denote the $N \times 2$ matrices with ith rows equal to X_i and Z_i, respectively, so that we can write in matrix notation

$$\mathbf{Y} = \iota_N \cdot \alpha + \mathbf{W} \cdot \beta + \varepsilon = \mathbf{X} \begin{pmatrix} \alpha & \beta \end{pmatrix}' + \varepsilon. \tag{3.2}$$

Let $N_1 = \sum_{i=1}^N W_i$, $N_0 = N - N_1$, $\overline{W} = N_1/N$, and $\overline{Y} = \sum_{i=1}^N Y_i/N$. We are interested in the distribution of the ordinary least squares estimators:

$$\hat{\beta}_{\text{ols}} = \frac{\sum_{i=1}^N (Y_i - \overline{Y}) \cdot (W_i - \overline{W})}{\sum_{i=1}^N (W_i - \overline{W})^2}, \qquad \text{and} \quad \hat{\alpha}_{\text{ols}} = \overline{Y} - \hat{\beta}_{\text{ols}} \cdot \overline{W}.$$

The starting point is the following model for the conditional distribution of $\mathbf{Y}$ given the location $\mathbf{Z}$ and the covariate $\mathbf{W}$:

Assumption 1. (Model)

$$\mathbf{Y} \,\Big|\, \mathbf{W} = \mathbf{w}, \mathbf{Z} = \mathbf{z} \sim \mathcal{N}(\iota_N \cdot \alpha + \ \mathbf{w} \cdot \beta, \Omega(\mathbf{z})).$$

Under this assumption we can infer the exact (finite sample) distribution of the least squares estimator, conditional on the covariates $\mathbf{X}$, and the locations $\mathbf{Z}$.

Lemma 2. (Distribution of Least Squares Estimator) *Suppose Assumption 1 holds. Then $\hat{\beta}_{\text{ols}}$ is unbiased and Normally distributed,*

$$\mathbb{E}\left[\hat{\beta}_{\text{ols}} \mid \mathbf{W}, \mathbf{Z}\right] = \beta, \quad \textit{and} \quad \hat{\beta}_{\text{ols}} \,\Big|\, \mathbf{W}, \mathbf{Z} \sim \mathcal{N}\left(\beta, \mathbb{V}_M(\mathbf{W}, \mathbf{Z})\right), \tag{3.3}$$

where

$$\mathbb{V}_M(\mathbf{W}, \mathbf{Z}) = \frac{1}{N^2 \cdot \overline{W}^2 \cdot (1 - \overline{W})^2} \begin{pmatrix} \overline{W} & -1 \end{pmatrix} \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix}' \Omega(\mathbf{Z}) \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix} \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix}. \tag{3.4}$$

We write the model-based variance $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$ as a function of $\mathbf{W}$ and $\mathbf{Z}$ to make explicit that this variance is conditional on both the treatment indicators $\mathbf{W}$ and the locations $\mathbf{Z}$. This lemma follows directly from the standard results on least squares estimation and is given without proof. Given Assumption 1, the exact distribution for the least squares coefficients $(\hat{\alpha}_{\text{ols}}, \hat{\beta}_{\text{ols}})'$ is Normal, centered at $(\alpha, \beta)'$ and with covariance matrix $(\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\Omega(\mathbf{Z})\mathbf{X}) (\mathbf{X}'\mathbf{X})^{-1}$. We then obtain the variance for $\hat{\beta}_{\text{ols}}$ in (3.4) by writing out the component matrices of the joint variance of $(\hat{\alpha}_{\text{ols}}, \hat{\beta}_{\text{ols}})'$.

It is also useful for the subsequent discussion to consider the variance of $\hat{\beta}_{\text{ols}}$, conditional on the locations $\mathbf{Z}$, and conditional on $N_1 = \sum_{i=1}^N W_i$, without conditioning on the entire vector $\mathbf{W}$. With some abuse of language, we refer to this as the unconditional variance $\mathbb{V}_U(\mathbf{Z})$ (although it is still conditional on $\mathbf{Z}$ and N_1). Because the conditional and unconditional expectation of $\hat{\beta}_{\text{ols}}$ are both equal to β, it follows that the unconditional variance is simply the expected value of the model-based variance:

$$\begin{aligned}\mathbb{V}_U(\mathbf{Z}) &= \mathbb{E}[\mathbb{V}_M(\mathbf{W},\mathbf{Z}) \mid \mathbf{Z}] \\ &= \frac{N^2}{N_0^2 \cdot N_1^2} \cdot \mathbb{E}\left[(\mathbf{W} - N_1/N \cdot \iota_N)'\Omega(\mathbf{Z})(\mathbf{W} - N_1/N \cdot \iota_N) \mid \mathbf{Z}\right].\end{aligned} \tag{3.5}$$

3.3 Spatial Correlation Patterns in Earnings

In this section we provide some evidence for the presence and structure of spatial correlations, that is, how Ω varies with $\mathbf{Z}$. Specifically we show in our application, first, that the structure is more general than the state-level correlations that are typically allowed for, and second, that this matters for inference.

We use data from the 5% public use sample from the 2000 census. Our sample consists of 2,590,190 men at least 20 and at most 50 years old, with positive earnings. We exclude individuals from Alaska, Hawaii, and Puerto Rico (these states share no boundaries with other states, and as a result spatial correlations may be very different than those for other states), and treat DC as a separate state, for a total of 49 "states". Table 3.1 presents some summary statistics for the sample. Our primary outcome variable is the logarithm of yearly earnings, in deviations from the overall mean, denoted by Y_i. The overall mean of log earnings is 10.17, the overall standard deviation is 0.97. We do not have individual level locations. Instead we know for each individual only the puma (public use microdata area) of residence, and so we take Z_i to be the latitude and longitude of the center of the puma of residence.

Table 3.1: Summary statistics

Average log earnings	10.17
Standard deviation of log earnings	0.97
Number of pumas in the sample	2,057
Average number of observations per puma	1,259
Standard deviation of number of observations per puma	409
Number of states (incl DC, excl AK, HA, PR) in the sample	49
Average number of observations per state	52,861
Standard deviation of number of observations per state	58,069
Number of divisions in the sample	9
Average number of observations per division	287,798
Standard deviation of number of observations per division	134,912

Notes: 2000 census data ($N = 2,590,190$)

Let $\mathbf{Y}$ be the variable of interest, in our case log earnings in deviations from the overall mean. Suppose we model the vector $\mathbf{Y}$ as

$$\mathbf{Y} \mid \mathbf{Z} \sim \mathcal{N}(0, \Omega(\mathbf{Z}, \gamma)).$$

If researchers have covariates that vary at the state level, the conventional strategy is to allow for correlation at the same level of aggregation ("clustering by state"), and model the covariance matrix as

$$\Omega_{ij}(\mathbf{Z}, \gamma) = \sigma_\varepsilon^2 \cdot \mathbf{1}_{i=j} + \sigma_S^2 \cdot \mathbf{1}_{S_i = S_j} = \begin{cases} \sigma_S^2 + \sigma_\varepsilon^2 & \text{if } i = j \\ \sigma_S^2 & \text{if } i \neq j,\ S_i = S_j \\ 0 & \text{otherwise,} \end{cases} \tag{3.6}$$

where $\Omega_{ij}(\mathbf{Z}, \gamma)$ is the (i, j)th element of $\Omega(\mathbf{Z}, \gamma)$. The first variance component, σ_ε^2, captures the variance of idiosyncratic errors, uncorrelated across different individuals. The second variance component, σ_S^2, captures correlations between individuals in the same state. Estimating σ_ε^2 and σ_S^2 on our sample of 2,590,190 individuals by maximum likelihood leads to $\hat{\sigma}_\varepsilon^2 = 0.929$ and $\hat{\sigma}_S^2 = 0.016$. The question addressed in this section is whether the covariance structure in (3.6) provides an accurate approximation to the true covariance matrix $\Omega(\mathbf{Z})$. We provide two pieces of evidence that it is not.

The first piece of evidence against the simple covariance matrix structure is based on simple descriptive measures of the correlation patterns as a function of distance between individuals. For a distance d (in miles), define the overall, within-state, and out-of-state covariances as

$$C(d) = \mathbb{E}\left[Y_i \cdot Y_j \middle|\, d(Z_i, Z_j) = d\right],$$

$$C_S(d) = \mathbb{E}\left[Y_i \cdot Y_j \middle|\, S_i = S_j, d(Z_i, Z_j) = d\right],$$

and

$$C_{\overline{S}}(d) = \mathbb{E}\left[Y_i \cdot Y_j \middle| S_i \neq S_j, d(Z_i, Z_j) = d \right].$$

If the model in (3.6) was correct, then $C_S(d)$ should be constant (but possibly non-zero) as a function of the distance d, and $C_{\overline{S}}(d)$ should be equal to zero for all d.

We estimate these covariances using averages of the products of individual level outcomes for pairs of individuals whose distance is within some bandwidth h of the distance d:

$$\widehat{C(d)} = \sum_{i<j} \mathbf{1}_{|d(Z_i,Z_j)-d|\leq h} \cdot Y_i \cdot Y_j \bigg/ \sum_{i<j} \mathbf{1}_{|d(Z_i,Z_j)-d|\leq h},$$

$$\widehat{C_S(d)} = \sum_{i<j, S_i=S_j} \mathbf{1}_{|d(Z_i,Z_j)-d|\leq h} \cdot Y_i \cdot Y_j \bigg/ \sum_{i<j, S_i=S_j} \mathbf{1}_{|d(Z_i,Z_j)-d|\leq h},$$

and

$$\widehat{C_{\overline{S}}(d)} = \sum_{i<j} \mathbf{1}_{S_i\neq S_j} \cdot \mathbf{1}_{|d(Z_i,Z_j)-d|\leq h} \cdot Y_i \cdot Y_j \bigg/ \sum_{i<j, S_i\neq S_j} \mathbf{1}_{|d(Z_i,Z_j)-d|\leq h}.$$

Figure 3.1 shows the covariance functions for $\widehat{C_S(d)}$ and $\widehat{C_{\overline{S}}(d)}$ for the bandwidth $h = 50$ miles for the within-state and out-of-state covariances. (Results based on a bandwidth $h = 20$ are similar.) The main conclusion from Figure 3.1a is that within-state correlations decrease with distance. Figure 3.1b suggests that correlations for individuals in different states are non-zero, also decrease with distance, and are of a magnitude similar to within-state correlations. Thus, these figures suggest that the simple covariance model in (3.6) is not an accurate representation of the true covariance structure.

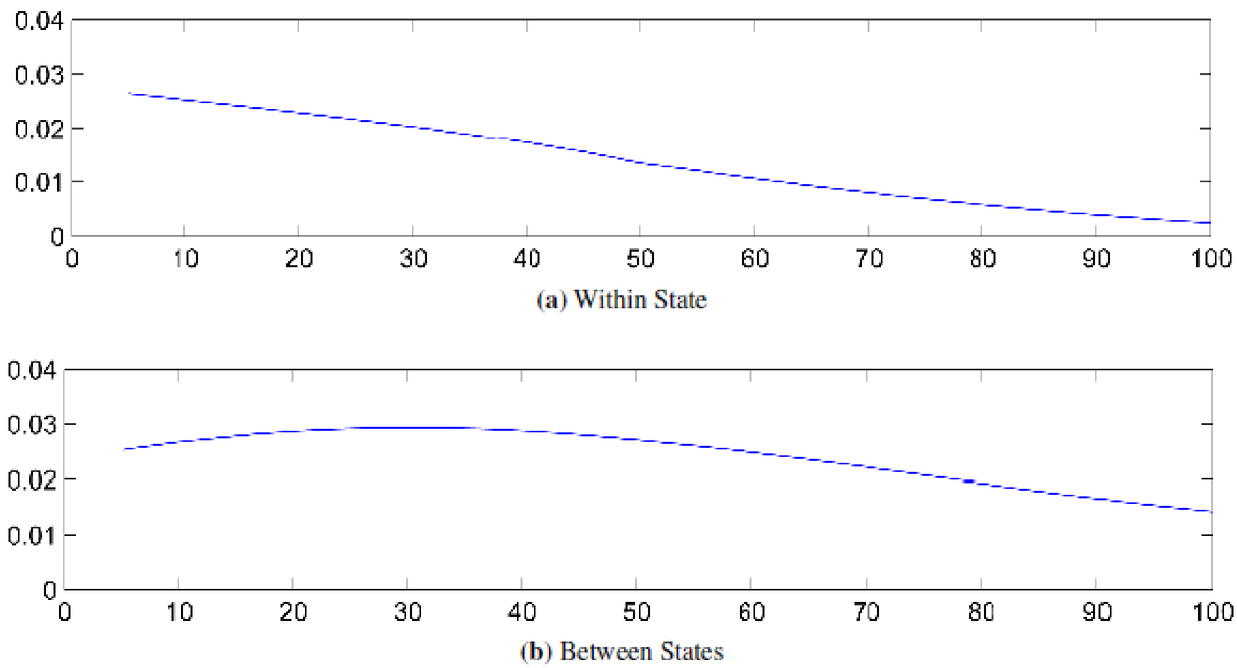


Figure 3.1: Covariance of demeaned log earnings of individuals as function of distance (in miles). Bandwidth $h = 50$ miles.

As a second piece of evidence we consider various parametric structures for the covariance matrix $\Omega(\mathbf{Z})$ that generalize (3.6). At the most general level, we specify the following form for $\Omega_{ij}(\mathbf{Z}, \gamma)$:

$$\Omega_{ij}(\mathbf{Z}, \gamma) = \begin{cases} \sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(Z_i, Z_j)) + \sigma^2_D + \sigma^2_S + \sigma^2_P + \sigma^2_\varepsilon & \text{if } i = j, \\ \sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(Z_i, Z_j)) + \sigma^2_D + \sigma^2_S + \sigma^2_P & \text{if } i \neq j, P_i = P_j, \\ \sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(Z_i, Z_j)) + \sigma^2_D + \sigma^2_S & \text{if } P_i \neq P_j, S_i = S_j, \\ \sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(Z_i, Z_j)) + \sigma^2_D & \text{if } S_i \neq S_j, D_i = D_j, \\ \sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(Z_i, Z_j)) & \text{if } D_i \neq D_j. \end{cases} \tag{3.7}$$

Beyond state level correlations the most general specification allows for correlations at the puma level (captured by σ^2_P) and at the division level (captured by σ^2_D). In addition we allow for spatial correlation as a smooth function geographical distance, declining at an exponential rate, captured by $\sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(z, z'))$. Although more general than the typical covariance structure allowed for, this model still embodies important restrictions, notably that correlations do not vary by location. A more general model might allow variances or covariances to vary directly by the location z, e.g., with correlations stronger or weaker in the Western versus the Eastern United States, or in more densely or sparsely populated parts of the country.

Table 3.2 gives maximum likelihood estimates for the covariance parameters γ given various restrictions, based on the log earnings data, with standard errors based on the second derivatives of the log likelihood function. To put these numbers in perspective, the estimated value for α in the most general model, $\hat{\alpha} = 0.0293$, implies that the pure spatial component, $\sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(z, z'))$, dies out fairly quickly: at a distance of about twenty-five miles the spatial covariance due to the $\sigma^2_{\text{dist}} \cdot \exp(-\alpha \cdot d(z, z'))$ component is half what it is at zero miles. The covariance of log earnings for two individuals in the same puma is 0.080/0.948 = 0.084. For these data, the covariance between log earnings and years of education is approximately 0.3, so the within-puma covariance is substantively important, equal to about 30% of the log earnings and education covariance. For two individuals in the same state, but in different pumas and ignoring the spatial component, the total covariance is 0.013. The estimates suggest that much of what shows up as within-state correlations in a model such as (3.6) that incorporates only within-state correlations, in fact captures much more local, within-puma, correlations.

To show that these results are typical for the type of correlations found in individual level economic data, we calculated results for the same models as in Table 3.2 for two other variables collected in the census, years of education and hours worked. Results for those variables are reported in an earlier version of the paper Barrios, Imbens, Diamond, and Kolesár (2010). In all cases puma-level correlations are an order of magnitude larger than within-state out-of-puma level correlations, and within-division correlations are of the same order of magnitude as within-state correlations.

The two sets of results, the covariances by distance and the model-based estimates of cluster contributions to the variance, both suggest that the simple model in (3.6) that assumes zero covariances for individuals in different states, and constant covariances for individuals in the same state irrespective of distance, is at odds with the data. Covariances vary substantially within states, and do not vanish at state boundaries.

Table 3.2: Estimates for clustering variances for demeaned log earnings

							$\widehat{\text{s.e.}}(\hat{\beta})$	
σ_ε^2	σ_D^2	σ_S^2	σ_P^2	σ_{dis}^2	α	Log Lik	MW	NE/ENC
0.931	0	0	0	0	0	-1213298	0.002	0.002
[0.001]								
0.929	0	0.016	0	0	0	-1200407	0.080	0.057
[0.001]		[0.002]						
0.868	0	0.011	0.066	0	0	-1116976	0.068	0.049
[0.001]		[0.003]	[0.002]					
0.929	0.006	0.011	0	0	0	-1200403	0.091	0.081
[0.001]	[0.002]	[0.002]						
0.868	0.006	0.006	0.066	0	0	-1116972	0.081	0.076
[0.001]	[0.003]	[0.002]	[0.002]					
0.868	0.005	0.006	0.047	0.021	0.029	-1116892	0.074	0.085
[0.001]	[0.005]	[0.001]	[0.002]	[0.003]	[0.005]			

Notes: Standard errors based on the second derivative of log-likelihood in square brackets. Log Lik refers to the value of the log-likelihood function evaluated at the maximum likelihood estimates. The last two columns refer to the implied standard errors if the regressor is and indicator for high state minimum wage (MW) or and indicator for the state being in New England or the East-North-Central Division (NE/ENC).

Now we turn to the second question of this section, whether the magnitude of the correlations we reported matters for inference. In order to assess this, we look at the implications of the models for the correlation structure for the precision of least squares estimates. To make this specific, we focus on the model in (3.1), with log earnings as the outcome Y_i, and W_i equal to an indicator that individual i lives in a state with a minimum wage that is higher than the federal minimum wage in the year 2000. This indicator takes on the value one for individuals living in nine states in our sample, California, Connecticut, Delaware, Massachusetts, Oregon, Rhode Island, Vermont, Washington, and DC, and zero for all other states in our sample (see Figure 3.2 for a visual impression).[49] In the second to last column in Table 3.2, under the label "MW," we report in each row the standard error for $\hat{\beta}_{\text{ols}}$ based on the specification for $\Omega(\mathbf{Z}, \gamma)$ in that row. To be precise, if $\hat{\Omega} = \Omega(\mathbf{Z}, \hat{\gamma})$ is the estimate for $\Omega(\mathbf{Z}, \gamma)$ in a particular specification, the standard error is

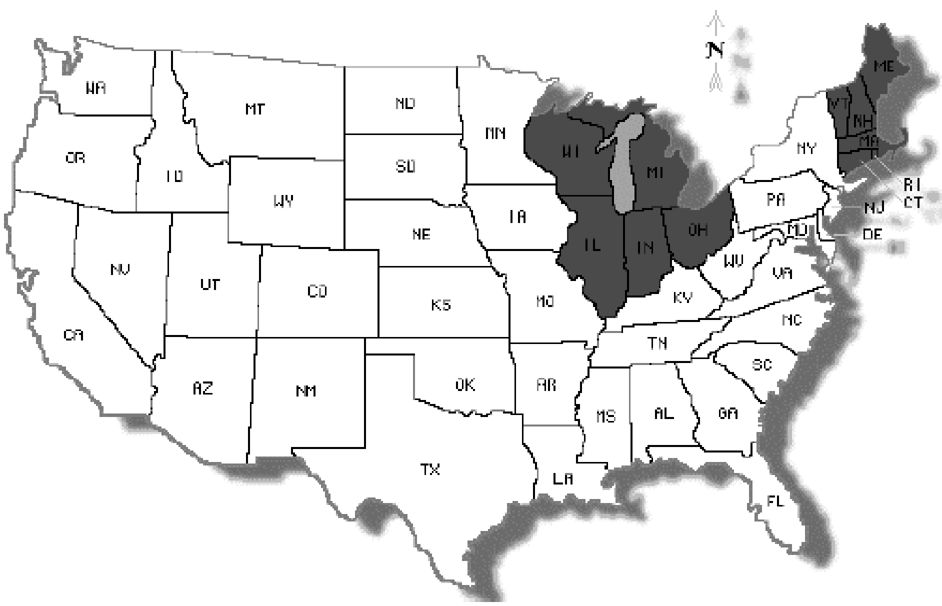
$$\text{s.e.}(\hat{\beta}_{\text{ols}}) = \left(\frac{1}{N^2 \overline{W}^2 (1 - \overline{W})^2} \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix}' \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix}' \Omega(\mathbf{Z}, \hat{\gamma}) \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix} \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix} \right)^{1/2}.$$

With no correlation between units at all, the estimated standard error is 0.002. If we allow only for state level correlations, Model (3.6), the estimated standard error goes up to 0.080, demonstrating the well known importance of allowing for correlation at the level that the covariate varies. There are two general points to take away from the column with standard errors. First, the biggest impact on the standard errors comes from incorporating state-level correlations (allowing σ_S^2 to differ from zero), even though according to the variance component estimates other variance components are substantially more important. Second, among the specifications that allow for $\sigma_S^2 \neq 0$, however, there is still a substantial amount of variation in the implied standard errors. Incorporating only σ_S^2 leads to a standard error around 0.080, whereas also including division-level correlations ($\sigma_D^2 \neq 0$) increase that

[49] The data come from the website `http://www.dol.gov/whd/state/stateMinWageHis.htm`. To be consistent with the 2000 census, we use the information from 2000, not the current state of the law.

to approximately 0.091, an increase of 15%. We repeat this exercise for a second binary covariate, with the results reported in the last column of Table 3.2. In this case the covariate takes on the value one only for the New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont) and East-North-Central states (Illinois, Indiana, Michigan, Ohio, and Wisconsin), collectively referred to as the NE/ENC states from here on. This set of states corresponds to more geographical concentration than the set of minimum wage states (see Figure 3.2). In this case, the impact on the standard errors of mis-specifying the covariance structure $\Omega(\mathbf{Z})$ is even larger, with the most general specification leading to standard errors that are almost 50% larger than those based on the state-level correlations specification (3.6). In the next three sections we explore theoretical results that provide some insight into these empirical findings.

(a) State minimum wage higher than federal minimum wage (CA, CT, DC, DE, MA, OR, RI, VT, WA)



(b) New England/East North Central states (CT, IL, IN, MA, ME, MI, OH, RI, NH, VT, WI)

Figure 3.2: Spatial correlation of regressors

3.4 Randomization Inference

In this section we consider a different approach to analyzing the distribution of the least squares estimator, based on randomization inference Rosenbaum (2002). Recall the linear model (3.1),

$$Y_i = \alpha + \beta \cdot W_i + \varepsilon_i, \qquad \text{with } \varepsilon|\mathbf{W}, \mathbf{Z} \sim \mathcal{N}(0, \Omega(\mathbf{Z})).$$

In Section 3.2 we analyzed the properties of the least squares estimator $\hat{\beta}_{\text{ols}}$ under repeated sampling. To be precise, the sampling distribution for $\hat{\beta}_{\text{ols}}$ was defined by repeated sampling in which we keep both the vector of treatments $\mathbf{W}$ and the location $\mathbf{Z}$ fixed on all draws, and redraw only the vector of residuals ε for each sample. Under this repeated sampling thought-experiment, the exact variance of $\hat{\beta}_{\text{ols}}$ is $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$ as given in Lemma 2.

It is possible to construct confidence intervals in a different way, based on a different repeated sampling thought-experiment. Instead of conditioning on the vector $\mathbf{W}$ and $\mathbf{Z}$, and resampling the ε, we can condition on ε and $\mathbf{Z}$, and resample the vector $\mathbf{W}$. To be precise, let $Y_i(0)$ and $Y_i(1)$ denote the potential outcomes under the two levels of the treatment W_i, and let $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$ denote the corresponding N-vectors. Then let $Y_i = Y_i(W_i)$ be the realized outcome. We assume that the effect of the treatment is constant, $Y_i(1) - Y_i(0) = \beta$. Defining $\alpha = \mathbb{E}[Y_i(0)]$, the residual is $\varepsilon_i = Y_i - \alpha - \beta \cdot W_i$. In this section we focus on the simplest case, where the covariate of interest W_i is completely randomly assigned, conditional on $\sum_{i=1}^{N} W_i = N_1$.

Assumption 3. Randomization

$$\text{pr}(\mathbf{W} = \mathbf{w} \mid \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) = 1 \Big/ \begin{pmatrix} N \\ N_1 \end{pmatrix}, \quad \textit{for all } \mathbf{w} \textit{ s.t. } \sum_{i=1}^{N} w_i = N_1.$$

Under this assumption we can infer the exact (finite sample) variance for the least squares estimator for $\hat{\beta}_{\text{ols}}$ conditional on $\mathbf{Z}$ and $(\mathbf{Y}(0), \mathbf{Y}(1))$:

Lemma 4. *Suppose that Assumption 3 holds and that the treatment effect $Y_i(1) - Y_i(0) = \beta$ is constant for all individuals. Then (i), $\hat{\beta}_{\text{ols}}$ conditional on $(\mathbf{Y}(0), \mathbf{Y}(1))$ and $\mathbf{Z}$ is unbiased for β,*

$$\mathbb{E}\left[\hat{\beta}_{\text{ols}}\middle|\,\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}\right] = \beta, \tag{3.8}$$

and, (ii), its exact conditional (randomization-based) variance is

$$\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) = \mathbb{V}\left(\hat{\beta}_{\text{ols}}\middle|\,\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}\right) = \frac{N}{N_0 \cdot N_1 \cdot (N-2)} \sum_{i=1}^{N} \left(\varepsilon_i - \overline{\varepsilon}\right)^2, \tag{3.9}$$

where $\overline{\varepsilon} = \sum_{i=1}^{N} \varepsilon_i / N$.

Because this result direct follows from results by Neyman (1990) on randomization inference for average treatment effects, specialized to the case with a constant treatment effect, the proof is omitted. Note that although the variance is exact, we do not have exact Normality, unlike the result in Lemma 2.

In the remainder of this section we explore two implications of the randomization perspective. First of all, although the model and randomization variances $\mathbb{V}_M$ and $\mathbb{V}_R$ are exact if both Assumptions 1 and 3 hold, they differ because they refer to different repeated sampling thought experiments, or, alternatively, to different conditioning sets. To illustrate this, let us consider the bias and variance under a third repeated sampling thought experiment, without conditioning on either $\mathbf{W}$ or ε, just conditioning on the locations $\mathbf{Z}$ and (N_0, N_1), maintaining both the model and the randomization assumption.

Lemma 5. *Suppose Assumptions 1 and 3 hold. Then (i), $\hat{\beta}_{\text{ols}}$ is unbiased for β,*

$$\mathbb{E}\left[\hat{\beta}_{\text{ols}}\middle|\,\mathbf{Z}, N_0, N_1\right] = \beta, \tag{3.10}$$

(ii), its exact unconditional variance is:

$$\mathbb{V}_U(\mathbf{Z}) = \left(\frac{1}{N-2}\,\text{trace}(\Omega(\mathbf{Z})) - \frac{1}{N \cdot (N-2)} \iota_N' \Omega(\mathbf{Z}) \iota_N\right) \cdot \frac{N}{N_0 \cdot N_1}, \tag{3.11}$$

and (*iii*),

$$\mathbb{V}_U(\mathbf{Z}) = \mathbb{E}\left[\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z})|\mathbf{Z}, N_0, N_1\right] = \mathbb{E}\left[\mathbb{V}_M(\mathbf{W}, \mathbf{Z})|\mathbf{Z}, N_0, N_1\right].$$

Thus, in expectation, $\mathbb{V}_R(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z})$, is equal to the expectation of $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$.

For the second point, suppose we had focused on the repeated sampling variance for $\hat{\beta}_{\text{ols}}$ conditional on $\mathbf{W}$ and $\mathbf{Z}$, but possibly erroneously modeled the covariance matrix as constant times the identify matrix, $\Omega(\mathbf{Z}) = \sigma^2 \cdot I_N$. Using such a (possibly incorrect) model a researcher would have concluded that the exact sampling distribution for $\hat{\beta}_{\text{ols}}$ conditional on the covariates would be

$$\hat{\beta}_{\text{ols}}\Big| \mathbf{W}, \mathbf{Z} \sim \mathcal{N}\left(\beta, \mathbb{V}_{\text{INC}}\right), \qquad \text{where} \quad \mathbb{V}_{\text{INC}} = \sigma^2 \cdot \frac{N}{N_0 \cdot N_1}. \tag{3.12}$$

If $\Omega(\mathbf{Z})$ differs from $\sigma^2 \cdot I_N$, then $\mathbb{V}_{\text{INC}}$ is not in general the correct (conditional) distribution for $\hat{\beta}_{\text{ols}}$. However, in some cases the misspecification need not lead to invalid inferences in large samples. To make that precise, we first need to define precisely how inference is performed. Implicitly the maximum likelihood estimator for the misspecified variance defines σ^2 as the probability limit of the estimator:

$$\hat{\sigma}^2 = \arg\max \left\{ \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(Y_i - \hat{\alpha}_{\text{ols}} - \hat{\beta}_{\text{ols}} W_i\right)^2 \right\}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{\alpha}_{\text{ols}} - \hat{\beta}_{\text{ols}} W_i\right)^2.$$

The probability limit for this estimator $\hat{\sigma}^2$, under Assumptions given in the Lemma below, is $\text{plim}(\ \text{trace}(\Omega(Z))/N)$. Then the probability limit of the normalized variance based on the possibly incorrect model is

$$N \cdot \mathbb{V}_{\text{INC}} = \text{plim}(\text{trace}(\Omega(Z))/N)\,\text{plim}\left(\frac{N^2}{N_0 \cdot N_1}\right).$$

The following result clarifies the properties of this probability limit.

Lemma 6. *Suppose Assumption 1 holds with* $\Omega(\mathbf{Z})$ *satisfying* $\mathrm{trace}(\Omega(\mathbf{Z}))/N \to c$ *for some* $0 < c < \infty$*, and* $\iota_N'\Omega(\mathbf{Z})\iota_N/N^2 \to 0$*, and Assumption 3 holds with* $N_1/N \to p$ *for some* $0 < p < 1$*. Then*

$$N \cdot (\mathbb{V}_{\mathrm{INC}} - \mathbb{V}_U(\mathbf{Z})) \xrightarrow{p} 0, \qquad \text{and} \quad N \cdot \mathbb{V}_{\mathrm{INC}} \xrightarrow{p} \frac{c}{p \cdot (1-p)}.$$

Hence, and this is a key insight of this section, if the assignment $\mathbf{W}$ is completely random, and the treatment effect is constant, one can, at least in large samples, ignore the off-diagonal elements of $\Omega(\mathbf{Z})$, and (mis-)specify $\Omega(\mathbf{Z})$ as $\sigma^2 \cdot I_N$. Although the resulting variance estimator will *not* be estimating the variance under the repeated sampling thought experiment that one may have in mind, (namely $\mathbb{V}_M(\mathbf{W}, \mathbf{Z})$), it leads to valid confidence intervals under the randomization distribution. The result that the mis-specification of the covariance matrix need not lead to inconsistent standard errors if the covariate of interest is randomly assigned has been noted previously. Greenwald (1983) writes: "when the correlation patterns of the independent variables are unrelated to those across the errors, then the least squares variance estimates are consistent." Angrist and Pischke (2009) write, in the context of clustering, that: "if the [covariate] values are uncorrelated within the groups, the grouped error structure does not matter for standard errors." The preceding discussion interprets this result formally from a randomization perspective.

3.5 Randomization Inference with Cluster-level Randomization

Now let us return to the setting that is the main focus of the paper. The covariate of interest, W_i, varies only between clusters (states), and is constant within clusters. Instead of assuming that W_i is randomly assigned at the individual level, we now assume that it is randomly assigned at the cluster level. Let M be the number of clusters, M_1 the number of clusters with all individuals assigned $W_i = 1$, and M_0 the number of clusters with all individuals assigned to $W_i = 0$. The cluster indicator is

$$C_{im} = \mathbf{1}_{S_i=m} = \begin{cases} 1 & \text{if individual } i \text{ is in cluster/state } m, \\ 0 & \text{otherwise,} \end{cases}$$

with $\mathbf{C}$ the $N \times M$ matrix with typical element C_{im}. For randomization inference we condition on $\mathbf{Z}$, ε, and M_1. Let N_m be the number of individuals in cluster m. We now look at the properties of $\hat{\beta}_{\text{ols}}$ over the randomization distribution induced by this assignment mechanism. To keep the notation precise, let $\tilde{\mathbf{W}}$ be the M-vector of assignments at the cluster level, with typical element $\tilde{W}_m$. Let $\tilde{\mathbf{Y}}(0)$ and $\tilde{\mathbf{Y}}(1)$ be M-vectors, with m-th element equal to $\tilde{Y}_m(0) = \sum_{i:C_{im}=1} Y_i(0)/N_m$, and $\tilde{Y}_m(1) = \sum_{i:C_{im}=1} Y_i(1)/N_m$ respectively. Similarly, let $\tilde{\varepsilon}$ be an M-vector with m-th element equal to $\tilde{\varepsilon}_m = \sum_{i:C_{im}=1} \varepsilon_i/N_m$, and let $\overline{\tilde{\varepsilon}} = \sum_{m=1}^{M} \tilde{\varepsilon}_m/M$.

Formally the assumption on the assignment mechanism is now:

Assumption 7. (Cluster Randomization)

$$\text{pr}(\tilde{\mathbf{W}} = \tilde{\mathbf{w}} \mid \mathbf{Z} = \mathbf{z}) = 1 \Big/ \begin{pmatrix} M \\ M_1 \end{pmatrix}, \quad \textit{for all } \mathbf{w} \textit{ s.t. } \sum_{m=1}^{M} \tilde{w}_m = M_1, \textit{ and } 0 \textit{ otherwise.}$$

We also make the assumption that all clusters are the same size:

Assumption 8. (Equal Cluster Size) *$N_m = N/M$ for all $m = 1, \ldots, M$.*

Lemma 9. *Suppose Assumptions 7 and 8 hold, and the treatment effect $Y_i(1) - Y_i(0) = \beta$ is constant. Then (i), the exact sampling variance of $\hat{\beta}_{\text{ols}}$, conditional on $\mathbf{Z}$ and ε, under the cluster randomization distribution is*

$$\mathbb{V}_{CR}(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) = \frac{M}{M_0 \cdot M_1 \cdot (M-2)} \sum_{m=1}^{M} \left(\tilde{\varepsilon}_m - \overline{\tilde{\varepsilon}}\right)^2, \tag{3.13}$$

(ii) if also Assumption 1 holds, then the unconditional variance is

$$\mathbb{V}_U(\mathbf{Z}) = \mathbb{E}\left[\mathbb{V}_{CR}(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) | \mathbf{Z}, M_1\right] = \frac{M^2}{M_0 \cdot M_1 \cdot (M-2) \cdot N^2} \cdot \left(M \cdot \mathrm{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right) - \iota'\Omega(\mathbf{Z})\iota\right). \quad (3.14)$$

The unconditional variance is a special case of the expected value of the unconditional variance in (3.5), with the expectation taken over $\mathbf{W}$ given the cluster-level randomization.

3.6 Variance Estimation Under Misspecification

In this section we present the main theoretical result in the paper. It extends the result in Section 3.4 on the robustness of model-based variance estimators under complete randomization to the case where the model-based variance estimator accounts for clustering, but not necessarily for all spatial correlations, and that treatment is randomized at cluster level.

Suppose the model generating the data is the linear model in (3.1), with a general covariance matrix $\Omega(\mathbf{Z})$, and Assumption 1 holds. The researcher estimates a parametric model that imposes a potentially incorrect structure on the covariance matrix. Let $\Omega(\mathbf{Z}, \gamma)$ be the parametric model for the error covariance matrix. The model is misspecified in the sense that there need not be a value γ such that $\Omega(\mathbf{Z}) = \Omega(\mathbf{Z}, \gamma)$. The researcher then proceeds to calculate the variance of $\hat{\beta}_{\text{ols}}$ as if the postulated model is correct. The question is whether this implied variance based on a misspecified covariance structure leads to correct inference.

The example we are most interested in is characterized by a clustering structure by state. In that case $\Omega(\mathbf{Z}, \gamma)$ is the $N \times N$ matrix with $\gamma = (\sigma_\varepsilon^2, \sigma_S^2)'$, where

$$\Omega_{ij}(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_S^2) = \begin{cases} \sigma_\varepsilon^2 + \sigma_S^2 & \text{if } i = j \\ \sigma_S^2 & \text{if } i \neq j, S_i = S_j, \\ 0 & \text{otherwise.} \end{cases} \quad (3.15)$$

Initially, however, we allow for any parametric structure $\Omega(\mathbf{Z}, \gamma)$. The true covariance matrix

$\Omega(\mathbf{Z})$ may include correlations that extend beyond state boundaries, and that may involve division-level correlations or spatial correlations that decline smoothly with distance as in the specification (3.7).

Under the (misspecified) parametric model $\Omega(\mathbf{Z},\gamma)$, let $\tilde{\gamma}$ be the pseudo true value, defined as the value of γ that maximizes the expectation of the logarithm of the likelihood function,

$$\tilde{\gamma} = \arg\max_{\gamma} \mathbb{E}\left[-\frac{1}{2}\cdot \ln\left(\det\left(\Omega(\mathbf{Z},\gamma)\right)\right) - \frac{1}{2}\cdot \varepsilon'\Omega(\mathbf{Z},\gamma)^{-1}\varepsilon \,\middle|\, \mathbf{Z}\right].$$

Given the pseudo true error covariance matrix $\Omega(\tilde{\gamma})$, the corresponding pseudo-true model-based variance of the least squares estimator, conditional on $\mathbf{W}$ and $\mathbf{Z}$, is

$$\mathbb{V}_{\text{INC,CR}} = \frac{1}{N^2\overline{W}^2(1-\overline{W})^2}\begin{pmatrix}\overline{W}\\ -1\end{pmatrix}'\begin{pmatrix}\iota_N & \mathbf{W}\end{pmatrix}'\Omega(\mathbf{Z},\tilde{\gamma})\begin{pmatrix}\iota_N & \mathbf{W}\end{pmatrix}\begin{pmatrix}\overline{W}\\ -1\end{pmatrix}.$$

Because for some $\mathbf{Z}$ the true covariance matrix $\Omega(\mathbf{Z})$ differs from the misspecified one, $\Omega(\mathbf{Z},\tilde{\gamma})$, it follows that in general this pseudo-true conditional variance $\mathbb{V}_M(\Omega(\mathbf{Z},\tilde{\gamma}),\mathbf{W},\mathbf{Z})$ will differ from the true variance $\mathbb{V}_M(\Omega(\mathbf{Z}),\mathbf{W},\mathbf{Z})$. Here we focus on the expected value of $\mathbb{V}_M(\Omega(\mathbf{Z},\tilde{\gamma}),\mathbf{W},\mathbf{Z})$, conditional on $\mathbf{Z}$, under assumptions on the distribution of $\mathbf{W}$. Let us denote this expectation by $\mathbb{V}_U(\Omega(\mathbf{Z},\tilde{\gamma}),\mathbf{Z}) = \mathbb{E}[\mathbb{V}_M(\Omega(\mathbf{Z},\tilde{\gamma}),\mathbf{W},\mathbf{Z})|\mathbf{Z}]$. The question is under what conditions on the specification of the error-covariance matrix $\Omega(\mathbf{Z},\gamma)$, in combination with assumptions on the assignment process, this unconditional variance is equal to the expected variance with the expectation of the variance under the correct error-covariance matrix, $\mathbb{V}_U(\Omega(\mathbf{Z}),\mathbf{Z}) = \mathbb{E}[\mathbb{V}_M(\Omega(\mathbf{Z}),\mathbf{W},\mathbf{Z})|\mathbf{Z}]$.

The following theorem shows that if the randomization of $\mathbf{W}$ is at the cluster level, then solely accounting for cluster level correlations is sufficient to get valid confidence intervals.

Theorem 10. (Clustering with Misspecified Error-Covariance Matrix) *Suppose Assumption 1 holds with* $\Omega(\mathbf{Z})$ *satisfying* $\text{trace}(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C})/N \to c$ *for some* $0 < c < \infty$*, and* $\iota_N'\Omega(\mathbf{Z})\iota_N/N^2 \to 0$*, Assumption 7 holds with* $M_1/M \to p$ *for some* $0 < p < 1$*, and*

Assumption 8 holds. Suppose also that that $\Omega(\mathbf{Z},\gamma)$ is specified as in (3.15). Then

$$N\cdot\left(\mathbb{V}_{\text{INC,CR}}-\mathbb{V}_U(\mathbf{Z})\right)\xrightarrow{p}0,\qquad\text{and}\quad N\cdot\mathbb{V}_{\text{INC,CR}}\xrightarrow{p}\frac{c}{N_m^2\cdot p\cdot(1-p)}.$$

This is the main theoretical result in the paper. It implies that if cluster level explanatory variables are randomly allocated to clusters, there is no need to consider covariance structures beyond those that allow for cluster level correlations. In our application, if the covariate (state minimum wage exceeding federal minimum wage) were as good as randomly allocated to states, then there is no need to incorporate division or puma level correlations in the specification of the covariance matrix. It is in that case sufficient to allow for correlations between outcomes for individuals in the same state. Formally the result is limited to the case with equal sized clusters. There are few exact results for the case with variation in cluster size, although if the variation is modest, one might expect the current results to provide useful guidance.

In many econometric analyses researchers specify the conditional distribution of the outcome given some explanatory variables, and ignore the joint distribution of the explanatory variables. The result in Theorem 10 shows that it may be useful to pay attention to this distribution. Depending on the joint distribution of the explanatory variables, the analyses may be robust to mis-specification of particular aspects of the conditional distribution. In the next section we discuss some methods for assessing the relevance of this result.

3.7 Spatial Correlation in State Averages

The results in the previous sections imply that inference is substantially simpler if the explanatory variable of interest is randomly assigned, either at the unit or cluster level. Here we discuss tests originally introduced by Mantel (1967) Schabenberger and Gotway (2004) to analyze whether random assignment is consistent with the data, against the alternative hypothesis of some spatial correlation. These tests allow for the calculation of exact, finite sample, p-values. To implement these tests we use the location of the units. To make the discussion more specific, we test the random assignment of state-level variables against the alternative of spatial correlation.

Let Y_s be the variable of interest for state s, for $s = 1, \ldots, S$, where state s has location Z_s (the centroid of the state). In the illustrations of the tests we use an indicator for a state-level regulation, and the state-average of an individual-level outcome. The null hypothesis of no spatial correlation in the Y_s can be formalized as stating that conditional on the locations $\mathbf{Z}$, each permutation of the values $(Y_1, \ldots, Y_S)$ is equally likely. With S states, there are $S!$ permutations. We assess the null hypothesis by comparing, for a given statistic $M(\mathbf{Y}, \mathbf{Z})$, the value of the statistic given the actual $\mathbf{Y}$ and $\mathbf{Z}$, with the distribution of the statistic generated by randomly permuting the $\mathbf{Y}$ vector.

The tests we focus on in the current paper are based on Mantel statistics Mantel (1967), Schabenberger and Gotway (2004). These general form of the statistics we use is Geary's c (also known as a Black-White or BW statistic in the case of binary outcomes), a proximity-weighted average of squared pairwise differences:

$$G(\mathbf{Y}, \mathbf{Z}) = \sum_{s=1}^{S-1} \sum_{t=s+1}^{S} (Y_s - Y_t)^2 \cdot d_{st}, \tag{3.16}$$

where $d_{st} = d(Z_s, Z_t)$ is a non-negative weight monotonically related to the proximity of the states s and t. Given a statistic, we test the null hypothesis of no spatial correlation by comparing the value of the statistic in the actual data set, G^{obs}, to the distribution of the statistic under random permutations of the Y_s. The latter distribution is defined as follows. Taking the S units, with values for the variable $Y_1, \ldots, Y_S$, we randomly permute the values $Y_1, \ldots, Y_S$ over the S units. For each of the $S!$ permutations g we re-calculate the Mantel statistic, say G_g. This defines a discrete distribution with $S!$ different values, one for each allocation. The one-sided exact p-value is defined as the fraction of allocations g (out of the set of $S!$ allocations) such that the associated Mantel statistic G_g is less than or equal to the observed Mantel statistic G^{obs}:

$$p = \frac{1}{S!} \sum_{g=1}^{S!} \mathbf{1}_{G^{\text{obs}} \geq G_g}. \tag{3.17}$$

A low value of the p-value suggests rejecting the null hypothesis of no spatial correlation in the variable of interest. In practice the number of allocations is often too large to calculate the exact p-value and so we approximate the p-value by drawing a large number of allocations, and calculating the proportion of statistics less than or equal to the observed Mantel statistic. In the calculations below we use $10,000,000$ draws from the randomization distribution.

We use six different measures of proximity. First, we define the proximity d_{st} as states s and t sharing a border:

$$d_{st}^{B} = \begin{cases} 1 & \text{if } s, t \text{ share a border,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.18}$$

Second, we define d_{st} as an indicator for states s and t belonging to the same census division of states (recall that the US is divided into 9 divisions):

$$d_{st}^{D} = \begin{cases} 1 & \text{if } D_s = D_t, \\ 0 & \text{otherwise.} \end{cases} \tag{3.19}$$

The last four proximity measures are functions of the geographical distance between states s and t:

$$d_{st}^{GD} = -d\left(Z_s, Z_t\right), \qquad \text{and} \quad d_{st}^{\alpha} = \exp\left(-\alpha \cdot d\left(Z_s, Z_t\right)\right) \tag{3.20}$$

where $d(z, z')$ is the distance in miles between two locations z and z', and Z_s is the latitude and longitude of state s, measured as the latitude and longitude of the centroid for each state. We use $\alpha = 0.00138$, $\alpha = 0.00276$, and $\alpha = 0.00693$. For these values the proximity index declines by 50% at distances of 500, 250, and 100 miles.

We calculate the p-values for the Mantel test statistic based on three variables. First, an indicator for having a state minimum wage higher than the federal minimum wage. This indicator takes on the value 1 in nine out of the forty nine states in our sample, with these nine states mainly concentrated in the North East and the West Coast. Second, we calculate the p-values for the average of the logarithm of yearly earnings. Third, we calculate the p-values for the indicator for NE/ENC states. The results for the three variables and six statistics are presented in Table 3.3. All three variables exhibit considerable spatial correlation. Interestingly the results are fairly sensitive to the measure of proximity. From these limited calculations, it appears that sharing a border is a measure of proximity that is sensitive to the type of spatial correlations in the data.

Table 3.3: p-values for Geary's c, one-sided alternatives (10,000,000 draws)

				Distance weights, 50% decline at:		
Proximity	Border	Division	Distance	500 miles	250 miles	100 miles
Minimum wage	< 0.0001	0.0028	0.9960	0.0093	0.0365	0.4307
Log wage	0.0005	0.0239	0.0692	0.0276	0.0298	0.1644
NE/ENC	< 0.0001	< 0.0001	0.0967	0.0877	0.0692	0.0321

3.8 A Small Simulation Study

We carried out a small simulation study to investigate the relevance of the theoretical results from Section 3.6. In all cases the model was

$$Y_i = \alpha + \beta \cdot W_i + \varepsilon_i,$$

with $N = 2,590,190$ observations to mimic our actual data. In our simulations every state has the same number of individuals, and every puma within a given state has the same number of individuals. We considered three distributions for W_i. In all cases W_i varies only at the state level. In the first case $W_i = 1$ for individuals in nine randomly chosen states. In the second case $W_i = 1$ for the the nine minimum wage states. In the third case $W_i = 1$ for the eleven NE/ENC states. The distribution for ε is in all cases Normal with mean zero and covariance matrix Ω. The general specification we consider for Ω is

$$\Omega_{ij}(\mathbf{Z}, \gamma) = \begin{cases} \sigma_D^2 + \sigma_S^2 + \sigma_P^2 + \sigma_\varepsilon^2 & \text{if } i = j, \\ \sigma_D^2 + \sigma_S^2 + \sigma_P^2 & \text{if } i \neq j, P_i = P_j, \\ \sigma_D^2 + \sigma_S^2 & \text{if } P_i \neq P_j, S_i = S_j, \\ \sigma_D^2 & \text{if } S_i \neq S_j, D_i = D_j, \end{cases}$$

We look at two different sets of values for $(\sigma_\varepsilon^2, \sigma_P^2, \sigma_S^2, \sigma_D^2)$, $(0.929, 0, 0.016, 0)$ (only state level correlations, corresponding to the second pair of rows in Table 3.4) and $(0.868, 0.005, 0.005, 0.066)$ (puma, state and division level correlations, corresponding to the fifth pair of rows in Table 3.4).

Given the data, we consider five methods for estimating the variance of the least squares estimator $\hat{\beta}_{\text{ols}}$, and thus for constructing confidence intervals. The first is based on the randomization distribution:

$$\hat{\mathbb{V}}_{CR}(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}) = \frac{M}{M_0 \cdot M_1 \cdot (M-2)} \sum_{m=1}^{M} \hat{\bar{\varepsilon}}_m^2,$$

where $\hat{\bar{\varepsilon}}_m$ is the average value of the residual $\hat{\varepsilon}_i = Y_i - \hat{\alpha}_{\text{ols}} - \hat{\beta}_{\text{ols}} \cdot W_i$ over cluster m. The second, third and fourth variances are model-based:

$$\hat{\mathbb{V}}_M(\hat{\Omega}(\mathbf{Z}), \mathbf{W}, \mathbf{Z}) = \frac{1}{N^2 \cdot \overline{W}^2 \cdot (1-\overline{W})^2} (\overline{W} \quad -1) \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix}' \hat{\Omega}(\mathbf{Z}) \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix} \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix},$$

using different estimates for $\hat{\Omega}(\mathbf{Z})$. First we use an infeasible estimator, namely the true value for $\Omega(\mathbf{Z})$. Second, we specify

$$\Omega_{ij}(\mathbf{Z}, \gamma) = \begin{cases} \sigma_S^2 + \sigma_\varepsilon^2 & \text{if } i = j, \\ \sigma_S^2 & \text{if } i \neq j, S_i = S_j. \end{cases}$$

We estimate σ_P^2 and σ_S^2 using moment-based estimators, and plug that into the expression for the covariance matrix. For the third variance estimator in this set of three variance estimators we specify

$$\Omega_{ij}(\mathbf{Z}, \gamma) = \begin{cases} \sigma_D^2 + \sigma_S^2 + \sigma_P^2 + \sigma_\varepsilon^2 & \text{if } i = j, \\ \sigma_D^2 + \sigma_S^2 + \sigma_P^2 & \text{if } i \neq j, P_i = P_j, \\ \sigma_D^2 + \sigma_S^2 & \text{if } P_i \neq P_j, S_i = S_j, \\ \sigma_D^2 & \text{if } S_i \neq S_j, D_i = D_j, \end{cases}$$

and again use moment-based estimators.

The fifth and last variance estimator allows for more general variance structures within states, but restricts the correlations between individuals in different states to zero. This estimator assumes Ω is block diagonal, with the blocks defined by states, but does not impose constant correlations within the blocks. The estimator for Ω takes the form

$$\hat{\Omega}_{\text{STATA},ij}(\mathbf{Z}) = \begin{cases} \hat{\varepsilon}_i^2 & \text{if } i = j, \\ \hat{\varepsilon}_i \cdot \hat{\varepsilon}_j & \text{if } i \neq j, S_i = S_j, \\ 0 & \text{otherwise,} \end{cases}$$

leading to

$$\hat{\mathbb{V}}_{\text{STATA}} = \frac{1}{N^2 \cdot \overline{W}^2 (1 - \overline{W})^2} \cdot (\overline{W} \quad -1) \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix}' \Omega_{\text{STATA}}(\mathbf{Z}) \begin{pmatrix} \iota_N & \mathbf{W} \end{pmatrix} \begin{pmatrix} \overline{W} \\ -1 \end{pmatrix}.$$

This is the variance estimator implemented in STATA and widely used in empirical work.

In Table 3.4 we report the actual level of tests of the null hypothesis that $\beta = \beta_0$ with a nominal level of 5%. First consider the two columns with random assignment of states to the treatment. In that case all variance estimators lead to tests that perform well, with actual levels between 5.0 and 7.6%. Excluding the STATA variance estimator the actual levels are below 6.5%. The key finding is that even if the correlation pattern involves pumas as well as divisions, variance estimators that ignore the division level correlations do very well.

Table 3.4: Size of t-Tests (in %) Using Different Variance Estimators

Treatment type	Random		Min. Wage		NE/ENC	
Shock type	S	PSD	S	PSD	S	PSD
$\hat{\mathbb{V}}_{\text{CR}}(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z})$	5.6	5.6	5.6	16.2	5.6	26.3
$\hat{\mathbb{V}}_{M}(\Omega(\mathbf{Z}), \mathbf{W}, \mathbf{Z})$	5.0	5.0	5.0	5.0	5.0	5.0
$\hat{\mathbb{V}}_{M}(\Omega(\hat{\sigma}^2_{\epsilon}, \hat{\sigma}^2_{S}), \mathbf{W}, \mathbf{Z})$	6.1	6.1	6.1	17.1	6.1	27.2
$\hat{\mathbb{V}}_{M}(\Omega(\hat{\sigma}^2_{\epsilon}, \hat{\sigma}^2_{P}, \hat{\sigma}^2_{S}, \hat{\sigma}^2_{D}), \mathbf{W}, \mathbf{Z})$	6.1	6.5	5.7	9.0	5.4	13.8
Stata	7.6	7.6	8.5	18.5	7.7	30.4

Notes: Models with only state level correlations (S), and models with puma, state, and division level correlations (PSD). 500,000 draws.

When we do use the minimum wage states as the treatment group the assignment is no longer completely random. If the correlations are within state, all variance estimators still perform well. However, if there are correlations at the division level, now only the variance estimator using the true variance matrix does well. The estimator that estimates the division level correlations does best among the feasible estimators, but, because the data are not informative enough about these correlations to precisely estimate the variance components, even this estimator exhibits substantial size distortions. The same pattern, but even stronger, emerges with the NE/ENC states as the treatment group.

3.9 Conclusion

In empirical studies with individual level outcomes and state level explanatory variables, researchers often calculate standard errors allowing for within-state correlations between individual-level outcomes. In many cases, however, the correlations may extend beyond state boundaries. Here we explore the presence of such correlations, and investigate the implications of their presence for the calculation of standard errors. In theoretical calculations we show that under some conditions, in particular random assignment of regulations, correlations in outcomes between individuals in different states can be ignored. However, state level variables often exhibit considerable spatial correlation, and ignoring out-of-state correlations of the magnitude found in our application may lead to substantial underestimation of standard errors.

In practice we recommend that researchers explicitly explore the spatial correlation structure of both the outcomes as well as the explanatory variables. Statistical tests based on Mantel statistics, with the proximity based on shared borders, or belonging to a common division, are straightforward to calculate and lead to exact p-values. If these test suggest that both outcomes and explanatory variables exhibit substantial spatial correlation, we recommend that one should explicitly account for the spatial correlation by allowing for a more flexible specification than one that only accounts for state level clustering.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010, June). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association 105*(490), 493–505.

Acemoglu, D. (2002). Directed technical change. *The Review of Economic Studies 69*(4), 781–809.

Acemoglu, D. and J. Angrist (2001). How large are human-capital externalities? evidence from compulsory-schooling laws. In *NBER Macroeconomics Annual 2000, Volume 15*, NBER Chapters, pp. 9–74. National Bureau of Economic Research, Inc.

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. Volume 4 of *Handbook of Labor Economics*, Chapter 12, pp. 1043–1171. Elsevier.

Albouy, D. (2008). Are big cities bad places to live? estimating quality of life across metropolitan areas. Working Paper 14472, National Bureau of Economic Research.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Autor, D., L. Katz, and M. Kearney (2008). Trends in us wage inequality: Revising the revisionists. *Review of Economics and Statistics*.

Autor, D. H. and D. Dorn (2012). The growth of low skill service jobs and the polarization of the u.s. labor market. *Massuchusetts Institute of Technology Mimeo*.

Autor, D. H., D. Dorn, and G. H. Hanson (2012). The china syndrome: Local labor market effects of import competition in the united states. NBER Working Papers 18054, National Bureau of Economic Research, Inc.

Barrios, T., G. W. Imbens, R. Diamond, and M. Kolesár (2010). Clustering, Spatial Correlations and Randomization Inference. NBER Working Paper, 15760.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* Books from Upjohn Press. W.E. Upjohn Institute for Employment Research.

Bayer, P., F. Ferreira, and R. McMillan (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy 115*(4), 588–638.

Bayer, P., N. Keohane, and C. Timmins (2009). Migration and hedonic valuation: The case of air quality. *Journal of Environmental Economics and Management 58*(1), 1 – 14.

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*(5), 988 – 1036.

Becker, G. and K. Murphy (2000). *Social Economics: Market Behavior in a Social Environment*. Belknap Series. Belknap Press of Harvard University Press.

Bennett, J. (2010). As detroit faces cutbacks, this one hits a sour note. *The Wall Street Journal*.

Berry, C. R. and E. L. Glaeser (2005). The divergence of human capital levels across cities*. *Papers in Regional Science 84*(3), 407–444.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–90.

Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy 112*(1), pp. 68–105.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics 119*(1), 249–275.

Bester, C. A., T. G. Conley, and C. B. Hansen (2011). Inference with dependent data using cluster covariance estimators. *Journal of Econometrics 165*(2), 137–151.

Bishop, K. (2010). A dynamic model of location choice and hedonic valuation. *Washington University Mimeo*.

Blanchard, O. J. and L. F. Katz (1992). Regional evolutions. *Brookings Papers on Economic Activity 23*(1), 1–76.

Bollinger, C. and B. Hirsch (2006). Match bias from earnings imputation in the current population survey: The case of imperfect matching. *Journal of Labor Economics 24*(3), 483–519.

Brueckner, J. and D. Neumark (2011). Beaches, sunshine, and public-sector pay: theory and evidence on amenities and rent extraction by government workers. Technical report, National Bureau of Economic Research.

Brueckner, J. K. (1987). The structure of urban equilibria: A unified treatment of the muth-mills model. In E. S. Mills (Ed.), *Handbook of Regional and Urban Economics*, Volume 2 of *Handbook of Regional and Urban Economics*, Chapter 20, pp. 821–845. Elsevier.

Card, D. (2009). Immigration and inequality. *American Economic Review 99*(2), 1–21.

Card, D. and T. Lemieux (2001). Can falling supply explain the rising return to college for younger men? a cohort-based analysis. *The Quarterly Journal of Economics 116*(2), pp. 705–746.

Card, D., A. Mas, and J. Rothstein (2008). Tipping and the dynamics of segregation. *The Quarterly Journal of Economics 123*(1), 177–218.

Chetty, R. and A. Szeidl (2007). Consumption commitments and risk preferences. *The Quarterly Journal of Economics 122*(2), 831–877.

Ciccone, A. and G. Peri (2006). Identifying human-capital externalities: Theory with applications. *The Review of Economic Studies 73*(2), 381–412.

Costa, D. and M. Kahn (2000). Power couples: changes in the locational choice of the college educated, 1940–1990. *The Quarterly Journal of Economics 115*(4), 1287–1315.

Cronin, J. (2011). *Reforming Boston Schools, 1930 to the Present: Overcoming Corruption and Racial Segregation*. Palgrave Studies in Urban Education. Palgrave Macmillan.

Diggle, P., P. Heagerty, K.-Y. Liang, and S. L. Zeger (2002). *Analysis of longitudinal data*. Oxford University Press.

Epple, D. and H. Sieg (1999). Estimating equilibrium models of local jurisdictions. *The Journal of Political Economy 107*(4), 645–681.

Epple, D. and A. Zelenitz (1981). The implications of competition among jurisdictions: does tiebout need politics? *The Journal of Political Economy*, 1197–1217.

Fisher, R. A. (1925). *The Design of Experiments* (first ed.). London: Oliver and Boyd.

Freeman, R. (1986). Unionism comes to the public sector. *Journal of Economic Literature*, 41–86.

Gelfand, A. E., P. Diggle, P. Guttorp, and M. Fuentes (2010). *Handbook of Spatial Statistics*. Chapman & Hall.

Gittleman, M. and B. Pierce (2012). Compensation for state and local government workers. *The Journal of Economic Perspectives*, 217–242.

Glaeser, E., J. Scheinkman, and A. Shleifer (1995). Economic growth in a cross-section of cities. *Journal of Monetary Economics 36*(1), 117–143.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer (1992). Growth in cities. *Journal of Political Economy 100*(6), 1126–52.

Goldin, C. and L. Katz (2008). *The Race Between Education and Technology*. Belknap Press for Harvard University Press.

Goldin, C. and L. F. Katz (2007). Long-run changes in the wage stucture: Narrowing, widening, polarizing. *Brookings Papers on Economic Activity*, 135–65.

Greenwald, B. C. (1983). A general analysis of bias in the estimated standard errors of least squares coefficients. *Journal of Econometrics 22*(3), 323–338.

Greenwood, M. J. (1997). Chapter 12 internal migration in developed countries. Volume 1, Part B of *Handbook of Population and Family Economics*, pp. 647 – 720. Elsevier.

Gregory, R. G. and J. Borland (1999). Chapter 53 recent developments in public sector labor markets. Volume 3, Part C of *Handbook of Labor Economics*, pp. 3573 – 3630. Elsevier.

Guerrieri, V., D. Hartley, and E. Hurst (2011). Endogenous gentrification and housing price dynamics. *Booth School of Business Mimeo*.

Gyourko, J., A. Saiz, and A. Summers (2008). A new measure of the local regulatory environment for housing markets: The wharton residential land use regulatory index. *Urban Studies 45*(3), 693–729.

Gyourko, J. and J. Tracy (1991). The structure of local public finance and the quality of life. *Journal of Political Economy*, 774–806.

Handbury, J. (2012). Are poor cities cheap for everyone? non-homotheticity and the cost of living across u.s. cities. *Columbia University Mimeo*.

Hansen, C. B. (2007). Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics 140*(2), 670–694.

Ibragimov, R. and U. Muller (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics 28*(4), 453–468.

Kahn, M. E. (1999). The silver lining of rust belt manufacturing decline. *Journal of Urban Economics 46*(3), 360–376.

Katz, L. F. and D. H. Autor (1999). Chapter 26 changes in the wage structure and earnings inequality. Volume 3, Part A of *Handbook of Labor Economics*, pp. 1463 – 1555. Elsevier.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Kennan, J. and J. R. Walker (2011). The effect of expected income on individual migration decisions. *Econometrica 79*(1), 211–251.

Krueger, A. (1988). Are public sector workers paid more than their alternative wage? evidence from longitudinal data and job queues. In *When public sector workers unionize*, pp. 217–242. University of Chicago Press.

Liang, K.-Y. and S. L. Zeger (1986, January). Longitudinal data analysis for generalized linear models. *Biometrika 73*(1), 13–22.

Mantel, N. (1967, February). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer research 27*(2), 209–20.

Martelle, S. (2012). *Detroit: A Biography.* Chicago Review Press.

McFadden, D. (1973). Conditional logit analyisis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. pp. 105–142. Academic Press.

Mirel, J. (1999). *The rise and fall of an urban school system : Detroit, 1907-81.* Ann Arbor : University of Michigan Press.

Moretti, E. (2004a). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics 121*(1-2), 175–212.

Moretti, E. (2004b). Human capital externalities in cities. *Handbook of regional and urban economics 4*, 2243–2291.

Moretti, E. (2004c). Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review 94*(3), 656–690.

Moretti, E. (2011a). Local labor markets. Volume 4 of *Handbook of Labor Economics*, Chapter 14, pp. 1237–1313. Elsevier.

Moretti, E. (2011b). Real wage inequality. *University of California Berkeley Mimeo*.

Moretti, E. (2012). *The New Geography of Jobs*. Houghton Mifflin Harcourt.

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics 32*(3), 385–397.

Nevo, A. (2010). Empirical models of consumer behavior. NBER Working Papers 16511, National Bureau of Economic Research, Inc.

Neyman, J. (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on principles. Section 9. *Statistical Science 5*(4), 465–480.

Notowidigdo, M. J. (2011). The incidence of local labor demand shocks. *Booth School of Business Mimeo*.

Ottaviano, G. I. P. and G. Peri (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association 10*(1), 152–197.

Pakes, A. and P. McGuire (2001). Stochastic algorithms, symmetric markov perfect equilibrium, and the 'curse' of dimensionality. *Econometrica 69*(5), 1261–81.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy 90*(6), pp. 1257–1278.

Rosen, S. (1979). Wages-based indexes of urban quality of life. In P. Mieszkowski and M. Straszheim (Eds.), *Current Issues in Urban Economics*. John Hopkins Univ. Press.

Rosenbaum, P. R. (2002). *Observational Studies*. New York: Springer-Verlag.

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek. (2010). *Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]*..

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics 125*(3), pp. 1253–1296.

Saks, R. E. (2008). Job creation and housing construction: Constraints on metropolitan area employment growth. *Journal of Urban Economics 64*(1), 178–195.

Schabenberger, O. and C. A. Gotway (2004). *Statistical Methods for Spatial Data Analysis* (first ed.). Chapman and Hall/CRC.

Shapiro, J. (2006). Smart cities: Quality of life, productivity, and the growth effects of human capital. *The Review of Economics and Statistics 88*(2), 324–335.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics 20*(4), 518–29.

Topel, R. H. (1986). Local labor markets. *Journal of Political Economy 94*(3), pp. S111–S143.

Winerip, M. (2011). For detroit schools, mixed picture on reforms. *The New York Times*.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

A Chapter 1 Appendix

A.1 Data Appendix

Variable	Source	Sample	Notes
Metropolitain Statistical Area	US Census 1980, 1990, 2000	All MSAs identified across all 3 censuses. Rural areas of each state are included as additional geopgraphic units.	MSAs identified in some, but not all of the censuses are included in rural areas of each state.
Local High Skill and Low Skill Wages	US Census 1980, 1990, 2000	All individuals ages 25-55 working at least 35 hours per week and 48 weeks per year.	Local wages in each MSA are averages of workers for each skill level living in each city. High skill worker is defined as a worker with at least a 4 year college degree. All other workers are considered low skill.
Local Housing Rent	US Census 1980, 1990, 2000	All households where the head-of-household is between the ages of 25 and 55 and works at least 35 hours per week and 48 weeks per year.	Rental rates are measured as the gross rent, which includes both the housing rent and the cost of utilities. Rents are imputed for households which own their home. Imputed rents are converted from housing values using a discount rate of 7.85 percent (Peiser and Smith 1985), to which electricity and gas utility costs are added.
Local College Employment Ratio	US Census 1980, 1990, 2000	All employed workers between the ages of 25 and 55.	College employment ratio is defined as the ratio of number of employed workers in the city with a 4 year college degree to the number of employed lower skill workers living in the city.
Worker's Race	US Census 1980, 1990, 2000	All households where the head-of-household is between the ages of 25 and 55 with positive wage earnings.	A household is classifed as black if the head of household reports his race as black.
Worker's Immigrant Status	US Census 1980, 1990, 2000	All households where the head-of-household is between the ages of 25 and 55 with positive wage earnings.	A household is classifed as an immigrant if the head-of-household was born outside of the United States.

Variable	Source	Sample	Notes
Property Crimes and Violent Crimes per 1000 Residents	FBI Uniform Crime Reports 1980, 2000	All non-rural MSAs which the FBI data covers.	
EPA Air Quality Index	Envionmental Protection Agency	All non-rural MSAs which the EPA reports data on.	

A.2 Estimation Appendix

A.2.1 Labor Demand

The parameters to be estimated in cities' labor demand curves are $(\rho, \gamma_H, \gamma_L)$. ρ governs the slope of firms' labor demand curves and γ_H and γ_L measure endogenous productivity spillovers driven by cities' college employment ratios.

Throughout the appendix, I use a change of variables defining the high and low skill exogenous productivities:

$$\begin{aligned} \ln\left(\theta_{jt}^{1/\alpha}\left(1-\lambda_{jt}\right)\right) &= \varepsilon_{jt}^{H} \\ \ln\left(\theta_{jt}^{1/\alpha}\lambda_{jt}\right) &= \varepsilon_{jt}^{L}. \end{aligned}$$

These variable definitions are useful for showing how N_{jt} is computed for estimation, but have no economic impact on the model. City $j's$ inverse labor demand is:

$$\begin{aligned} \Delta w_{jt}^{H} &= \Delta c_t + \Delta \ln\left(\theta_{jt}^{1/\alpha}\left(1-\lambda_{jt}\right)\right) + (1-\rho)\Delta \ln N_{jt} + (\rho-1)\Delta \ln H_{jt} + \gamma_H \Delta \ln\left(\frac{H_{jt}}{L_{jt}}\right) \\ \Delta w_{jt}^{L} &= \Delta c_t + \Delta \ln\left(\theta_{jt}^{1/\alpha}\lambda_{jt}\right) + (1-\rho)\Delta \ln N_{jt} + (\rho-1)\Delta \ln L_{jt} + \gamma_L \Delta \ln\left(\frac{H_{jt}}{L_{jt}}\right) \\ N_{jt} &= \left(\lambda_{jt}\left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma_L} L_{jt}^{\rho} + (1-\lambda_{jt})\left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma_H} H_{jt}^{\rho}\right)^{\frac{1}{\rho}}. \end{aligned} \tag{A.1}$$

A contributor to local productivity change is changes in the productivity levels of the industries located within each city. I measure these observable local productivity shocks by interacting cross-sectional differences in industrial employment composition with national changes in industry wage levels, separately for high and low skill workers. I refer to these as Bartik shocks, as represented by ΔB_{jt}^{H} and ΔB_{jt}^{L}. The aggregate Bartik shock, which weights the two skill group specific Bartik shock by the share of a city's employment made up of each skill group is represented by ΔB_{jt}^{all}. Industry classifications are those used in the 1980, 1990 and 2000 Census, which is very close to 3 digit SIC code classifications. See equations (1.18) and (1.21) in the main text for the exact definition.

Since I do not observe all other contributors to local productivity changes, I rewrite the labor demand equations by explicitly defining $\Delta\tilde{\varepsilon}_{jt}^{H}$ and $\Delta\tilde{\varepsilon}_{jt}^{L}$, which represent all productivity changes of high and low skill workers, respectively, unaccounted for by the Bartik shocks $\Delta B_{jt}^{H}, \Delta B_{jt}^{L}$:

$$\Delta\tilde{\varepsilon}_{jt}^{H} = \Delta w_{jt}^{H} - \begin{pmatrix} \Delta c_t + \beta_B^H \Delta B_{jt}^H + (1-\rho)\Delta \ln N_{jt} \\ +(\rho-1)\Delta \ln H_{jt} + \gamma_H \Delta \ln\left(\frac{H_{jt}}{L_{jt}}\right) \end{pmatrix} \tag{A.2}$$

$$\Delta\tilde{\varepsilon}_{jt}^{L} = \Delta w_{jt}^{L} - \begin{pmatrix} \Delta c_t + \beta_B^L \Delta B_{jt}^L + (1-\rho)\Delta \ln N_{jt} \\ +(\rho-1)\Delta \ln L_{jt} + \gamma_L \Delta \ln\left(\frac{H_{jt}}{L_{jt}}\right) \end{pmatrix}. \tag{A.3}$$

While the direct impact of Bartik shocks lead to shifts in local labor demand curves, the interaction of the Bartik shocks with cities' housing markets creates variation that can identify the labor demand parameters. Heterogeneity in housing supply elasticity will lead to differences in population changes in response to a given Bartik shock. Land unavailable for housing development due to geographic features, x_j^{geo}, as well as land-use regulation, x_j^{reg}, impact local housing supply elasticity. The interaction of these housing supply elasticity measures with local Bartik shocks can be used as instruments for quantities of labor within the city.

This leads to the following moments:

- Bartik Shocks uncorrelated with unobserved productivity changes:[50]

[50] These moments only identify the coefficients which measure the Bartik shocks direct impact on wages. They do not directly identify any of the structural labor demand parameters.

$$
\begin{aligned}
E\left(\Delta\tilde{\varepsilon}_{jt}^{H}\Delta B_{jt}^{H}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{L}\Delta B_{jt}^{L}\right) &= 0
\end{aligned}
$$

- Bartik shocks interacted with housing supply elasticity measures are uncorrelated with unobserved productivity changes:

$$
\begin{aligned}
E\left(\Delta\tilde{\varepsilon}_{jt}^{H}\Delta B_{jt}^{H}x_{j}^{geo}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{H}\Delta B_{jt}^{H}x_{j}^{reg}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{L}\Delta B_{jt}^{L}x_{j}^{geo}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{L}\Delta B_{jt}^{L}x_{j}^{reg}\right) &= 0
\end{aligned}
$$

- The aggregate Bartik shock and interactions with housing supply elasticity are uncorrelated with high and low skill unobserved productivity changes:

$$
\begin{aligned}
E\left(\Delta\tilde{\varepsilon}_{jt}^{L}\Delta B_{jt}^{all}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{H}\Delta B_{jt}^{all}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{L}\Delta B_{jt}^{all}x_{j}^{geo}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{L}\Delta B_{jt}^{all}x_{j}^{reg}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{H}\Delta B_{jt}^{all}x_{j}^{geo}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{H}\Delta B_{jt}^{all}x_{j}^{reg}\right) &= 0.
\end{aligned}
$$

- High skill Bartik shock and interactions with housing supply elasticity are uncorrelated with low skill unobserved productivity changes, and vice versa:

$$
\begin{aligned}
E\left(\Delta\tilde{\varepsilon}_{jt}^{L}\Delta B_{jt}^{H}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{H}\Delta B_{jt}^{L}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{L}\Delta B_{jt}^{H}x_{j}^{geo}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{L}\Delta B_{jt}^{H}x_{j}^{reg}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{H}\Delta B_{jt}^{L}x_{j}^{geo}\right) &= 0 \\
E\left(\Delta\tilde{\varepsilon}_{jt}^{H}\Delta B_{jt}^{L}x_{j}^{reg}\right) &= 0.
\end{aligned}
$$

I also include decade fixed effects.

To compute the moments, I need to be able to compute the unobserved changes in cities' productivities $\left(\Delta\tilde{\varepsilon}_{jt}^{H}, \Delta\tilde{\varepsilon}_{jt}^{L}\right)$, given a guess of the labor demand parameters $(\rho, \gamma_H, \gamma_L)$, and the data $\left(w_{jt}^{\text{col}}, w_{jt}^{hs}, H_{jt}, L_{jt}\right)$, as seen in equations (1.19) and (1.20) .$\Delta\tilde{\varepsilon}_{jt}^{H}$ and $\Delta\tilde{\varepsilon}_{jt}^{L}$ depend on the CES aggregate labor quantity in the city, N_{jt}. N_{jt} is a function of λ_{jt}, which measures the relative productivity of high and low skill workers, and it is not directly observable. However, the CES functional form allows one to solve for λ_{jt} in terms of observable data. Taking the ratio of high skill wages to low skill wages in city j gives:

$$
\frac{w_{jt}^{H}}{w_{jt}^{L}} = \frac{\theta_{jt}\alpha N_{jt}^{1-\rho}\left(\frac{\theta_{jt}(1-\alpha)}{r}\right)^{\frac{1-\alpha}{\alpha}}\left(1-\lambda_{jt}\right)H_{jt}^{\rho-1}\left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma_H}}{\theta_{jt}\alpha N_{jt}^{1-\rho}\left(\frac{\theta_{jt}(1-\alpha)}{r}\right)^{\frac{1-\alpha}{\alpha}}\lambda_{jt}L_{jt}^{\rho-1}\left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma_L}}.
$$

Solving for λ_{jt} :

$$
\lambda_{jt} = \frac{w_{jt}^{L}H_{jt}^{\rho-1+\gamma_H-\gamma_L}}{w_{jt}^{H}L_{jt}^{\rho-1+\gamma_H-\gamma_L}+w_{jt}^{L}H_{jt}^{\rho-1+\gamma_H-\gamma_L}}.
$$

Plugging in this expression for λ_{jt} into equation $(A.1)$,the CES aggregate labor quantity in the city, and rearranging gives:

$$
N_{jt} = \left(\frac{w_{jt}^{L}H_{jt}^{\rho-1+\gamma_H}L_{jt}^{\rho-\gamma_L}+w_{jt}^{H}L_{jt}^{\rho-1-\gamma_L}H_{jt}^{\rho+\gamma_H}}{w_{jt}^{H}L_{jt}^{\rho-1+\gamma_H-\gamma_L}+w_{jt}^{L}H_{jt}^{\rho-1+\gamma_H-\gamma_L}}\right)^{1/\rho}. \tag{A.4}
$$

Thus, given a guess of the parameter values $(\rho, \gamma_H, \gamma_L)$ and data $\left(w_{jt}^H, w_{jt}^L, H_{jt}, L_{jt}\right)$ from each decade 1980, 1990, and 2000, one can compute $\Delta\tilde{\varepsilon}_{jt}^H$ and $\Delta\tilde{\varepsilon}_{jt}^L$ using equations $(A.2)$, $(A.3)$, and $(A.4)$. These residuals can the be used to form the labor demand moments:

$$
\begin{aligned}
m^{H\_labor}(\rho, \gamma_H, \gamma_L) &= \frac{1}{JT}\sum_{jt}\left(\Delta\tilde{\varepsilon}_{jt}^H \Delta Z_{jt}\right) \\
m^{L\_labor}(\rho, \gamma_H, \gamma_L) &= \frac{1}{JT}\sum_{jt}\left(\Delta\tilde{\varepsilon}_{jt}^L \Delta Z_{jt}\right) \\
\Delta Z_{jt} &\in \left\{\begin{array}{c} \Delta B_{jt}^{all}, \Delta B_{jt}^H, \Delta B_{jt}^L \\ \Delta B_{jt}^{all} x_j^{reg}, \Delta B_{jt}^H x_j^{reg}, \Delta B_{jt}^L x_j^{reg} \\ \Delta B_{jt}^{all} x_j^{geo}, \Delta B_{jt}^H x_j^{geo}, \Delta B_{jt}^L x_j^{geo} \end{array}\right\}.
\end{aligned}
$$

A.2.2 Housing Supply

To identify the elasticity of housing supply, one needs variation in a city's total population which is unrelated to changes in unobserved construction costs. Defining changes in unobserved construction costs as $\Delta\varepsilon_{jt}^{CC}$, I rewrite the housing supply curve as:

$$
\begin{aligned}
\Delta\varepsilon_{jt}^{CC} &= \Delta \ln (CC_{jt}) \\
&= \Delta r_{jt} - \left(\Delta \ln(\iota_t) + \left(\gamma + \gamma^{geo}\exp\left(x_j^{geo}\right) + \gamma^{reg}\exp\left(x_j^{reg}\right)\right)\Delta \ln (H_{jt} + L_{jt})\right).
\end{aligned}
$$

I instrument for housing demand changes using the Bartik shocks discussed above. This leads to moments:

$$
\begin{aligned}
E\left(\Delta\varepsilon_{jt}^{CC}\Delta Z_{jt}\right) &= 0 \\
\Delta Z_{jt} &\in \left\{\begin{array}{c} \Delta B_{jt}^{all}, \Delta B_{jt}^H, \Delta B_{jt}^L \\ \Delta B_{jt}^{all} x_j^{reg}, \Delta B_{jt}^H x_j^{reg}, \Delta B_{jt}^L x_j^{reg} \\ \Delta B_{jt}^{all} x_j^{geo}, \Delta B_{jt}^H x_j^{geo}, \Delta B_{jt}^L x_j^{geo} \end{array}\right\}.
\end{aligned}
$$

The sample analog for each of these moments is measured by:

$$m^{\text{housing}}\left(\gamma,\gamma^{geo},\gamma^{reg}\right)=\frac{1}{JT}\sum_{jt}\left(\Delta\varepsilon_{jt}^{CC}\Delta Z_{jt}\right).$$

A.2.3 Labor Supply

Recall that the utility of worker i in city j at time t is a function of the city's wage, rent, amenities offered, the worker's demographic characteristics, z_i, and his state and Census division of birth, st_i, div_i:

$$\begin{aligned} V_{ijt} &= \delta_{jt}^{z_i}+\beta^{st}z_i st_i x_j^{st}+\beta^{\text{div}}z_i\,\text{div}_i\,x_j^{\text{div}}+\varepsilon_{ijt} \\ \delta_{jt}^{z_i} &= \beta^{w}z_i w_{jt}^{edu}-\beta^{r}z_i r_{jt}+\beta^{A}z_i x_{jt}^{A}+\beta^{\text{col}}z_i\frac{H_{jt}}{L_{jt}}. \end{aligned}$$

The utility a worker with demographics z gains from any city can be written as the sum of the common utility value of that city for all workers with demographics z,δ_{jt}^{z}, the utility value of whether the city is located close to the workers' state of birth, $\beta^{st}z_i st_i x_j^{st}+\beta^{\text{div}}z_i\,\text{div}_i\,x_j^{\text{div}}$, and a worker specific idiosyncratic taste for the city ε_{ijt}.

To estimate workers' preferences for cities, I use a two-step estimator similar to the methods used by Berry, Levinsohn, and Pakes (2004) to estimate households' demand for automobiles. In the first step, I use a maximum likelihood estimator, where I treat the common utility value of each city for each demographic group, δ_{jt}^{z}, as a parameter to be estimated. The distributional assumption on the iid taste preferences for cities (ε_{ijt}) allow me to form a likelihood function to maximize.[51] Since ε_{ijt} is distributed Type I extreme

[51] While one often worries about bias in estimating fixed effects using a non-linear objective function, such as maximum likelihood, I have a very large sample of individual level data (over 2 million observations for each decade), relative to about 1000 estimated fixed effects per decade. The asymptotics assumes the number of individuals goes to infinity, either holding the number of cities fixed or at a faster rate than the number of cities goes to infinity.

value, the likelihood function is:

$$LLH(\delta^{z}_{jt}, \beta^{st}, \beta^{\text{div}}) = \sum_{i=1}^{n} \log\left(\frac{\exp(\delta^{z_i}_{jt} + \beta^{st} z_i st_i x^{st}_j + \beta^{\text{div}} z_i \,\text{div}_i\, x^{\text{div}}_j) * 1(j_i = j)}{\sum_k^J \exp(\delta^{z_i}_{kt} + \beta^{st} z_i st_i x^{st}_k + \beta^{\text{div}} z_i \,\text{div}_i\, x^{\text{div}}_k)}\right),$$

where $1(j_i = j)$ is an indicator function of whether worker i did, in fact, choose to live in city j and n is the total number of workers in the data. This setup allows the econometrician to only need to observe the location choices of workers and the workers' demographics, but not cities' characteristics, to estimate how different types of workers value different cities.[52] The maximum likelihood estimation identifies a mean utility level for each city, for each demographic group, for each decade of data. This amounts to estimating 1608 mean utility parameters (268 cities x 6 demographic groups) in each decade. A desirable property of the conditional logit model is that the likelihood function is globally concave. Optimizing the likelihood function to find the global maximum is relatively simple, since there are no local minima.

A.2.4 Estimation of wage, local price, and amenity preferences

The second step of estimation decomposes the mean utility estimates into how workers value wages, rents and amenities. For each demographic group, the maximum likelihood estimation produced mean utility estimates in 1980, 1990, and 2000 for each city. I first difference the mean utility estimates for a given city, for each demographic group across decades. The change in a city's mean utility across a decade for workers with demographics z is driven by changes in wages, rents, and amenities:

[52] An issue I do not explicitly consider is the presence of unobserved household characteristics which impact preferences over cities. Many discrete choice models incorporate this by adding random coefficients on some product characteristics in the utility function. These can be useful for estimating realistic substitution elasticities across products. The random coefficients would allow the characteristics of the city a worker chose to live in to inform what types of city characteristics that worker is likely to prefer. This break the undesirable independence of irrelevent alternatives (IIA) property of conditional logit models. However, since I do not observe any information about workers' preferences over cities beyond their first choice cities, any identification of the variance of random coefficient parameters would be solely due to functional form restrictions.

$$\Delta\delta^z_{jt} = \beta^w z\Delta w^{edu}_{jt} - \beta^r z\Delta r_{jt} + \beta^A z\Delta x^A_{jt} + \beta^{\text{col}} z\Delta\left(\frac{H_{jt}}{L_{jt}}\right). \tag{A.5}$$

Changes in cities' wages, rents, and college employment ratios are observed in the data. However, other amenity changes are unobserved by the econometrician. Define $\Delta\xi^z_{jt}$ as the change in utility value of city $j's$ amenities unobserved to the econometrician across decades for workers with demographics $z.$ Formally, $\Delta\xi^z_{jt}$ is defined by:

$$\Delta\xi^z_{jt} = \beta^A z\Delta x^A_{jt}.$$

Plugging this into equation $(A.5)$:

$$\Delta\delta^z_{jt} = \beta^w z\Delta w^{edu}_{jt} - \beta^r z\Delta r_{jt} + \beta^{\text{col}} z\Delta\left(\frac{H_{jt}}{L_{jt}}\right) + \Delta\xi^z_{jt}.$$

If the changes in unobserved amenities across decades were uncorrelated with the changes in local wages, rents, and endogenous amenities, then the equation above could be estimated by regressing the changes in mean utilities on the changes in local wages, rents, and college concentration. However, since wages, rents, and the college employment ratios are simultaneously determined with workers' city choices, exogenous changes in local amenities will be correlated wages, rent, and college employment ratio changes.

To identify workers' preferences for cities' wages, rents, and endogenous amenities, Bartik productivity shocks and their interaction with housing supply characteristics are used as instruments for these endogenous outcomes. Since Bartik productivity shocks are driven by national changes in industrial productivity, they are unrelated to local exogenous amenity changes. While local housing supply elasticity characteristics, such as coastal proximity and mountains, likely are amenities of a city, they do not change over time. The identifying assumption is that housing supply elasticity characteristics are independent of changes in local exogenous amenities.

Identifying workers' preferences for endogenous amenities is equivalent to identifying workers' preferences for a city's college employment ratio. The lefthand side of the estimating equation is equal to a demographic group's mean utility level for each city, which is identified off of population differences across cities. A city has a high mean utility if a large share of all workers nationwide, within a demographic group, choose to live there. The populations of all demographic groups within a city also determine a city's college employment ratio. While population is both on the left and right hand side of the estimating equation, variation in cities' college employment ratios is not solely driven by any given demographic group. For example, variation in cities' low skill populations impacts the college employment ratio, holding the high skill local population fixed. Variation in cities' college employment ratios driven by Bartik shocks and their interactions to housing supply elasticity is driven by how these local shocks differentially impact high and low skill workers. Thus, the migration response of a given demographic group to the set of Bartik productivity shocks and their interactions with housing supply elasticity depends on how these shocks jointly influence local wages, rents, and college employment ratios. Intuitively, workers' preferences for a city's college employment ratio is identified off how other demographic groups respond to the local Bartik shocks.

This leads to the following moment restrictions:

$$\begin{aligned} E\left(\Delta\xi^{z}_{jt}\Delta Z_{jt}\right) &= 0 \\ \Delta Z_{jt} &\in \left\{ \begin{array}{c} \Delta B^{all}_{jt}, \Delta B^{H}_{jt}, \Delta B^{L}_{jt} \\ \Delta B^{all}_{jt} x^{reg}_{j}, \Delta B^{H}_{jt} x^{reg}_{j}, \Delta B^{L}_{jt} x^{reg}_{j} \\ \Delta B^{all}_{jt} x^{geo}_{j}, \Delta B^{H}_{jt} x^{geo}_{j}, \Delta B^{L}_{jt} x^{geo}_{j} \end{array} \right\}. \end{aligned}$$

Decade fixed effects are also included.

The sample analog for each of these moments is measured by:

$$m_z^{\text{util}}\left(\beta^w, \beta^r, \beta^{\text{col}}\right) = \frac{1}{JT}\sum_{jt}\left(\Delta\xi_{jt}^z \Delta Z_{jt}\right).$$

The labor demand, housing supply, and labor supply moments are all used together to jointly estimate all parameters using a 2-step GMM estimator. Standard errors are clustered by MSA in all estimating equations. Standard errors for the worker preference estimates account for the fact that the mean utilities are estimated. However, since the first step is estimated off such a large dataset, this additional source of error does not have a meaningful effect on the standard errors of worker preference estimates.

A.3 Dynamic Adjustment & Equilibrium Stability

Since a city's college share endogenously determines a city's productivity and amenity level, there could be many possible equilibrium outcomes. The equilibrium selected will depend on initial conditions and the dynamic adjustment mechanism of wages, rents, amenities, and population. Under the true dynamic adjustment process, the equilibrium selected in the data is likely to be stable. While I have not modeled or estimated the determinants of the dynamic adjustment process, I will assume a simple dynamic adjustment process to assess whether the observed equilibrium outcomes in the data are stable, given the model's parameter estimates.

Consider a dynamic adjustment mechanism where in each year t $\alpha\%$ of the US population considers moving to a new city, while $(1-\alpha)\%$ remain fixed in their current location. Let $g\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix}$ be a function that maps the population distribution of migrants across cities in year $t-1$, $\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix}$, to the city choices of the migrate population in year t. Total population in each city in year t is then the combination of the migrant's city choices and the population distribution of non-migrants from the previous year, as represented by the function f:

$$\begin{pmatrix} H_t \\ L_t \end{pmatrix} = f\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix}, \tag{A.6}$$

$$f\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix} = \alpha g\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix} + (1-\alpha) * \begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix}. \tag{A.7}$$

$$H_t = (H_{1t}, H_{2t}, ...H_{Jt}),$$

$$L_t = (L_{1t}, L_{2t}, ...L_{Jt}).$$

The migrants choose where to live in year t based off of the wages, rents, and amenities observed in year $t-1$. Workers do not take into account the impact of their migration or the expected migration of others on cities' wages, rents, and amenities in the following period. The wages, rents, and amenities in each city in year $t-1$ are the city's labor demand, housing supply, and amenity levels given the population in year $t-1$. Thus, I am assuming labor demand, housing supply, and amenity supply fully respond to population changes from the previous period. An alternative approach would be to allow wages, rents, and amenities to sluggishly adjust to population changes each period. Each migrant moves to the city which offers him the highest utility. Formally, g is parameterized as:

$$
\begin{aligned}
g\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix} &= \left(g_1 \begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix}, g_2 \begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix}, ...g_J \begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix} \right) \\
g_j\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix} &= \begin{pmatrix} \sum_{i\in\mathcal{H}} \frac{\exp\left(u_{ij}\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix}\right)}{\sum_k \exp(u_{ik}\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix})} \\ \sum_{i\in\mathcal{L}} \frac{\exp\left(u_{ij}\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix}\right)}{\sum_k \exp(u_{ik}\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix})} \end{pmatrix}
\end{aligned}
$$

$\mathcal{H}$ and $\mathcal{L}$ are the sets of high and low skill workers nationwide, respectively. The utility each migrant worker would receive from each city in period $t-1$ is:

$$
\begin{aligned}
V_{ijt} &= u_{ijt} + \varepsilon_{ijt}, \\
u_{ij}\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix} &= \beta_i^w w_j^{edu}\left(H_{jt-1}, L_{jt-1}\right) + \beta_i^r r_j\left(H_{jt-1}, L_{jt-1}\right) + \beta_i^{\text{col}} \frac{H_{jt-1}}{L_{jt-1}} + \xi_{ij} \\
&\quad + \beta_i^{st} st_i x_j^{st} + \beta_i^{\text{div}} \operatorname{div}_i x_j^{\text{div}},
\end{aligned}
$$

where $\left(\beta_i^w, \beta_i^r, \beta_i^{\text{col}}, \xi_{ij}, \beta_i^{st}, \beta_i^{\text{div}}\right)$ are the migrant's preferences as determined by his demographics.

The wages and rents workers expect to receive in each city are defined as:

$$
\begin{aligned}
w_j^H\left(H_{jt-1}, L_{jt-1}\right) &= c+(1-\rho)\ln N_j\left(H_{jt-1}, L_{jt-1}\right)+(\rho-1)\ln H_{jt-1}+\gamma_H \ln\left(\frac{H_{jt-1}}{L_{jt-1}}\right)+\varepsilon_j^H, \\
w_j^L\left(H_{jt-1}, L_{jt-1}\right) &= c+(1-\rho)\ln N_j\left(H_{jt-1}, L_{jt-1}\right)+(\rho-1)\ln L_{jt-1}+\gamma_L \ln\left(\frac{H_{jt-1}}{L_{jt-1}}\right)+\varepsilon_j^L, \\
r_j\left(H_{jt-1}, L_{jt-1}\right) &= a+\left(\gamma+\gamma^{geo}\exp\left(x_j^{geo}\right)+\gamma^{reg}\exp\left(x_j^{reg}\right)\right)\ln\left(H_{jt-1}+L_{jt-1}\right)+\varepsilon_j^{CC}.
\end{aligned}
$$

These represent labor demands and housing supplies which fully adjust to the population levels in year $t-1$.

Migrant workers gain utility in period $t-1$ based on the city they choose to migrate to. Worker are myopic under this dynamic because they do not take into account that they will only have a $\alpha\%$ chance of being able to migrate again the following period. They migrate based on the current period's wages, rents, and amenities. However, other workers may also find the same city appealing, and the aggregate migration responses of all workers will impact the wages, rents, and amenities in the following period. I do not argue that this dynamic adjustment process mimics the actual migration dynamics of workers in reality, but it is a simple dynamic and useful for considering the stability of equilibria. Blanchard & Katz (1992) and other work finds that migration only partially responds to utility differences across states each year. This fact could motivate using this dampened migration dynamic.

Fixed points of this dynamic adjustment mechanism are equilibria of the model:

$$
\begin{pmatrix} H_t \\ L_t \end{pmatrix} = f\begin{pmatrix} H_t \\ L_t \end{pmatrix}.
$$

To calculate f, I plug in the model's parameter estimates of labor demand, labor supply, housing supply, and the inferred levels of exogenous productivity, amenities, and construction costs. By construction, the population distribution of workers observed in the data is in equilibrium. To test for stability, I hold the population distribution fixed at that observed in the data and test whether the magnitude of all eigenvalues of the Jacobian of f are less than 1. When $0 \leq \alpha \leq 0.07$, the equilibrium observed in the data in 1980, 1990, and 2000 are all stable. However, for $\alpha > 0.07$, the Jacobian of f has eigenvalues below -1 in some years. While this model can potentially have a large set of possible equilibria, the equilibria observed in the data are stable under this assumed migration dynamic.

If a large fraction of the population can migrate each period ($\alpha > 0.07$) and workers do not take into account the expected migration of others, the dynamic adjustment process can cycle, where cities oscillate between being high college share, desirable cities, to low college share, undesirable cities.

To build intuition for why the dynamic can oscillate, I simulate this dynamic adjustment process for a simple two city case. I simplify the model to a two skill group model, ignoring preference heterogeneity of immigrants and Blacks. I also ignore workers' preferences to live their state of birth.

I create two cities with identical housing supply curves and exogenous amenity levels. I allow city 1 to exogenously be more productive than city 2 for college workers, but both cities are equally productive for non-college workers. The cities' exogenous productivities are set to:

$$\begin{aligned} \varepsilon_1^H &= 1.5 \\ \varepsilon_2^L &= 1 \\ \varepsilon_1^L &= \varepsilon_2^L = .5 \end{aligned}$$

Housing supply elasticities are set to the estimated average city's housing supply elasticity:

$$\gamma_1 = \gamma_2 = 0.16.$$

The cities' labor demand and housing supply curves are:

$$\begin{aligned}
w_j^H(H_{jt-1}, L_{jt-1}) &= (1-\rho)\ln N_j(H_{jt-1}, L_{jt-1}) + (\rho-1)\ln H_{jt-1} + \gamma_H \ln\left(\frac{H_{jt-1}}{L_{jt-1}}\right) + \varepsilon_j^H, \\
w_j^L(H_{jt-1}, L_{jt-1}) &= (1-\rho)\ln N_j(H_{jt-1}, L_{jt-1}) + (\rho-1)\ln L_{jt-1} + \gamma_L \ln\left(\frac{H_{jt-1}}{L_{jt-1}}\right) + \varepsilon_j^L, \\
r_j(H_{jt-1}, L_{jt-1}) &= \gamma_j \ln(H_{jt-1} + L_{jt-1}).
\end{aligned}$$

$(\rho, \gamma_H, \gamma_L)$, are set the estimated values $(0.563, 0.555, 0.020)$,as reported in Table 1.3. I set exogenous amenities equal in both cities, and I normalize them to 0. Workers' (non-random) utility value for each city is then:

$$u_{ij}\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix} = \beta_i^w w_j^{edu}(H_{jt-1}, L_{jt-1}) + \beta_i^r r_j(H_{jt-1}, L_{jt-1}) + \beta_i^{\text{col}} \frac{H_{jt-1}}{L_{jt-1}}.$$

Plugging in the model estimated worker preferences from Table 1.3 gives:

$$\begin{aligned}
u_{Hj}\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix} &= 1.41 w_j^H(H_{jt-1}, L_{jt-1}) - .95 r_j(H_{jt-1}, L_{jt-1}) + 3.38\frac{H_{jt-1}}{L_{jt-1}}, \\
u_{Lj}\begin{pmatrix} H_{t-1} \\ L_{t-1} \end{pmatrix} &= 4.12 w_j^L(H_{jt-1}, L_{jt-1}) - 2.80 r_j(H_{jt-1}, L_{jt-1}) + 1.36\frac{H_{jt-1}}{L_{jt-1}}.
\end{aligned}$$

I set the mass of college workers in the nation equal to one and the mass of non-college workers equal to 3. This gives a nationwide college share of 0.25, which is the share of college workers observed in the data from 1990.

I simulate this dynamic starting the population distribution equal across both cities:

$$\begin{aligned}
H_{1t=0} &= H_{2t=0} = .5 \\
L_{1t=0} &= L_{2t=0} = 1.5.
\end{aligned}$$

I simulate the dynamic adjustment varying the share of workers which can migrate period, as measured by α in equation $(A.7)$. I plot the dynamic adjustments in Figure A.1. The figure shows that for $\alpha \in \{0.05, 0.25, 0.5\}$, the dynamic adjustment process converges to an equilibrium. However, for $\alpha = 0.75$, the dynamic oscillates between each city being high wage, high rent, high skill to low wage, low rent, low skill. Intuitively, a city that has a high college share appears desirable to high and low skill workers. Both groups migrate in to the desirable city, but the in-migration of the low skill overwhelms the migration of the high skill. This sharply lowers the college share in the following period, making the city undesirable, leading to large out-migration. By dampening the migration process, the migration response each period to desirable cities does not drive down the utility of that city as much in the following period. This allows the dynamic to converge to an equilibrium.

While I have only evaluated dynamic adjustment and stability under a simple assumed dynamic, future work could attempt to estimate the dynamics of this system. Under a known dynamic adjustment process, counterfactual policy experiments, such as place-based policies, could be run to assess the possible equilibrium outcomes for different cities.

A.4 Comparison of Productivity Estimates to Outside Research

The large literature studying the nationwide increases in the college wage gap finds that the advent of computers and the internet has led to increases in high skill productivity and decreases in low skill productivity.[53] Computers substitute for routine tasks often performed by low skill workers such as the tasks performed by bookkeepers, bank tellers, and secretaries. Computers complement non-routine abstract tasks often performed by high skill workers such as computer scientists, engineers, and physicians. Recent work by Autor and Dorn (2012) shows that the commuting zones (a geographic unit similar to MSAs) that historically employed workers in routine tasks experienced more wage growth for high skill workers, along

[53] See Acemoglu and Autor (2011) for a review of this literature.

with larger wage declines for low skill workers employed in routine tasks from 1980-2005.[54]

Additionally, work by Beaudry, Doms, and Lewis (2010) shows how cities most likely to adopt new computer technology are those with higher skill workforces, since the returns to the technology adoption in high skill areas are higher. Thus, a validation check of my model is whether the locations which Autor and Dorn (2012) measure to have a historically high share of employment in occupations requiring routine tasks and the locations which Beaudry, Doms, and Lewis (2010) measure to have the highest rates of computer technology adoption coincide with my model's predictions of which cities experienced the largest amounts of skill biased technological change.

Appendix Table A.5 reports the MSA which had the largest increases in the gap between high skill productivity and low skill productivity, which measures how skill biased local productivity change was. While Autor and Dorn (2012) and Beaudry, Doms, and Lewis (2010) do not directly report a city ranking based on their metrics, from looking at cities labeled on their graphs, the cities ranking highest on routine tasks share include New York, NY, Chicago, IL, San Francisco, CA, Detroit, MI, Milwaukee, WI, and Providence, RI. My model ranks New York, Chicago, San Francisco, and Milwaukee in the top ten largest increases in the skill biased productivity change. Cities ranking lowest for routine task share include: Fresno, CA, West Palm Beach, FL, New Orleans, LA, Virginia Beach, VA, Orlando, FL, and Scranton, PA. The model predicts Fresno, CA and New Orleans LA to be in the top ten cities with the least amount of skill bias productivity change.

The cities ranking highest on computer adoption, as measured by Beaudry, Doms, and Lewis (2010) are: San Francisco Bay Area, CA, Seattle, WA, Washington DC, Houston, TX, Boston, MA and Dallas, TX. The model ranks San Francisco-Oakland-Vallejo, CA, San Jose, CA, Seattle-Everett, WA, Houston-Brazoria, TX, Dallas-Fort Worth, TX and Boston,

[54] Autor and Dorn (2012) also focus on the polarization of the labor market by comparing wage changes of non-routine, low skill jobs, mainly comprised of local service occupations, with the wages of routine, low-skill jobs. Modeling local wage changes for different types of low skill workers within an MSA is beyond the scope of this paper and a point of future research.

MA-NH in the top ten largest increases in the local productivity gap.[55]

The cities with the largest increases in low skill productivity are very different locations. The city with the largest increase in low skill productivity, a mere 5.3% increase, was Greensboro-Winston Salem-High Point, NC, followed by Austin, TX, Riverside-San Bernardino,CA, Fresno, CA, and Providence-Fall River-Pawtucket, MA/RI. Greensboro-Winston Salem-High Point, NC is known for large textile, tobacco, and furniture corporations. While some manufacturing jobs in these industries have left, the area remains the national center for textile manufacturing. The area has more recently focused on attracting technology industries to the area. For example the computer manufacturer Dell Inc opened a computer assembly plant in the area during the early 2000s. This area is still highly focused on manufacturing and has fared better than other manufacturing towns during the 80s and 90s. The cities with the largest decreases in low skill productivity represent the areas hit hardest by the manufacturing decline. The largest low skill productivity declines were in Pittsburgh, PA, followed by Youngstown-Warren, OH-PA, San Jose, CA, Milwaukee, WI, and Seattle-Everett, WA.

One reason some manufacturing towns fared better than others during the 80s and 90s is based on their exposure to Chinese exports. Work by Autor, Dorn, and Hanson (2012) analyzes which cities' had large amounts of their labor force employed in industries that faced the most competition from the increased penetration of Chinese exports. They find the cities which were most exposed to Chinese export competition experienced wage and employment declines, particularly for low skill workers. Autor, Dorn, and Hanson (2012) reports the top 10 Commuting Zones of the 40 largest commuting zones which experienced the largest exposure to Chinese export competition from 1990 to 2007. I compare this to my list of locations with top 10 largest productivity declines for low skill workers. My ranking is taken from the sample of the 75 largest MSAs, unlike Autor, Dorn, and Hanson (2012)'s sample, which contains only the largest 40 commuting zones. Three cities in my top 10 list

[55] Beaudry, Doms, and Lewis (2010) does not explicitly report which cities had the least computer adoption, which prevents doing a comparison.

are too small to be included in Autor, Dorn, and Hanson (2012)'s list. Of the remaining seven cities, four cities appear in both lists: San Jose, CA, Milwaukee, WI, Boston, MA-NH, and Chicago, IL. This provides further evidence that the model predictions about productivity differences across space fits well with outside knowledge on which cities are thought to have higher or lower productivity levels.

A.5 College Share Changes and Housing Supply Elasticity

While local productivity changes appear to be the predominant driver of changes in cities' employment, cities with less elastic housing supplies have also experienced larger increases in their college shares. Appendix Table A.6 shows that cities with less elastic housing supply have had larger increases in wages, rents, and college employment ratios. This phenomenon can be explained by the national increase in population over time. As the nation's population grows larger, the cities with less elastic housing supply will experience relatively higher increases in housing rents. Since low skill workers have a higher demand elasticity with respect to rent than higher skill workers, the rent increases in inelastic cities will lead to more out migration of low skill than high skill workers. This, in turn, leads to larger increases in the college employment ratios of housing inelastic cities. Increases in a city's college employment ratio raises all workers' wages, as shown by the estimates for cities' labor demand curve, as well as increase the city's amenities. These effects will mitigate the out-migration driven by the initial increase in the city's rent, keeping the rent high in these housing inelastic cities. Thus, nationwide population growth over time should lead to larger increases in wages, rents, and college employment ratios in housing inelastic cities.

To test whether the observed relationship between college share growth and housing supply elasticity is consistent with the model's predicted equilibrium response to nationwide population growth, I simulate this counterfactual scenario. I hold fixed the exogenous productivity and amenities levels of cities in 1980, but I allow the nationwide population to shift from its 1980 level to the level observed in 2000. This population change includes both the overall increase in the number of people, as well as the changes in characteristics of the population, such as the increase in the average education level of the population.

Using the 2000 population but the 1980 exogenous amenity and productivity levels, I solve for the equilibrium wages, rents, and population distribution. Since this model can have multiple equilibria, I must select between equilibria. The equilibrium is selected by assuming a process of how workers adjust to local wages, rents, and amenities. I assume an adjustment process which is a simplified version of the iterated best response learning dynamic often used to solve dynamic games in industrial organization. (See Pakes and McGuire (2001).) To solve for the equilibrium, I take the worker population of 2000 and assume these workers start out living in the cities based on the population distribution observed in 1980. I then compute what local wages, rents, and amenities would be based on this population distribution across cities. I then allow 5% of workers to migrate between cities based on this new set of wages, rents, and amenities.[56] I don't allow 100% of workers to adjust to cities' wages, rents, and amenities instantaneously to allow for a "sluggish adjustment process." There is a large amount of evidence that workers slowly adjust to wages across space.[57] After these 5% of workers have migrated, I recompute new local wages, rents, and amenities and iterate on this process until it converges. I have tried a variety of alternative starting population distributions and population adjustments rates. While the final equilibrium can differ slightly based on these choices, all equilibria which I have been able to compute exhibit very similar patterns of how workers sort into cities with different housing supply elasticities.

Using the counterfactual equilibrium college employment ratios across cities, I measure the change in each city's college employment ratio from the observed level in 1980 to the counterfactual level in 2000. Appendix Table A.7 shows regressions of these counterfactual college employment ratio changes on each city's inverse housing supply elasticity. I find

[56]The 1980-2000 census show that 12% of workers move between MSAs within a 5 year period. I have tried a variety of migration adjustment rates between 1% and 10% and all yield similar results.

[57]See Blanchard and Katz (1992) for a study on the properties of worker migration dynamics.

that cities with less elastic housing supplies would have had larger increases their college employment ratios.

To gauge whether the magnitude of sorting across areas with differing housing supply elasticities in this counterfactual is similar to that observed in the real data, I run the same regression using the actual changes in cities' college employment ratios on cities' inverse housing supply elasticities. Appendix Table A.7 shows that the coefficient on this regression is almost identical to the regression using the counterfactual college shares. A test of whether these coefficients are the same cannot be rejected. As further evidence, I regress the counterfactual college employment ratio changes on the actual college employment ratio changes across cities. Appendix Table A.7 shows that a coefficient of 1 cannot be rejected, meaning that the magnitude of sorting across cities with differing housing supplies can be explained by the nationwide population change from 1980 to 2000. However, the R squared of this regression is only 0.084. The interaction of nationwide population growth with housing supply elasticity heterogeneity mechanism does not explain very much of the observed changes in city' college employment ratios, the overwhelming driver is local productivity change.

A.6 Microfounding Endogenous Amenities

I microfound the endogenous amenities in a city by modeling them as local goods, such as bars, restaurants, and dry cleaners. While this setup provides a possible mechanism of how increases in a city's college share could lead to more desirable amenities, I do not intend to argue that this mechanism is the main driver behind my estimated results. Other important possible mechanisms include direct preferences for more educated neighbors, peer effects within the local school system leading to better schools, or changes in crime or air quality. Since many of my observable amenities involve the amount of local goods per-capita, I focus on that mechanism here.

I assume that all workers have a taste for variety in the local goods available. By modeling the firms which produce local goods as monopolistically competitive, increases in a city's total income across all residents leads to a larger variety of local goods and services available. However, for a given number of local goods available in the city, I assume that increases in population leads to congestion within the city, which lowers workers' utility. For a given number of stores, increases in population will lead to longer lines within each store, as well as more traffic congestion on city streets. Since college educated workers earn more income than lower educated workers, an increase in a city's college population will increase the variety of goods available in the city by a larger amount than an increase in the population of lower skill workers. However, increases in the population of each education group lead to the same amount of congestion in the city. If the increase in the variety of local goods outweighs the congestion costs for high skill workers, but not for low skill workers, then the overall desirability of the cities' local goods is increasing is college workers and decreasing in non-college workers.

A.6.1 Workers' demand for local goods

Let worker a of type edu, earn a wage W_{jt}^{edu} in city j. He consumes a national good O, housing M, and local goods k where the total number of local goods is K_{jt}.[58] The price of the national good is P_t, the price of housing is R_{jt}, and the price of each local good is P_{kjt}. The worker also consumes the exogenous amenities in city. The consumption value of local amenities, such as parks and climate, are freely consumed by all city residents He maximizes his utility function subject to his budget constraint:

$$\max_{O,H,q_1,..q_K} \ln\left(O^{1-\alpha-\beta}\right) + \ln\left(M^{\alpha}\right) + \ln\left(\sum_k q_k^{\rho}\right)^{\frac{\beta}{\rho}} - c\frac{(H_{jt}+L_{jt})}{K_{jt}} + A_{jt}$$
$$s.t.\ P_tO + R_{jt}M + \sum_k P_{kjt}q_k \leq W_{jt}^{edu}.$$

[58]I only allow workers to be heterogeneous based in their wage in this setup. Allowing richer preferenence heterogeneity does not change the results, but complicates exposition.

Utility from the consumption of local goods is measured by $\ln\left(\sum_k q_k^{\rho}\right)^{\frac{\beta}{\rho}}$. Workers have a love of variety for local goods k, as represented by the CES functional form. The amount of consumption of good k is represented by q_k. Higher levels of q_k measures both consumption of higher qualities of k as well as higher quantities. For example, consider the consumption of coffee at a coffee shop. Higher values of q could represent buying a large size instead of a small, or it could represent buying a higher quality coffee such as a latte, instead of drip coffee.

To consume the local goods, the worker must "pay" a congestion cost, $c\frac{(H_{jt}+L_{jt})}{K_{jt}}$, proportional to the number of shoppers per store in the city. For a given number of local goods, K_{jt}, increases in the city population creates crowding at local stores, as well as more traffic congestion in the city streets. If there are a fewer number of stores per-capita, each consumer must drive further, on average, within the city to get to each store, causing more congestion. The congestion cost is proportional to the number of consumers, but does not depend on the amount each consumer spends on each good. For example, driving a given distance to either Walmart or Whole Foods leads to similar amounts of traffic, even if the Whole Foods consumer will spend more money on their journey.

Demand for each local product k, as a function of its price p_{kjt}, is equal to the sum of each individual's demand. Since workers are identical, except that workers with different education earn different wages, the product demand function is:

$$q\left(p_{kjt}\right) = \frac{\beta\left(p_{kjt}\right)^{-\frac{1}{1-\rho}}\left(W_{jt}^{H}H_{jt} + W_{jt}^{L}L_{jt}\right)}{\left(\sum_k p_{kjt}^{-\frac{\rho}{1-\rho}}\right)^{\rho}}.$$

A.6.2 Firms' supply of local goods

There are a large number of potential entrants into the local goods market in city j. Each firm has identical production costs equal to:

$$c\left(q_{kjt}\right) = F_t + \tau R_{jt} q_{kjt}.$$

There is a fixed cost of operation, F_t. The marginal cost of production depends on the housing rental rate R_{jt} and τ measures the amount of store space needed to stock amount q_{jkt}. Local good producers must rent a storefront to sell their products and they must rent a larger space to sell a larger quantity. Since q_{jkt} measures both quality and quantity, this can be thought of both the need for space to store larger quantities of goods, as well as the extra needs for space to store higher quality goods. For example, fresh juice requires a refrigerator, while juice from concentrate only requires shelf space.

Firms maximize profits by pricing such that marginal cost equals marginal revenue:

$$\max_{P_{kjt}} \pi_{kjt} = \max_{p_{kjt}} pq\left(P_{kjt}\right) - F_t - \tau R_{jt} q\left(P_{kjt}\right).$$

Each firm's pricing equation is:

$$P_{jkt} = P_{jt} = \frac{\tau R_{jt}}{\rho}.$$

Since firms' all have identical marginal costs, and workers have CES preferences, all goods within the same city have the same price, which I will refer to as P_{jt}.

The number of local good produced in the city is pinned down from a free entry condition. Since firms can freely enter, equilibrium profits must equal zero.

$$\pi_{kjt} = \frac{\tau R_{jt}}{\rho} q\left(\frac{\tau R_{jt}}{\rho}\right) - F_t - \tau R_{jt} q\left(\frac{\tau R_{jt}}{\rho}\right) = 0.$$

Since all products have identical prices and workers' preferences are symmetric across products, each firm's total revenue is equal to the total amount of expenditure on local goods within the city, $\beta\left(W_{jt}^H H_{jt} + W_{jt}^L L_{jt}\right)$, divided by the total number of products for sale in the local goods market, K.

$$\frac{\beta\left(W_{jt}^H H_{jt} + W_{jt}^L L_{jt}\right)}{K} - F_t - \tau R_{jt} \frac{\beta\left(W_{jt}^H H_{jt} + W_{jt}^L L_{jt}\right)}{\frac{\tau R_{jt}}{\rho} K} = 0.$$

Solving for K :

$$K = \frac{(1-\rho)\,\beta\left(W_{jt}^{H}H_{jt} + W_{jt}^{L}L_{jt}\right)}{F_t}. \tag{A.8}$$

The number of local goods for sale within a city is increasing in population and wages.

A.6.3 Comparative Statics

Returning to workers' utility functions of each city, I solve for the worker's indirect utility function. Since the utility function is Cobb-Douglas, the worker's expenditure share on housing is equal to α, the total expenditure share on local goods is β, which makes the expenditure share on national goods equal to $(1-\alpha-\beta)$. The indirect utility function is:

$$\begin{aligned} V_{ijt} &= (1-\alpha-\beta)\ln\left(\frac{(1-\alpha-\beta)\,W_{jt}^{edu}}{P_t}\right) + \alpha\ln\left(\frac{\alpha W_{jt}^{edu}}{R_{jt}}\right) \\ &\quad + \beta\ln\left(\sum_k\left(\frac{\beta W_{jt}^{edu}\rho}{K\tau R_{jt}}\right)^{\rho}\right)^{\frac{1}{\rho}} - c\frac{(H_{jt}+L_{jt})}{K_{jt}} + A_{jt}. \end{aligned}$$

Since all local products have the same price and worker preferences are symmetric across products, each worker divides his local goods expenditure equally across all local goods. Dropping terms which do not vary across cities and grouping like terms gives:

$$V_{ijt} = \ln\left(W_{jt}^{edu}\right) - (\alpha+\beta)\ln\left(R_{jt}\right) + \beta\left(\frac{1-\rho}{\rho}\right)\ln\left(K_{jt}\right) - c\frac{(H_{jt}+L_{jt})}{K_{jt}} + A_{jt}. \tag{A.9}$$

Since both housing and local goods' prices reflect housing costs, the coefficient on the housing rental rate reflects consumption both on housing and local goods. Equation $(A.9)$ also explicitly shows the trade off between increases in the number of local good providers, K_{jt}, and increases in congestion costs $c\frac{(H_{jt}+L_{jt})}{K_{jt}}$.

I now plug in the equilibrium number of local good providers in the city, equation $(A.8)$, and drop terms which do not vary across cities. This leads to the indirect utility function rewritten as:

$$\begin{aligned} V_{jt} &= \ln\left(W_{jt}^{edu}\right) - (\alpha+\beta)\ln\left(R_{jt}\right) + \beta\left(\frac{1-\rho}{\rho}\right)\ln\left(W_{jt}^{H}H_{jt} + W_{jt}^{L}L_{jt}\right) \\ & -c\frac{F_t\left(H_{jt}+L_{jt}\right)}{(1-\rho)\beta\left(W_{jt}^{H}H_{jt}+W_{jt}^{L}L_{jt}\right)} + A_{jt}. \end{aligned}$$

Taking comparative statics of utility with respect to high skill population gives:

$$\frac{\partial V_{ijt}}{\partial H_{jt}} = \underbrace{\beta\left(\frac{1-\rho}{\rho}\right)\frac{W_{jt}^{H}}{W_{jt}^{H}H_{jt}+W_{jt}^{L}L_{jt}}}_{\text{Utility gain from increase in variety}} - \underbrace{\frac{cF_t}{(1-\rho)\beta}\frac{\left(W_{jt}^{L}-W_{jt}^{H}\right)L_{jt}}{\left(W_{jt}^{H}H_{jt}+W_{jt}^{L}L_{jt}\right)^2}}_{\text{Utility gain from decreased congestion}} > 0. \tag{A.10}$$

Equation $(A.10)$ show that increases in a city's high skill population leads to a net utility gain from the expansion of local product variety and changes in the congestion costs. Since high skill workers' incomes allows them to expand the product market more than lower skill workers, population per local good provider falls. This lowers congestions costs by decreasing the number of shoppers per store.

Similarly, the utility impact from increases in non-college population is:

$$\frac{\partial V_{ijt}}{\partial L_{jt}} = \underbrace{\beta\left(\frac{1-\rho}{\rho}\right)\frac{W_{jt}^{L}}{W_{jt}^{H}H_{jt}+W_{jt}^{L}L_{jt}}}_{\text{Utility gain from increase in variety}} - \underbrace{\frac{cF_t}{(1-\rho)\beta}\frac{\left(W_{jt}^{H}-W_{jt}^{L}\right)H_{jt}}{\left(W_{jt}^{H}H_{jt}+W_{jt}^{L}L_{jt}\right)^2}}_{\text{Utility loss from increased congestion}}.$$

An increase in low skill workers in the city will expand the local good market. However, since the low skill population cannot support as many stores per worker as the high skill population, increases in the low skill population lead to more crowding in each store. The overall utility impact of low skill population increases is ambiguous, and depends on whether the utility gain from the love of variety outweighs the utility cost of additional congestion. If the overall utility impact of low skill population changes is negative, then "amenities," as measured by the desirability of the local goods market, is increasing in high skill population and decreasing in low skill population.

Appendix Table 1.8 shows the reduced form causal effect of a city's college employment ratio on many local amenities. Using high and low skill industrial labor demand shocks as instruments for cities' college employment ratios, I find high college employment ratios leads to more apparel stores per capita, eating and drinking places per capita, and movie theaters per capita. The college employment ratio also decreases property crime rates and improves local air quality. This motivates using the ratio of college to non-college residents of a city as an index for the desirability of local endogenous amenities.

A.7 Supplementary Tables and Figures

Table A.1: Summary Statistics of Household Head City Choice Samples

	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
A. 1980										
	Non-College Sample					College Sample				
Black	1491803	0.121	0.327	0	1	490596	0.051	0.219	0	1
Immigrant	1491803	0.075	0.263	0	1	490596	0.079	0.269	0	1
Live in State of Birth	1380427	0.644	0.479	0	1	451954	0.488	0.500	0	1
B. 1990										
	Non-College Sample					College Sample				
Black	1697029	0.107	0.308	0	1	651938	0.053	0.225	0	1
Immigrant	1697029	0.090	0.286	0	1	651938	0.097	0.296	0	1
Live in State of Birth	1544252	0.661	0.473	0	1	588876	0.485	0.500	0	1
C.2000										
	Non-College Sample					College Sample				
Black	1900470	0.124	0.329	0	1	823856	0.067	0.250	0	1
Immigrant	1900470	0.132	0.338	0	1	823856	0.135	0.341	0	1
Live in State of Birth	1650555	0.669	0.470	0	1	712929	0.499	0.500	0	1

Notes: Sample is all heads of household between ages 25 and 55 with positive wage earnings. This sample is used to estimate workers' preferences for cities. Summary statistics for whether a worker lives in his state of birth is restricted to non-immigrant workers. College is defined as having a 4 year college degree.

Table A.2: Amenity Changes Associated with College and Non-College Population Changes: 1980-2000

	(1)	(2)	(3)	(4)	(5)
	Grocery Stores per 1000 Residents	Apparel Stores per 1000 Residents	Eating and Drinking Places per 1000 Residents	Book Stores per 1000 Residents	Dry Cleaners per 1000 Residents
Δ Log Non-College Population	0.0195	-0.441***	-0.212***	-0.118	-0.394***
	[0.0847]	[0.0913]	[0.0532]	[0.162]	[0.141]
Δ Log College Population	-0.225**	0.405***	0.170***	0.131	0.419***
	[0.0896]	[0.0967]	[0.0563]	[0.172]	[0.149]
Constant	0.344***	-0.413***	0.234***	0.146**	-0.0251
	[0.0383]	[0.0413]	[0.0240]	[0.0733]	[0.0636]
Observations	218	218	218	218	218
R-squared	0.083	0.097	0.071	0.003	0.038

	(6)	(7)	(8)	(9)	(10)
	Movie Thearters per 1000 Residents	Museums and Art Galleries per 1000 Residents	Propery Crimes per 1000 Residents	Violent Crimes Per 1000 Residents	EPA Air Quality Index
Δ Log Non-College Population	-0.340*	-0.366	0.147	-0.145	0.311*
	[0.175]	[0.246]	[0.124]	[0.164]	[0.175]
Δ Log College Population	0.310*	0.428*	-0.245*	0.123	-0.287
	[0.185]	[0.253]	[0.131]	[0.174]	[0.177]
Constant	-0.849***	0.649***	0.252***	-0.0304	-0.0649
	[0.0793]	[0.108]	[0.0563]	[0.0746]	[0.0759]
Observations	218	174	215	215	177
R-squared	0.017	0.016	0.019	0.004	0.018

Notes: Standard errors in brackets. Changes measured between 1980-2000. Retail and local service establishments per capita data come from County Business Patters 1980, 2000. Crime data is from the FBI. Air Quality Index in from the EPA. Higher values of of the air quality index indicate more pollution. *** p<0.01, ** p<0.05, * p<0.1

Table A.3: Largest and Smallest Amenity Changes across 75 Largest Cities

Largest Increases in College Amenities	Δ Amenity	Largest Increases in Non-College Amenities	Δ Amenity
Raleigh-Durham, NC	0.904	Raleigh-Durham, NC	0.439
Las Vegas, NV	0.827	Boston, MA-NH	0.410
Charlotte-Gastonia-Rock Hill, NC-SC	0.804	Scranton-Wilkes-Barre, PA	0.389
Boston, MA-NH	0.742	Rochester, NY	0.382
Providence-Fall River, MA/RI	0.738	Harrisburg-Lebanon--Carlisle, PA	0.352
Orlando, FL	0.722	Atlanta, GA	0.346
Tacoma, WA	0.682	Syracuse, NY	0.336
West Palm Beach-Boca Raton, FL	0.679	Allentown-Bethlehem-Easton, PA/NJ	0.335
Atlanta, GA	0.670	Charlotte-Gastonia-Rock Hill, NC-SC	0.333
Scranton-Wilkes-Barre, PA	0.665	Pittsburgh, PA	0.328
Largest Decreases in College Amenities	**Δ Amenity**	**Largest Decreases in Non-College Amenities**	**Δ Amenity**
Fresno, CA	-	Los Angeles-Long Beach, CA	-
Baton Rouge, LA	0.006	Ventura-Oxnard-Simi Valley, CA	0.027
Los Angeles-Long Beach, CA	0.011	San Diego, CA	0.031
Tulsa, OK	0.038	Fresno, CA	0.040
San Jose, CA	0.052	San Jose, CA	0.057
Oklahoma City, OK	0.054	San Francisco-Oakland-Vallejo, CA	0.061
Houston-Brazoria, TX	0.114	Sacramento, CA	0.093
New Orleans, LA	0.127	Honolulu, HI	0.106
Miami-Hialeah, FL	0.140	Miami-Hialeah, FL	0.121
Hartford-Bristol-Middleton, CT	0.158	Fort Lauderdale-Hollywood, FL	0.122
Best Amenities, College Workers, 1980	**Amenity**	**Best Amenities, Non-College Workers, 1980**	**Amenity**
Los Angeles-Long Beach, CA	2.889	Los Angeles-Long Beach, CA	1.213
San Francisco-Oakland-Vallejo, CA	2.530	San Francisco-Oakland-Vallejo, CA	0.956
Washington, DC/MD/VA	2.492	San Diego, CA	0.910
Denver-Boulder, CO	2.285	Phoenix, AZ	0.855
New York-Northeastern NJ	2.167	Honolulu, HI	0.829
Seattle-Everett, WA	2.143	Denver-Boulder, CO	0.818
Chicago, IL	2.126	San Jose, CA	0.816
Dallas-Fort Worth, TX	2.117	Tampa-St. Pete-Clearwater, FL	0.773
Atlanta, GA	2.039	New York-Northeastern NJ	0.728
Houston-Brazoria, TX	2.021	Fort Lauderdale-Hollywood, FL	0.728

Table A.3 (Continued)

Worst Amenities, College Workers, 1980	Amenity	Worst Amenities, Non-College Workers, 1980	Amenity
Scranton-Wilkes-Barre, PA	-	Syracuse, NY	-
Youngstown-Warren, OH-PA	0.075	Rochester, NY	0.013
Allentown-Bethlehem-Easton, PA/NJ	0.112	Allentown-Bethlehem-Easton, PA/NJ	0.042
Syracuse, NY	0.117	Toledo, OH/MI	0.044
Harrisburg-Lebanon--Carlisle, PA	0.183	Harrisburg-Lebanon--Carlisle, PA	0.045
Toledo, OH/MI	0.210	Scranton-Wilkes-Barre, PA	0.050
Rochester, NY	0.457	Youngstown-Warren, OH-PA	0.051
Albany-Schenectady-Troy, NY	0.464	Albany-Schenectady-Troy, NY	0.120
Buffalo-Niagara Falls, NY	0.507	Buffalo-Niagara Falls, NY	0.126
Providence-Fall River-Pawtucket, MA/I	0.583	Grand Rapids, MI	0.148
Best Amenities, College Workers, 2000	**Amenity**	**Best Amenities, Non-College Workers, 2000**	**Amenity**
Los Angeles-Long Beach, CA	2.473	Los Angeles-Long Beach, CA	0.892
Washington, DC/MD/VA	2.414	Phoenix, AZ	0.820
Boston, MA-NH	2.290	Denver-Boulder, CO	0.737
Atlanta, GA	2.282	Seattle-Everett, WA	0.714
San Francisco-Oakland-Vallejo, CA	2.279	Boston, MA-NH	0.710
Denver-Boulder, CO	2.277	San Francisco-Oakland-Vallejo, CA	0.695
Seattle-Everett, WA	2.259	Tampa-St. Pete-Clearwater, FL	0.695
Phoenix, AZ	2.202	Atlanta, GA	0.691
New York-Northeastern NJ	2.102	Las Vegas, NV	0.689
Dallas-Fort Worth, TX	2.096	New York-Northeastern NJ	0.658
Worst Amenities, College Workers, 2000	**Amenity**	**Worst Amenities, Non-College Workers, 2000**	**Amenity**
Youngstown-Warren, OH-PA	-	Youngstown-Warren, OH-PA	-
Allentown-Bethlehem-Easton, PA/NJ	0.121	Toledo, OH/MI	0.008
Syracuse, NY	0.200	Syracuse, NY	0.015
Harrisburg-Lebanon--Carlisle, PA	0.221	Buffalo-Niagara Falls, NY	0.042
Scranton-Wilkes-Barre, PA	0.238	Allentown-Bethlehem-Easton, PA/NJ	0.056
Toledo, OH/MI	0.288	Albany-Schenectady-Troy, NY	0.062
Buffalo-Niagara Falls, NY	0.431	Rochester, NY	0.073
Akron, OH	0.436	Harrisburg-Lebanon--Carlisle, PA	0.075
Albany-Schenectady-Troy, NY	0.440	Grand Rapids, MI	0.096
Fresno, CA	0.485	Akron, OH	0.117

Notes: Sample reports top and bottom 10 from the 75 biggest cities by 1980 population. Local amenities are inferred from model estimates. Local high and low skill amenities are normalized to 0 in city least with the least desirable amenities in 1980 and 2000. Units measure the log wage value equilvalent to the utility difference between the amenities in the given city and the city normalized to 0. See text for further details.

Table A.4: Largest and Smallest Productivity Changes across 75 Largest MSAs

Largest Increases in College Productivity		Largest Increases in Non-College Productivity	
	Δ Prod.		Δ Prod.
San Jose, CA	0.442	Greensboro-Winston Salem, NC	0.053
Boston, MA-NH	0.423	Austin, TX	0.050
San Francisco-Oakland-Vallejo, CA	0.402	Riverside-San Bernardino,CA	0.029
Tampa-St. Petersburg-Clearwater, FL	0.356	Fresno, CA	0.021
Providence-Fall River, MA/RI	0.344	Providence-Fall Rivert, MA/RI	0.017
Austin, TX	0.341	San Antonio, TX	0.017
Greenville-Spartanburg-Anderson SC	0.340	Tacoma, WA	0.016
New York-Northeastern NJ	0.337	Memphis, TN/AR/MS	0.012
Ventura-Oxnard-Simi Valley, CA	0.334	Las Vegas, NV	0.008
Chicago, IL	0.332	Salt Lake City-Ogden, UT	-0.002
Largest Decreases in College Productivity		**Largest Decreases in Non-College Productivity**	
	Δ Prod.		Δ Prod.
Rochester, NY	0.089	Pittsburgh, PA	-0.176
Baton Rouge, LA	0.125	Youngstown-Warren, OH-PA	-0.133
New Orleans, LA	0.138	San Jose, CA	-0.131
Fresno, CA	0.160	Milwaukee, WI	-0.114
Syracuse, NY	0.173	Seattle-Everett, WA	-0.112
Toledo, OH/MI	0.176	Boston, MA-NH	-0.111
Oklahoma City, OK	0.187	Chicago, IL	-0.106
Knoxville, TN	0.189	Akron, OH	-0.104
Albany-Schenectady-Troy, NY	0.202	Louisville, KY/IN	-0.101
Honolulu, HI	0.202	Detroit, MI	-0.098
Most Productive, College Workers, 1980		**Most Productive, Non-College Workers, 1980**	
	Prod.		Prod.
Washington, DC/MD/VA	0.518	Detroit, MI	0.463
San Jose, CA	0.439	Youngstown-Warren, OH-PA	0.444
Houston-Brazoria, TX	0.394	Toledo, OH/MI	0.394
New York-Northeastern NJ	0.386	Allentown-Bethlehem-Easton, PA/NJ	0.357
Rochester, NY	0.379	Riverside-San Bernardino,CA	0.354
Hartford-Bristol-Middleton, CT	0.359	Cleveland, OH	0.346
San Francisco-Oakland-Vallejo, CA	0.345	Milwaukee, WI	0.345
Ventura-Oxnard-Simi Valley, CA	0.330	Las Vegas, NV	0.345
Denver-Boulder, CO	0.328	Akron, OH	0.342
Los Angeles-Long Beach, CA	0.327	Pittsburgh, PA	0.339

Table A.4 (Continued)

Least Productive, College Workers, 1980		Least Productive, Non-College Workers, 1980	
	Prod.		Prod.
Scranton-Wilkes-Barre, PA	0.000	Austin, TX	0.000
Youngstown-Warren, OH-PA	0.058	Raleigh-Durham, NC	0.048
Las Vegas, NV	0.080	Washington, DC/MD/VA	0.064
Providence-Fall River, MA/RI	0.084	San Antonio, TX	0.121
Tampa-St. Petersburg-Clearwater, FL	0.099	Knoxville, TN	0.148
Greenville-Spartanburg-Anderson SC	0.101	Greensboro-Winston Salem, NC	0.151
Riverside-San Bernardino,CA	0.108	Boston, MA-NH	0.161
Springfield-Holyoke-Chicopee, MA	0.108	Miami-Hialeah, FL	0.163
Tacoma, WA	0.124	Atlanta, GA	0.163
Buffalo-Niagara Falls, NY	0.132	Memphis, TN/AR/MS	0.168
Most Productive, College Workers, 2000		**Most Productive, Non-College Workers, 2000**	
	Prod.		Prod.
San Jose, CA	0.611	Riverside-San Bernardino,CA	0.361
San Francisco-Oakland-Vallejo, CA	0.477	Detroit, MI	0.343
New York-Northeastern NJ	0.453	Las Vegas, NV	0.331
Washington, DC/MD/VA	0.451	Tacoma, WA	0.324
Boston, MA-NH	0.448	Ventura-Oxnard-Simi Valley, CA	0.293
Hartford-Bristol-Middleton, CT	0.398	Youngstown-Warren, OH-PA	0.290
Ventura-Oxnard-Simi Valley, CA	0.395	Toledo, OH/MI	0.288
Chicago, IL	0.375	Grand Rapids, MI	0.273
Houston-Brazoria, TX	0.356	Allentown-Bethlehem-Easton, PA/NJ	0.251
Philadelphia, PA/NJ	0.354	Salt Lake City-Ogden, UT	0.243
Least Productive, College Workers, 2000		**Least Productive, Non-College Workers, 2000**	
	Prod.		Prod.
Youngstown-Warren, OH-PA	0.000	Raleigh-Durham, NC	0.000
Las Vegas, NV	0.024	Washington, DC/MD/VA	0.015
Fresno, CA	0.039	Austin, TX	0.028
Scranton-Wilkes-Barre, PA	0.056	Boston, MA-NH	0.029
Tacoma, WA	0.058	Knoxville, TN	0.097
Riverside-San Bernardino,CA	0.073	San Jose, CA	0.097
Toledo, OH/MI	0.076	Miami-Hialeah, FL	0.100
Oklahoma City, OK	0.106	Tucson, AZ	0.103
Tucson, AZ	0.118	Richmond-Petersburg, VA	0.111
Baton Rouge, LA	0.119	Birmingham, AL	0.112

Notes: Sample reports top and bottom 10 from the 75 biggest cities by 1980 population. Local productivity is inferred from model estimates. Local high and low skill productivities are normalized to 0 in city least productive in 1980 and 2000. Unit measure differnece in log wages between cities directly due to productivity differences. Some MSA names shortened to fit in table column. See text for further details.

Table A.5: Largest and Smallest Productivity Gap Changes across 75 Largest Cities

Largest Exogenous Changes in Log College/Non-College Productivity Gap		Largest Total Changes in Log College/Non-College Productivity Gap	
MSA	Δ Log Prod. Gap	MSA	Δ Log Prod. Gap
San Jose, CA	0.320	San Jose, CA	0.580
San Francisco-Oakland-Vallejo, CA	0.272	Boston, MA-NH	0.520
Boston, MA-NH	0.270	San Francisco-Oakland-Vallejo, CA	0.468
Houston-Brazoria, TX	0.246	Chicago, IL	0.437
Los Angeles-Long Beach, CA	0.237	Louisville, KY/IN	0.398
Chicago, IL	0.235	Seattle-Everett, WA	0.395
Pittsburgh, PA	0.232	Milwaukee, WI	0.388
Phoenix, AZ	0.225	Pittsburgh, PA	0.387
Seattle-Everett, WA	0.222	Buffalo-Niagara Falls, NY	0.376
Dallas-Fort Worth, TX	0.219	New York-Northeastern NJ	0.366
Smallest Exogenous Changes in Log College/Non-College Productivity Gap		Smallest Total Changes in Log College/Non-College Productivity Gap	
MSA	Δ Log Prod. Gap	MSA	Δ Log Prod. Gap
Fort Wayne, IN	0.052	Fort Wayne, IN	0.055
Columbia, SC	0.062	Rochester, NY	0.123
Greensboro-Winston Salem, NC	0.063	Greensboro-Winston Salem, NC	0.127
Colorado Springs, CO	0.084	Fresno, CA	0.138
Tacoma, WA	0.088	New Orleans, LA	0.160
Albany-Schenectady-Troy, NY	0.094	Syracuse, NY	0.171
San Antonio, TX	0.094	Tacoma, WA	0.178
Las Vegas, NV	0.099	Albany-Schenectady-Troy, NY	0.194
Rochester, NY	0.099	Columbia, SC	0.207
Charlotte-Gastonia-Rock Hill, NC-SC	0.100	Las Vegas, NV	0.208

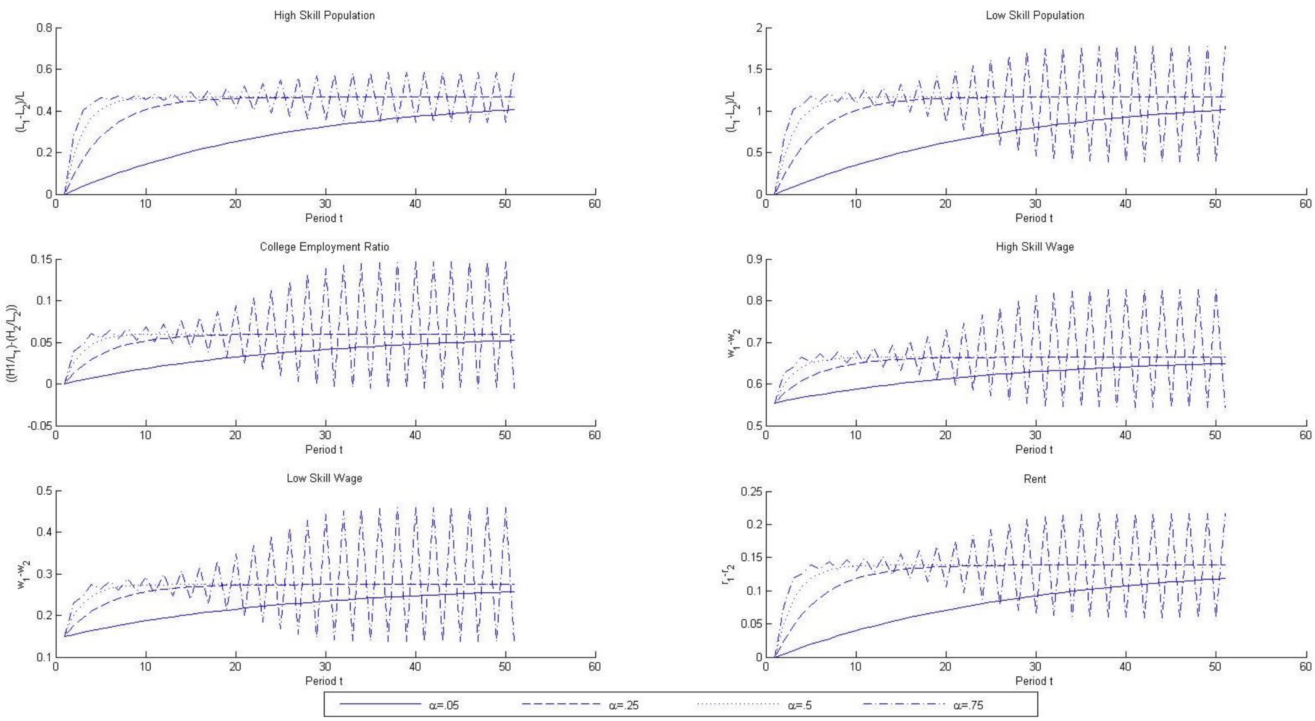
Notes: Sample reports top and bottom 10 from the 75 biggest cities by 1980 population. Local productivity is inferred from model estimates. Change in local productivity gap is defined as the difference between the change in local high skill productivity and low skill productive from 1980 to 2000. Unit measure differneces in the increase in local college wages gaps between cities directly due to productivity differences. Exogenous productivity changes are inferred from the residual of labor demand equations, while total productivity changes include by the exogenous change and the endogenous response due to changes in the local college employment ratio. See text for further details.

Table A.6: Treatment Effect of College Ratio Changes on Amenity Changes: 1980-2000

	(1)	(2)	(3)	(4)	(5)
	Grocery Stores per 1000 Residents	Apparel Stores per 1000 Residents	Eating and Drinking Places per 1000 Residents	Book Stores per 1000 Residents	Dry Cleaners per 1000 Residents
Δ College Emp Ratio	0.176	0.607***	0.322***	-0.139	0.0951
	[0.166]	[0.170]	[0.0998]	[0.301]	[0.263]
Constant	0.158***	-0.485***	0.175***	0.231**	0.0807
	[0.0523]	[0.0536]	[0.0314]	[0.0948]	[0.0826]
Observations	218	218	218	218	218
	(6)	(7)	(8)	(9)	(10)
	Movie Thearters per 1000 Residents	Museums and Art Galleries per 1000 Residents	Propery Crimes per 1000 Residents	Violent Crimes Per 1000 Residents	EPA Air Quality Index
Δ College Emp Ratio	0.732**	0.568	-0.582**	-0.211	-0.583*
	[0.326]	[0.493]	[0.238]	[0.309]	[0.301]
Constant	-0.986***	0.624***	0.323***	0.0636	0.0373
	[0.103]	[0.162]	[0.0752]	[0.0978]	[0.101]
Observations	218	174	215	215	177

Notes: Standard erorrs in brackets. Changes measured between 1980-2000. College employment ratio is defined as the ratio of number of employed college workers to the number of employed lower skill workers living in the city. Instuments are high skill Bartik labor demand shock and low skill Bartik labor demand shock measured from 1980 to 2000. First stage F statistic of instruments strength is 42. Retail and local service establishments per capita data come from County Business Patters 1980, 2000. Crime data is from the FBI. Air Quality Index in from the EPA. Higher values of of the air quality index indicate more pollution. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure A.1: Equilibrium convergence across various migration dampening parameters, α



Notes: All figures plot difference in outcome variable between city 1 and city 2. Cities 1 and 2 have identical exogenous characteristics other than city 1 is more productive for high skill than city 2. See text for details on dynamic adjust process.

B Chapter 2 Appendix

B.1 Income Tax

B.1.1 Government

The local government of city j charges an income tax τ_j to workers who choose to reside within the city. The local government also produces government services, which cost s_j for each worker in the city. N_j measure the population of city j. The local rent seeking government maximizes:

$$\max_{\tau_j, s_j} \tau_j w_j N_j - s_j N_j$$

B.1.2 Workers

All workers are homogeneous. Workers living in city j inelastically supply one unit of labor, and earn wage w_j. Each worker must rent a house to live in the city at rental rate r_j and pay the local income tax τ_j. Workers value the local amenities as measure by A_j. The desirability of government services s_j is represented by $g(s_j)$. Thus, workers' utility from living in city j is:

$$U_j = w_j (1 - \tau_j) - r_j + A_j + g(s_j).$$

Workers maximize their utility by living in the city which they find the most desirable.

B.1.3 Firms

All firms are homogenous and produce a tradeable output Y. Cities exogenously differ in their productivity as measured by θ_j. Local government services impact firms productivity, as measured by $b(s_j)$. The production function is:

$$Y_j = \theta_j N_j + b(s_j) N_j + F(N_j),$$

where $F'(N_j) > 0$ and $F''(N_j) = 0$ in labor. I assume a completely elastic labor demand curve to focus on the role of housing supply elasticity in setting tax rates.

The labor market is perfectly competitive, so wages equal the marginal product of labor:

$$w_j = \theta_j + b(s_j) + F'(N_j).$$

B.1.4 Housing

The housing market is identical to the setting described in the main text in Section 2.2.4. The housing supply curve is:

$$\begin{aligned} r_j &= a_j + \gamma_j \log(N_j), \\ \gamma_j &= \gamma x_j^{\text{house}} \end{aligned}$$

where x_j^{house} is a vector of city characteristics which impact the elasticity of housing supply.

B.1.5 Equilibrium in Labor and Housing

Since all workers are identical, all cities with positive population must offer equal utility to workers. In equilibrium, all workers must be indifferent between all cities. Thus:

$$U_j = w_j(1 - \tau_j) - r_j + A_j + g(s_j) = \bar{U}.$$

Plugging in labor demand and housing supply gives:

$$(\theta_j + b(s_j) + F'(N_j))(1 - \tau_j) - a_j - \gamma_j \log N_j + A_j + g(s_j) = \bar{U}. \tag{B.1}$$

Equation $(B.1)$ determines the equilibrium distribution of workers across cities.

B.1.6 Government Tax Competition

The government maximizes:

$$\max_{s_j,\tau_j} \tau_j w_j N_j - s_j N_j.$$

The first order conditions are:

$$\begin{aligned} 0 &= w_j \tau_j \frac{\partial N_j}{\partial s_j} + \tau_j N_j \frac{\partial w_j}{\partial N_j}\frac{\partial N_j}{\partial s_j} - N_j - s_j \frac{\partial N_j}{\partial s_j} \\ 0 &= \tau_j \left(\frac{\partial w_j}{\partial N_j}\frac{\partial N_j}{\partial \tau_j} N_j + w_j \frac{\partial N_j}{\partial \tau_j} \right) + w_j N_j - s_j \frac{\partial N_j}{\partial \tau_j}. \end{aligned} \tag{B.2}$$

Differentiating equation $(B.1)$ to solve for $\frac{\partial N_j}{\partial s_j}$ and $\frac{\partial N_j}{\partial \tau_j}$ gives:

$$\begin{aligned} \frac{\partial N_j}{\partial s_j} &= N_j \frac{(1-\tau_j)\, b'(s_j) + g'(s_j)}{\gamma_j} > 0 \\ \frac{\partial N_j}{\partial \tau_j} &= -N_j \frac{(\theta_j + b(s_j) + F'(N_j))}{\gamma_j} < 0. \end{aligned} \tag{B.3}$$

Population increases with government services and decreases in taxes. Plugging these into $(B.2)$ and combining the first order conditions shows that government services are provided such that the marginal benefit $\left(\left(1-\tau_j^*\right) b'(s_j) + g'(s_j)\right)$ equals marginal cost (1) :

$$\left(1-\tau_j^*\right) b'\left(s_j^*\right) + g'\left(s_j^*\right) = 1.$$

This is the socially optimal level of government service, given the tax rate.

The equilibrium tax revenue per capita is:

$$w_j \tau_j^* = \gamma_j + s_j^*. \tag{B.4}$$

To analyze the effect of housing supply elasticity on governments' ability to extract rent from taxes, I differentiate the tax markup with respect to the slope of the inverse housing supply curve, γ_j.

$$\frac{\partial}{\partial \gamma_j}\left(w_j \tau_j^* - s_j^*\right) = 1 > 0. \tag{B.5}$$

The government can extract more rent through higher taxes in a city with a less elastic housing supply with a income tax instrument.

B.2 Property Tax

B.2.1 Government

The local government of city j charges a property tax τ_j to workers who choose to reside within the city. The local rent seeking government maximizes:

$$\max_{\tau_j, s_j} \tau_j r_j N_j - s_j N_j$$

B.2.2 Workers

Workers' utility from living in city j facing a property tax τ_j is:

$$U_j = w_j - r_j \left(1 + \tau_j\right) + A_j + g\left(s_j\right).$$

B.2.3 Firms

The production function is:

$$Y_j = \theta_j N_j + b(s_j) N_j + F\left(N_j\right),$$

where $F'\left(N_j\right) > 0$ and $F''\left(N_j\right) = 0$ in labor. I assume a completely elastic labor demand curve to focus on the role of housing supply elasticity in setting tax rates.

The labor market is perfectly competitive, so wages equal the marginal product of labor:

$$w_j = \theta_j + b(s_j) + F'(N_j).$$

B.2.4 Housing

The housing market is identical to the setting described in the main text in Section 2.2.4. The housing supply curve is:

$$\begin{aligned} r_j &= a_j + \gamma_j \log(N_j), \\ \gamma_j &= \gamma x_j^{\text{house}} \end{aligned}$$

where x_j^{house} is a vector of city characteristics which impact the elasticity of housing supply.

B.2.5 Equilibrium in Labor and Housing

Since all workers are identical, all cities with positive population must offer equal utility to workers. In equilibrium, all workers must be indifferent between all cities. Thus:

$$U_j = w_j - r_j(1+\tau_j) + A_j + g(s_j) = \bar{U}.$$

Plugging in labor demand and housing supply gives:

$$(\theta_j + b(s_j) + F'(N_j)) - \left(a_j + \gamma_j \log N_j\right)(1+\tau_j) + A_j + g(s_j) = \bar{U}. \tag{B.6a}$$

Equation (2.1) determines the equilibrium distribution of workers across cities.

B.2.6 Government Tax Competition

The government maximizes:

$$\max_{s_j,\tau_j} \tau_j r_j N_j - s_j N_j.$$

The first order conditions are:

$$0 = r_j \tau_j \frac{\partial N_j}{\partial s_j} + \tau_j N_j \frac{\partial r_j}{\partial N_j}\frac{\partial N_j}{\partial s_j} - N_j - s_j \frac{\partial N_j}{\partial s_j} \tag{B.7}$$

$$0 = \tau_j \left(\frac{\partial r_j}{\partial N_j}\frac{\partial N_j}{\partial \tau_j} N_j + r_j \frac{\partial N_j}{\partial \tau_j} \right) + r_j N_j - s_j \frac{\partial N_j}{\partial \tau_j}. \tag{B.8}$$

Differentiating equation (2.1) to solve for $\frac{\partial N_j}{\partial s_j}$ and $\frac{\partial N_j}{\partial \tau_j}$gives:

$$\begin{aligned} \frac{\partial N_j}{\partial s_j} &= N_j \frac{b'(s_j) + g'(s_j)}{\gamma_j (1+\tau_j)} > 0 \\ \frac{\partial N_j}{\partial \tau_j} &= -N_j \frac{r_j}{\gamma_j (1+\tau_j)} < 0. \end{aligned} \tag{B.9}$$

Combining the first order conditions shows that government services are provided such that the marginal benefit $(b'(s_j) + g'(s_j))$ equals marginal cost (1), which is the same finding for an income tax and head tax:

$$b'\left(s_j^*\right) + g'\left(s_j^*\right) = 1.$$

Plugging $(B.9)$ into $(B.8)$ and rearranging shows the equilibrium tax revenue per capita is:

$$r_j \tau_j^* = \gamma_j + s_j^*. \tag{B.10}$$

Differentiating the tax markup with respect to the slope of the inverse housing supply curve, γ_j.

$$\frac{\partial}{\partial \gamma_j}\left(w_j \tau_j^* - s_j^*\right) = 1 > 0. \tag{B.11}$$

The government can extract more rent through higher taxes in a city with a less elastic housing supply using a property tax instrument. In the case of a property tax, as opposed to a head tax, there are four mechanisms through which a tax rate change impacts government revenue. To break these down, I rewrite the tax rate first order condition:

$$0 = \underbrace{\tau_j r_j \frac{\partial N_j}{\partial \tau_j}}_{\substack{\text{Decline in revenue due} \\ \text{to population decrease}}} + \underbrace{\tau_j \frac{\partial r_j}{\partial N_j} \frac{\partial N_j}{\partial \tau_j} N_j}_{\substack{\text{Decline in revenue} \\ \text{due to rent decrease}}} + \underbrace{r_j N_j}_{\substack{\text{Additional tax revenue} \\ \text{from each resident}}} - \underbrace{s_j \frac{\partial N_j}{\partial \tau_j}}_{\substack{\text{Government services} \\ \text{cost savings}}} \tag{B.12}$$

First, the amount of out-migration driven by a tax hike is influenced by the local housing supply elasticity. This is the first term of equation $(B.12)$. Second, the out-migration lowers rents and directly impacts tax revenues since the tax revenue is a percentage of housing rents. This is the second term of equation $(B.12)$. However, the housing supply elasticity will not impact the size of the rental rate decrease in response to a tax hike. To see this, recall the equilibrium condition, equation $(B.6a)$. For workers to derive utility $\bar{U}$ from this local area, the utility impact of a tax increase must be perfectly offset by a rent decrease.[59] Thus, the equilibrium rental rate response to a given tax increase does not depend on the local housing supply elasticity. Indeed, the housing supply elasticity determines the migration response required to change housing rents in order to offset the utility impact of the tax increase. Thus, a more inelastic housing supply decreases the elasticity of government revenue with respect to the tax rate, giving the government more market power when using a property tax instrument.

The third and forth terms of equation $(B.12)$ show a tax increase raises government revenues from each household and lowers the cost of government services due to out-migration. These channels also appear in the case of a head tax instrument.

[59] Since I have assumed a perfectly elastic labor demand curve, the rental rate response to a tax increase would be the same in any city. However, if labor demand was not perfectly elastic, then the rental rate response to a tax increase could differ with housing supply elasticity, since housing supply elasticity would influence the relative incidence of the tax rate on wages versus rents.

C Chapter 3 Appendix

For the general model, leaving aside terms that do not involve unknown parameters, the log likelihood function is

$$L(\gamma|\mathbf{Y}) = -\frac{1}{2}\ln\left(\det(\Omega(\mathbf{Z},\gamma))\right) - \varepsilon'^{-1}(\mathbf{Z},\gamma)\varepsilon/2.$$

The matrix $\Omega(\mathbf{Z},\gamma)$ is large in our illustrations, with dimension 2,590,190 by 2,590,190. Direct maximization of the log likelihood function is therefore not feasible. However, because locations are measured by puma locations, $\Omega(\mathbf{Z},\gamma)$ has a block structure, and calculations of the log likelihood simplify and can be written in terms of first and second moments by puma. We first give a couple of preliminary results.

Theorem 11. (Sylvester's Determinant Theorem) *Let A and B be arbitrary $M\times N$ matrices. Then:*

$$\det(I_N + A'B) = \det(I_M + BA').$$

Proof of Theorem 11: Consider a block matrix $\left(\begin{smallmatrix} M_1 & M_2 \\ M_3 & M_4 \end{smallmatrix}\right)$. Then:

$$\det\left(\begin{smallmatrix} M_1 & M_2 \\ M_3 & M_4 \end{smallmatrix}\right) = \det\left(\begin{smallmatrix} M_1 & 0 \\ M_3 & I \end{smallmatrix}\right)\det\left(\begin{smallmatrix} I & M_1^{-1}M_2 \\ 0 & M_4 - M_3M_1^{-1}M_2 \end{smallmatrix}\right) = \det M_1 \det(M_4 - M_3M_1^{-1}M_2)$$

similarly

$$\det\left(\begin{smallmatrix} M_1 & M_2 \\ M_3 & M_4 \end{smallmatrix}\right) = \det\left(\begin{smallmatrix} I & M_2 \\ 0 & M_4 \end{smallmatrix}\right)\det\left(\begin{smallmatrix} M_1 - M_2M_4^{-1}M_3 & 0 \\ M_4^{-1}M_3 & I \end{smallmatrix}\right) = \det M_4 \det(M_1 - M_2M_4^{-1}M_3)$$

Letting $M_1 = I_M, M_2 = -B, M_3 = A', M_4 = I_N$ yields the result. □

Lemma 12. (Determinant of Cluster Covariance Matrix) *Suppose $\mathbf{C}$ is an $N\times M$ matrix of binary cluster indicators, with $\mathbf{C}'\mathbf{C}$ equal to a $M\times M$ diagonal matrix, Σ is an arbitrary $M\times M$ matrix, and I_N is the N-dimensional identity matrix. Then, for scalar σ_ε^2,* *and*

$$\Omega = \sigma_\epsilon^2 I_N + \mathbf{C}\Sigma\mathbf{C}' \qquad \Omega_C = \Sigma + \sigma_\epsilon^2(\mathbf{C}'\mathbf{C})^{-1},$$

we have

$$\det(\Omega) = (\sigma_\epsilon^2)^{N-M}\det(\mathbf{C}'\mathbf{C})\det(\Omega_C).$$

Proof of Lemma 12: By Sylvester's theorem:

$$\begin{aligned}\det(\Omega) &= (\sigma_\epsilon^2)^N \det(I_N + \mathbf{C}\Sigma/\sigma_\epsilon^2\mathbf{C}') = (\sigma_\epsilon^2)^N \det(I_M + \mathbf{C}'\mathbf{C}\Sigma/\sigma_\epsilon^2)\\ &= (\sigma_\epsilon^2)^N \det(I_M + \mathbf{C}'\mathbf{C}\Omega_C/\sigma_\epsilon^2 - I_M) = (\sigma_\epsilon^2)^N \det(\mathbf{C}'\mathbf{C})\det(\Omega_C/\sigma_\epsilon^2)\\ &= (\sigma_\epsilon^2)^{N-M}\left(\prod N_p\right)\det(\Omega_C).\end{aligned}$$

□

Lemma 13. *Suppose Assumptions 7 and 8 hold. Then for any $N \times N$ matrix Ω,*

$$\mathbb{E}\left[\mathbf{W}'\Omega\mathbf{W}\right] = \frac{M_1\cdot(M_1-1)}{M\cdot(M-1)}\cdot\iota_N'\Omega\iota_N + \frac{M_1\cdot M_0}{M\cdot(M-1)}\cdot\mathrm{trace}\left(\mathbf{C}'\Omega\mathbf{C}\right).$$

Proof of Lemma 13: We have

$$\mathbb{E}[W_i\cdot W_j] = \begin{cases} M_1/M & \text{if } \forall m, C_{im} = C_{jm},\\ (M_1\cdot(M_1-1))/(M\cdot(M-1)) & \text{otherwise.}\end{cases}$$

it follows that

$$\mathbb{E}[\mathbf{W}\mathbf{W}'] = \frac{M_1\cdot(M_1-1)}{M\cdot(M-1)}\cdot\iota_N\iota_N' + \left(\frac{M_1}{M} - \frac{M_1\cdot(M_1-1)}{M\cdot(M-1)}\right)\cdot\mathbf{C}\mathbf{C}'$$

$$= \frac{M_1\cdot(M_1-1)}{M\cdot(M-1)}\cdot\iota_N\iota_N' + \frac{M_1\cdot M_0}{M\cdot(M-1)}\cdot\mathbf{C}\mathbf{C}'.$$

Thus

$$\mathbb{E}[\mathbf{W}'\Omega\mathbf{W}] = \mathrm{trace}\left(\mathbb{E}[\Omega\mathbf{W}\mathbf{W}']\right)$$

$$= \text{trace}\left(\Omega \cdot \left(\frac{M_1 \cdot (M_1 - 1)}{M \cdot (M-1)} \cdot \iota_N \iota_N' + \frac{M_1 \cdot M_0}{M \cdot (M-1)} \cdot \mathbf{CC}'\right)\right)$$

$$= \frac{M_1 \cdot (M_1 - 1)}{M \cdot (M-1)} \cdot \iota_N' \Omega \iota_N + \frac{M_1 \cdot M_0}{M \cdot (M-1)} \cdot \text{trace}\left(\mathbf{C}'\Omega\mathbf{C}\right). \qquad \square$$

Lemma 14. *Suppose the $N \times N$ matrix Ω satisfies*

$$\Omega = \sigma_\varepsilon^2 \cdot I_N + \sigma_C^2 \cdot \mathbf{CC}',$$

where I_N is the $N \times N$ identity matrix, and $\mathbf{C}$ is an $N \times M$ matrix of zeros and ones, with $\mathbf{C}\iota_M = \iota_N$ and $\mathbf{C}'\iota_N = (N/M)\iota_M$, so that,

$$\Omega_{ij} = \begin{cases} \sigma_\varepsilon^2 + \sigma_C^2 & \text{if } i = j \\ \sigma_C^2 & \text{if } i \neq j, \forall m, C_{im} = C_{jm}, \\ 0 & \text{otherwise,} \end{cases} \tag{C.1}$$

Then, (i)

$$\ln\left(\det\left(\Omega\right)\right) = N \cdot \ln\left(\sigma_\varepsilon^2\right) + M \cdot \ln\left(1 + \frac{N}{M} \cdot \frac{\sigma_C^2}{\sigma_\varepsilon^2}\right),$$

and, (ii)

$$\Omega^{-1} = \sigma_\varepsilon^{-2} \cdot I_N - \frac{\sigma_C^2}{\sigma_\varepsilon^2 \cdot (\sigma_\varepsilon^2 + \sigma_C^2 \cdot N/M)} \cdot \mathbf{CC}'$$

Proof of Lemma 14: First, consider the first part. Apply Lemma 12 with

$$\Sigma = \sigma_C^2 \cdot I_M, \quad \text{and} \quad \mathbf{C}'\mathbf{C} = \frac{N}{M} \cdot I_M, \quad \text{so that} \quad \Omega_C = \left(\sigma_C^2 + \sigma_\varepsilon^2 \cdot \frac{M}{N}\right) \cdot I_M.$$

Then, by Lemma 12, we have

$$\ln\det(\Omega) = (N - M) \cdot \ln(\sigma_\varepsilon^2) + M \cdot \ln(N/M) + \ln\det(\Omega_C)$$

$$= (N - M) \cdot \ln(\sigma_\varepsilon^2) + M \cdot \ln(N/M) + M \cdot \ln\left(\sigma_C^2 + \sigma_\varepsilon^2 \cdot \frac{M}{N}\right)$$

$$= (N-M)\cdot \ln(\sigma_\varepsilon^2) + M\cdot \ln\left(\frac{N}{M}\sigma_C^2+\sigma_\varepsilon^2\right)$$

$$= N\cdot \ln(\sigma_\varepsilon^2) + M\cdot \ln\left(1+\frac{N}{M}\cdot\frac{\sigma_C^2}{\sigma_\varepsilon^2}\right).$$

Next, consider part (ii). We need to show that

$$\left(\sigma_\varepsilon^2\cdot I_N+\sigma_C^2\cdot \mathbf{C}\mathbf{C}'\right)\left(\sigma_\varepsilon^{-2}\cdot I_N-\frac{\sigma_C^2}{\sigma_\varepsilon^2\cdot(\sigma_\varepsilon^2+\sigma_C^2\cdot N/M)}\cdot\mathbf{C}\mathbf{C}'\right)=I_N,$$

which amounts to showing that

$$-\frac{\sigma_\varepsilon^2\cdot\sigma_C^2}{\sigma_\varepsilon^2\cdot(\sigma_\varepsilon^2+\sigma_C^2\cdot N/M)}\cdot\mathbf{C}\mathbf{C}'+\sigma_C^2\cdot\mathbf{C}\mathbf{C}'\sigma_\varepsilon^{-2}-\mathbf{C}\mathbf{C}'\cdot\frac{\sigma_C^4}{\sigma_\varepsilon^2\cdot(\sigma_\varepsilon^2+\sigma_C^2\cdot N/M)}\cdot\mathbf{C}\mathbf{C}'=0.$$

This follows directly from the fact that $\mathbf{C}'\mathbf{C}=(N/M)\cdot I_M$ and collecting the terms. □

Proof of Lemma 5: The unbiasedness result directly follows from the conditional unbiasedness established in Lemma 4. Next we establish the second part of the Lemma. By the Law of Iterated Expectations,

$$\mathbb{V}_U(\mathbf{Z})=\mathbb{V}\left(\mathbb{E}\left[\hat\beta_{\mid \mathbf{Y}(0),\mathbf{Y}(1),\mathbf{Z}}\right]\ \mid \mathbf{Z},N_1\right)+\mathbb{E}\left[\mathbb{V}\left(\hat\beta\middle|\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{Z}\right)\ \mid \mathbf{Z},N_1\right] \tag{C.2}$$

$$=\mathbb{E}\left[\mathbb{V}\left(\hat\beta\middle|\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{Z}\right)\ \mid \mathbf{Z},N_1\right]$$

where the second line follows since $\hat\beta_{\text{ols}}$ is unbiased. By Lemma 4, we have:

$$\mathbb{E}\left[\mathbb{V}\left(\hat\beta\middle|\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{Z}\right)\ \mid \mathbf{Z},N_1\right]=\mathbb{E}\left[\frac{N}{N_0\cdot N_1\cdot(N-2)}\sum_{i=1}^{N}(\varepsilon_i-\overline{\varepsilon})^2\ \mid \mathbf{Z},N_1\right]$$

Observe that we can write:

$$\begin{aligned}\sum_{i=1}^{N}(\varepsilon_i-\overline{\varepsilon})^2 &= (\varepsilon-\iota_N\iota_N'\varepsilon/N)'(\varepsilon-\iota_N\iota_N'\varepsilon/N)\\ &= \varepsilon'\varepsilon-2\varepsilon'\iota_N\iota_N'\varepsilon/N+\varepsilon'\iota_N\iota_N\iota_N\iota_N'\varepsilon/N^2\\ &= \varepsilon'\varepsilon-\varepsilon'\iota_N\iota_N'\varepsilon/N.\end{aligned}$$

Hence:

$$\begin{aligned}\mathbb{V}_U(\mathbf{Z}) &= \frac{N}{N_0\cdot N_1\cdot(N-2)}\mathbb{E}\left[\varepsilon'\varepsilon-\varepsilon'\iota_N\iota_N'\varepsilon/N\,|\,\mathbf{Z},N_0,N_1\right]\\ &= \frac{N}{N_0\cdot N_1\cdot(N-2)}\,\mathrm{trace}\left(\mathbb{E}\left[\varepsilon\varepsilon'-\iota_N'\varepsilon\varepsilon'\iota_N/N\,|\,\mathbf{Z},N_0,N_1\right]\right)\\ &= \frac{N}{N_0\cdot N_1\cdot(N-2)}\left(\mathrm{trace}\left(\Omega(\mathbf{Z})\right)-\iota_N'\Omega(\mathbf{Z})\iota_N/N\right)\end{aligned}$$

which establishes (3.11). Finally, we prove the third part of the Lemma. By Lemma 2, $\hat{\beta}_{\text{ols}}$ is unbiased conditional on $\mathbf{Z},\mathbf{W}$, so that by argument like in Equation (C.2) above, we can also write:

$$\begin{aligned}\mathbb{V}_U(\mathbf{Z}) &= \mathbb{V}\left(\mathbb{E}\left[\hat{\beta}_{\,|\,\mathbf{Z},\mathbf{W}}\right]\ |\,\mathbf{Z},N_1\right)+\mathbb{E}\left[\mathbb{V}\left(\hat{\beta}\middle|\,\mathbf{Z},\mathbf{W}\right)\ |\,\mathbf{Z},N_1\right]\\ &= \mathbb{E}\left[\mathbb{V}\left(\hat{\beta}\middle|\,\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{Z}\right)\ |\,\mathbf{Z},N_1\right]\end{aligned}$$

which equals $\mathbb{E}\left[\mathbb{V}_R\left(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{Z}\right)\ |\,\mathbf{Z},N_1\right]$ by (C.2). □

Suppose Assumptions 1 holds with $\Omega(\mathbf{Z})$ satisfying $\mathrm{trace}(\Omega(\mathbf{Z}))/N\to c$ for some $0<c<\infty$, and $\iota_N'\Omega(\mathbf{Z})\iota_N/N^2\to 0$, and Assumption 3 holds with $N_1/N\to p$ for some $0<p<1$. Then

$$N\cdot(\mathbb{V}_{\text{INC}}-\mathbb{V}_U(\mathbf{Z}))\xrightarrow{p}0,\qquad \text{and}\ \ N\cdot\mathbb{V}_{\text{INC}}\xrightarrow{p}\frac{c}{p\cdot(1-p)}.$$

Proof of Lemma 6: We will first show that the second claim in the Lemma holds,

$$N\cdot\mathbb{V}_{\text{INC}}\xrightarrow{p}\frac{c}{p\cdot(1-p)},\tag{C.3}$$

and then show that

$$N \cdot \mathbb{V}_U \xrightarrow{p} \frac{c}{p \cdot (1-p)}, \tag{C.4}$$

which together prove the first claim in the Lemma.

Consider (C.3). By the conditions in the Lemma, $\hat{\alpha}$ and $\hat{\beta}$ are consistent for α and β, and therefore the probability limit of $\hat{\sigma}^2$ is the probability limit of $\sum_{i=1}^N \sum_{i=1}^N \varepsilon_i^2/N$ which is the probability limit of trace$(\Omega(\mathbf{Z})/N)$. Then

$$\text{plim}\,(N \cdot \mathbb{V}_{\text{INC}}) = \text{plim}\left(\frac{1}{N}\sum_{i=1}^N \varepsilon_i^2 \cdot \frac{N^2}{N_0 \cdot N - 1}\right)$$

$$= \text{plim}\left(\frac{\text{trace}(\Omega(\mathbf{Z}))}{N} \cdot \frac{N^2}{N_0 \cdot N_1}\right) = \frac{c}{p \cdot (1-p)}.$$

Now consider (C.4). By the conditions in the Lemma,

$$N \cdot \mathbb{V}_U = \left(\frac{1}{N-2}\,\text{trace}(\Omega(\mathbf{Z})) - \frac{1}{N \cdot (N-2)}\iota_N'\Omega(\mathbf{Z})\iota_N\right) \cdot \frac{N^2}{N_0 \cdot N_1}$$

$$= \frac{N}{N-2} \cdot \frac{1}{N}\,\text{trace}(\Omega(\mathbf{Z})) \cdot \frac{N^2}{N_0 \cdot N_1} - \frac{N}{N-2} \cdot \frac{1}{N^2}\iota_N'\Omega(\mathbf{Z})\iota_N \cdot \frac{N^2}{N_0 \cdot N_1}$$

$$\xrightarrow{p} \frac{c}{p \cdot (1-p)}.$$

□

Proof of Lemma 9: To show the first part of the Lemma, observe that under constant cluster size,

$$\hat{\beta}_{\text{ols}} = \frac{\sum_{m=1}^M (\tilde{Y}_m - \bar{\tilde{Y}})^2 (\tilde{W}_m - \bar{\tilde{W}})}{\sum_m (\tilde{W}_m - \bar{\tilde{W}})^2}$$

where $\tilde{Y}_m = (N/M)^{-1}\sum_{i:\,C_{im}=1} Y_i$, and $\bar{\tilde{Y}} = M^{-1}\sum_m \tilde{Y}_m = \bar{Y}$, and $\bar{\tilde{W}} = \bar{W}$. Therefore, we can apply Lemma 4, treating cluster averages $(\tilde{Y}_m, \tilde{W}_m, \tilde{\epsilon}_m)$ as a unit of observation, which yields the result.

To show the second part, again by Lemma 4, $\hat{\beta}_{\text{ols}}$ is unbiased, so that by the Law of

Iterated Expectations, and the first part of the Lemma,

$$\mathbb{V}_U(\mathbf{Z}) = \mathbb{V}\left(\mathbb{E}\left[\hat{\beta}_{\mid \mathbf{Y}(0),\mathbf{Y}(1),\mathbf{Z}}\right] \mid \mathbf{Z}, M_1\right) + \mathbb{E}\left[\mathbb{V}\left(\hat{\beta}\middle| \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}\right) \mid \mathbf{Z}, M_1\right]$$

$$= \mathbb{E}\left[\mathbb{V}\left(\hat{\beta}\middle| \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{Z}\right) \mid \mathbf{Z}, M_1\right]$$

$$= \mathbb{E}\left[\frac{M}{(M-2)\cdot M_0 \cdot M_1}\sum_{m=1}^{M}\left(\tilde{\varepsilon}_m - \overline{\tilde{\varepsilon}}\right)^2 \mid \mathbf{Z}, M_1\right]$$

Hence, it suffices to show that

$$\mathbb{E}\left[\sum_{s=1}^{M}\left(\tilde{\varepsilon}_s - \overline{\tilde{\varepsilon}}\right)^2\middle| \mathbf{Z}, M_1\right] = \left(\frac{M^2}{N^2}\cdot \text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right) - \frac{M}{N^2}\iota'\Omega(\mathbf{Z})\iota\right).$$

Note that in general $\mathbf{C}\iota_M = \iota_N$, and under Assumption 8, it follows that $\mathbf{C}'\mathbf{C} = (N/M)\cdot I_M$. We can write

$$\tilde{\varepsilon}_m = \left(\mathbf{C}'\mathbf{C}\right)^{-1}\mathbf{C}'\varepsilon = \frac{M}{N}\mathbf{C}'\varepsilon, \quad \text{and} \quad \overline{\tilde{\varepsilon}} = \frac{1}{M}\iota_M'\left(\mathbf{C}'\mathbf{C}\right)^{-1}\mathbf{C}'\varepsilon = \frac{1}{N}\iota_N'\varepsilon,$$

so that

$$\sum_{m=1}^{M}\left(\tilde{\varepsilon}_m - \overline{\tilde{\varepsilon}}\right)^2 = \left(\frac{M}{N}\mathbf{C}'\varepsilon - \frac{1}{M}\iota_M\iota_N'\varepsilon\right)'\left(\frac{M}{N}\mathbf{C}'\varepsilon - \frac{1}{M}\iota_M\iota_N'\varepsilon\right)$$

$$= \left(\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon\right)'\left(\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon\right).$$

$$= \varepsilon'\left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)'\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon.$$

Thus

$$\mathbb{E}\left[\sum_{m=1}^{M}\left(\tilde{\varepsilon}_s - \overline{\tilde{\varepsilon}}\right)^2\middle| \mathbf{Z}, M_1\right] = \mathbb{E}\left[\varepsilon'\left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)'\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon\middle| \mathbf{Z}, M_1\right]$$

$$= \text{trace}\left(\mathbb{E}\left[\left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)'\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\varepsilon\varepsilon'\middle| \mathbf{Z}, M_1\right]\right)$$

$$= \text{trace}\left(\left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)'\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\Omega(\mathbf{Z})\right)$$

$$= \text{trace}\left(\left(\frac{M}{N}\mathbf{C}' - \frac{1}{N}\iota_M\iota_N'\right)\Omega(\mathbf{Z})\left(\frac{M}{N}\mathbf{C} - \frac{1}{N}\iota_N\iota_M'\right)'\right)$$

$$= \frac{M^2}{N^2}\cdot \text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right) - \frac{M}{N^2}\cdot \iota_N'\Omega(\mathbf{Z})\iota_N. \qquad \square$$

Proof of Theorem 10: We show

$$N \cdot \mathbb{V}_U \xrightarrow{p} \frac{c}{N_m^2 \cdot p \cdot (1-p)},$$

and

$$N \cdot \mathbb{V}_{\text{INC,CR}} \xrightarrow{p} \frac{c}{N_m^2 \cdot p \cdot (1-p)},$$

which together imply the two claims in the Theorem. First consider the first claim. The normalized variance is

$$N \cdot \mathbb{V}_U(\mathbf{Z}) = \frac{M^2 \cdot N}{M_0 \cdot M_1 \cdot (M-2) \cdot N^2} \cdot \left(M \cdot \text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right) - \iota'\Omega(\mathbf{Z})\iota\right)$$

$$= \frac{M^2 \cdot N}{M_0 \cdot M_1 \cdot (M-2)} \cdot \left(\frac{M}{N} \cdot \frac{\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right)}{N} - \frac{\iota'\Omega(\mathbf{Z})\iota}{N^2}\right).$$

By the conditions in the Theorem the probability limit of this expression is

$$\text{plim}\left(\frac{M^2 \cdot N}{M_0 \cdot M_1 \cdot (M-2)} \cdot \left(\frac{M}{N} \cdot \frac{\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right)}{N} - \frac{\iota'\Omega(\mathbf{Z})\iota}{N^2}\right)\right)$$

$$= \text{plim}\left(\frac{M^2 \cdot N}{M_0 \cdot M_1 \cdot (M-2)}\right) \cdot \left(\text{plim}\left(\frac{M}{N} \cdot \frac{\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right)}{N}\right) - \text{plim}\left(\frac{\iota'\Omega(\mathbf{Z})\iota}{N^2}\right)\right)$$

$$= \frac{c}{N_m^2 \cdot p \cdot (1-p)}.$$

Next, consider the second claim. Now the probability limit of the model-based variance is

$$\text{plim}\left(N \cdot \mathbb{V}_{\text{INC,CR}}(\mathbf{Z})\right) =$$

$$\text{plim}\left(\frac{M^2 \cdot N}{M_0 \cdot M_1 \cdot (M-2) \cdot N^2} \cdot \left(M \cdot \text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z}, \tilde{\sigma}_\varepsilon, \tilde{\sigma}_S^2)\mathbf{C}\right) - \iota'\Omega(\mathbf{Z}, \tilde{\sigma}_\varepsilon, \tilde{\sigma}_S^2)\iota\right)\right)$$

$$= \text{plim}\left(\frac{M^2 \cdot N}{M_0 \cdot M_1 \cdot (M-2)} \cdot \left(\frac{M}{N} \cdot \frac{\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z}, \tilde{\sigma}_\varepsilon, \tilde{\sigma}_S^2)\mathbf{C}\right)}{N} - \frac{\iota'\Omega(\mathbf{Z}, \tilde{\sigma}_\varepsilon, \tilde{\sigma}_S^2)\iota}{N^2}\right)\right)$$

$$= \frac{1}{N_m \cdot p \cdot (1-p)} \cdot \left(\text{plim}\left(\frac{M}{N} \cdot \frac{\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z}, \tilde{\sigma}_\varepsilon, \tilde{\sigma}_S^2)\mathbf{C}\right)}{N}\right) - \text{plim}\left(\frac{\iota'\Omega(\mathbf{Z}, \tilde{\sigma}_\varepsilon, \tilde{\sigma}_S^2)\iota}{N^2}\right)\right)$$

$$= \frac{1}{N_m^2 \cdot p \cdot (1-p)} \cdot \text{plim}\left(\frac{\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z}, \tilde{\sigma}_\varepsilon, \tilde{\sigma}_S^2)\mathbf{C}\right)}{N}\right)$$

Hence, in order to prove the second claim it suffices to show that $\text{trace}(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}) = \text{trace}(\mathbf{C}'\Omega(\mathbf{Z}, (\tilde{\sigma}_\varepsilon, \tilde{\sigma}_S^2))\mathbf{C})$. The log likelihood function based on the specification (C.1) is

$$L(\sigma_\varepsilon^2, \sigma_S^2|\mathbf{Y}, \mathbf{Z}) = -\frac{1}{2} \cdot \ln\left(\Omega\left(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_S^2\right)\right) - \frac{1}{2} \cdot \mathbf{Y}'\Omega(\sigma_\varepsilon^2, \sigma_S^2)^{-1}\mathbf{Y}.$$

The expected value of the log likelihood function is

$$\mathbb{E}\left[L(\sigma_\varepsilon^2, \sigma_S^2|\mathbf{Y}, \mathbf{Z})\middle|\, \mathbf{Z}\right] = -\frac{1}{2}\ln\left(\Omega\left(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_S^2\right)\right) - \frac{1}{2} \cdot \mathbb{E}\left[\mathbf{Y}'\Omega(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_C^2)^{-1}\mathbf{Y}\right]$$

$$= -\frac{1}{2} \cdot \ln\left(\Omega\left(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_S^2\right)\right) - \frac{1}{2} \cdot \text{trace}\left(\mathbb{E}\left[\Omega\left(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_S^2\right)^{-1}\mathbf{Y}\mathbf{Y}'\right)\right]$$

$$= -\frac{1}{2} \cdot \ln\left(\Omega\left(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_S^2\right)\right) - \frac{1}{2} \cdot \text{trace}\left(\Omega\left(\mathbf{Z}, \sigma_\varepsilon^2, \sigma_S^2\right)^{-1}\Omega(\mathbf{Z})\right).$$

Using Lemma 14, this is equal to

$$\mathbb{E}\left[L(\sigma_\varepsilon^2, \sigma_S^2|\mathbf{Y}, \mathbf{Z})\middle|\, \mathbf{Z}\right] = -\frac{N}{2} \cdot \ln(\sigma_\varepsilon^2) - \frac{M}{2} \cdot \ln\left(1 + N/M \cdot \sigma_S^2/\sigma_\varepsilon^2\right)$$

$$-\frac{1}{2 \cdot \sigma_\varepsilon^2} \cdot \text{trace}(\Omega(\mathbf{Z})) + \frac{\sigma_S^2}{2 \cdot \sigma_\varepsilon^2 \cdot (\sigma_\varepsilon^2 + \sigma_S^2 \cdot N/M)} \cdot \text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right).$$

The first derivative of the expected log likelihood function with respect to σ_S^2 is

$$\frac{\partial}{\partial \sigma_S^2}\mathbb{E}\left[L(\sigma_\varepsilon^2, \sigma_S^2|\mathbf{Y}, \mathbf{Z})\middle|\, \mathbf{Z}\right] = -\frac{N}{2 \cdot (\sigma_\varepsilon^2 + N/M \cdot \sigma_S^2)} + \frac{\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right)}{(\sigma_\varepsilon^2 + \sigma_S^2 \cdot (N/M))^2}$$

Hence the first order condition for $\tilde{\sigma}_S^2$ implies that

$$\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right) = N \cdot (\tilde{\sigma}_\varepsilon^2 + \tilde{\sigma}_S^2 \cdot (N/M)).$$

For the misspecified error-covariance matrix $\Omega(\mathbf{Z}, \tilde{\gamma})$ we have

$$\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z}, \tilde{\gamma})\mathbf{C}\right) = \sum_{m=1}^{M} \left(N_m^2 \cdot \tilde{\sigma}_S^2 + N_m \cdot \tilde{\sigma}_\varepsilon^2\right).$$

By equality of the cluster sizes this simplifies to

$$\text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z}, \tilde{\gamma})\mathbf{C}\right) = N \cdot \left(\tilde{\sigma}_\varepsilon^2 + \tilde{\sigma}_S^2 \cdot (N/M)\right) = \text{trace}\left(\mathbf{C}'\Omega(\mathbf{Z})\mathbf{C}\right). \qquad \square$$