Why Has Urban Inequality Increased?*

Nathaniel Baum-Snow, Brown University Matthew Freedman, Cornell University
Ronni Pavan, Royal Holloway-University of London

June, 2014

Abstract

The increase in wage inequality since 1980 in the United States has been more pronounced in larger cities, even after accounting for differences in the composition of the workforce across locations. Using Census of Population and Census of Manufacturers data aggregated to the local labor market level, this paper examines the importance of changes in the factor bias of agglomeration economies, capital-skill complementarity, changes in the relative supply of skilled labor, and mutual interactions for understanding the more rapid increases in wage inequality in larger cities between 1980 and 2007. Parameter estimates of a production function that incorporates each of these mechanisms indicate strong evidence of capital-skill complementarity, increasing skill bias of agglomeration economies and declining capital bias of agglomeration economies. Immigration shocks serve as a source of exogenous variation across metropolitan areas in changes to the relative supply of skilled labor versus unskilled labor. The direct relative demand effect of the increasing skill bias of agglomeration economies rationalizes 79-91 percent of the more rapid increases in wage inequality in more populous local labor markets. Interactions between capital-skill complementarity and changes in the factor bias of agglomeration economies have generated outward and inward shifts in the relative demand for skilled labor in larger cities that almost offset.

*Adrian Rubli provided excellent research assistance.

1 Introduction

Since the seminal work of Katz & Murphy (1992), Bound & Johnson (1992) and Juhn, Murphy & Pierce (1993), economists have recognized that the structure of wages in the U.S. economy shifted markedly after 1980 toward greater inequality. Increases in wage inequality have occurred throughout the wage distribution in each decade, except during the 1990s in which there was stability throughout most of the wage distribution below the 75th percentile. Because the relative quantities and prices of skilled labor increased during the 1980s, these studies trace such rising inequality primarily to shifts in the relative demand for skill. Autor, Katz & Kearney (2008) reiterate this explanation and argue that its importance has persisted after 1990. Moretti (2013) and Baum-Snow & Pavan (2013) provide evidence that these relative demand shifts have occurred disproportionately in cities with higher costs of living and greater populations, respectively. Indeed, the elasticity of the college-high school wage ratio with respect to metropolitan area population for urban residents has grown in each of past three decades, from 0.019 in 1980 to 0.051 in the 2005-2007 period. Baum-Snow & Pavan (2013) finds that at least one-quarter of the increase in wage inequality nationwide since 1980 can be attributed to more rapid increases in skill prices in larger cities.

This paper formally investigates the relative importance of several mechanisms that may have generated more rapid increases in skill prices in larger cities. In particular, we examine the roles of capital-skill complementarity, changes in the nature of agglomeration spillovers in production, relative labor supply shifts that have differed across local labor markets and mutual interactions for generating this greater increase in wage inequality in larger cities over time. We employ a unified model that simultaneously incorporates these demand side mechanisms. This model makes use of a constant elasticity of substitution production function similar to that in Griliches (1969), Krusell et al. (2000) and Lewis (2011) with capital, skilled labor and unskilled labor as factors of production. To this standard specification, we add agglomeration economies that are allowed to be factor biased. Using factor quantity and price data in manufacturing for core based statistical areas (CBSAs) from 1980 to 2007, we estimate parameters of this production technology that capture elasticities of substitution between capital, skilled labor and unskilled labor. For econometric identification, we make use of immigration shocks as a source of exogenous variation across local labor markets in changes in the supply of skilled relative to unskilled labor, as in Card (2001) and Lewis (2011). Our analysis uses publicly available information about capital stocks in manufacturing aggregated to the CBSA level and public use census micro data.

Parameter estimates strongly indicate the existence of capital-skill complementarity, an increase in the bias of agglomeration economies toward skilled labor, little change in the bias toward unskilled labor, and a decline in the bias toward capital between 1980 and 2007. Decompositions implied by equilibrium conditions of the model reveal that the increase in the factor bias of agglomeration economies toward skilled workers is central for generating the increasingly positive relationship between skilled wage premia and city size among manufacturing workers since 1980. This operates primarily through a direct effect, with a small additional increase coming from interactions between this increased skill bias of agglomeration economies and capital skill complementarity. The greater complementarity between capital and skilled labor than capital and unskilled labor has generated more rapid capital accumulation in larger cities to keep up with the relative increases in the productivity of skilled workers in these locations. However the resulting relative increases in skilled labor demand in larger locations are almost offset by declines in the demand for skill that have come because of the greater increased productivity of skilled workers in larger locations than productivity of capital. That is, while changes in the factor bias of agglomeration economies interact with capital-skill complementarity to generate differential shifts in labor demand across locations, these differential shifts almost cancel out to leave only a small net positive effect. As a result, the direct effect of increases in the factor bias of agglomeration economies toward skilled labor accounts for 79-91 percent of the increasingly positive relationship between skilled-unskilled wage gaps and city size between 1980 and 2007. Given evidence in Baum-Snow & Pavan (2013) that the more rapid growth in inequality in larger cities has fed through to explain at least one-quarter of the nationwide rise in wage inequality since 1980, this paper's results indicate that the greater skill bias of agglomeration economies is an overlooked mechanism that has driven at least 20 percent of the nationwide increase in wage inequality since 1980.

Our evidence showing the existence of capital-skill complementarity is in line with results elsewhere in the literature, as in Goldin & Katz (1996), Autor, Katz & Krueger (1998), Krusell et al. (2000), Autor, Levy & Murnane (2003) and Dunne et al. (2004). Unlike these prior studies, however, we make use of cross-sectional empirical variation coupled with plausibly exogenous identifying variation across local labor markets to aid in recovery of our estimates. In these regards, this analysis most resembles that in Lewis (2011). However, our investigation examines a broader set of firms and capital stocks, though with more aggregation. Moreover, we recover the first estimates of specific production function parameters that govern capital-skill complementarity using panel data and exogenous shocks to local labor markets for econometric identification. The theoretical framework underlying our analysis is similar to that in Krusell et al. (2000). Our structural estimates are used to perform the first accounting in the literature of the extent to which capital-skill complementarity has interacted with differences in fundamentals across local labor markets to generate cross-sectional variation in wage inequality at the local labor market level.

In a broad sense, our evidence indicates the importance of considering the operations of local labor markets for understanding nationwide trends in wage inequality. Because, even with constant returns to scale, agglomeration economies render production technologies to be different across local markets, failure to consider local labor markets and local heterogeneity in production technologies may lead to model misspecification. Moreover, more variation in the data and sources of econometric identification are available at the local labor market level than nationally. Therefore, we hope that this analysis sparks additional research that drills deeper into the ways in which local heterogeneity in production processes may have influenced recent changes in the wage structure.

The remainder of this paper is structured as follows. Section 2 lays out the theoretical framework, including the production technology whose parameters we estimate. Section 3 presents the data and provides a descriptive picture of the changes in wage inequality since 1980 which points to the importance of considering local labor markets. Section 4 discusses identification and estimation. Section 5 discusses the results. Finally, Section 6 concludes.

2 Theoretical Framework

Our primary goal is to evaluate the relative importance of changes in the factor bias of agglomeration economies, capital-skill complementarity, changes in the relative supply of skilled versus unskilled labor, and interactions between these mechanisms, to understand changes in patterns of wage inequality across local labor markets since 1980.

In order to analyze these various potential causes of relative changes in wage inequality across different local labor markets, we begin with a standard nested constant elasticity of substitution production technology that incorporates capital-skill complementarity. We augment this standard specification to additionally incorporate agglomeration economies that may be factor biased. The following resulting specification is a generalization of the technology estimated in Krusell et al. (2000):

$$Y_j = A_j \left[c D_j^{\sigma\mu_u} U_j^{\sigma} + (1-c) \left(\lambda D_j^{\rho\mu_k} K_j^{\rho} + (1-\lambda) D_j^{\rho\mu_s} S_j^{\rho} \right)^{\frac{\sigma}{\rho}} \right]^{\frac{1}{\sigma}}. \tag{1}$$

In (1), U_j is unskilled labor efficiency units, S_j is skilled labor efficiency units and K_j is capital, all as chosen by firms in location j. These inputs combined with total factor productivity (TFP) A_j produce output Y_j. D_j denotes the location specific agglomeration force, which can be measured using metropolitan area population level or population density. We observe some measure of all variables in (1) except A_j in the aggregate data, though potentially with some error. Because this technology is constant returns to scale, one can estimate its parameters with data aggregated to the local labor market level. However, because D_j is exogenous and differs across local labor markets, it would not be possible to estimate parameters of this technology using more highly aggregated data.

In (1), the elasticity of substitution between capital or skilled labor and unskilled labor is $\frac{1}{1-\sigma}$ while that between capital and skilled labor is $\frac{1}{1-\rho}$. If capital-skill complementarity exists then $\sigma > \rho$. If either σ or ρ are equal to zero, the corresponding nesting is Cobb-Douglas. Agglomeration forces are governed by μ_k, μ_s and μ_u. If these parameters are equal, agglomeration is factor neutral. Skill-biased agglomeration requires that $\mu_s > \mu_u$ and $\mu_s > \mu_k$. We can think of changes in the skill bias of agglomeration forces as capturing a particular type of directed technical change, as in Acemoglu (1998).

In order to utilize and eventually estimate this production function, we assume that firms cost minimize and that each location j represents a local labor market. We also assume a national market for capital. This treatment is thus most consistent with K capturing capital equipment rather than capital structures. First order conditions from cost minimization are totally differentiated to analyze their changes over time. In doing so, we assume that the rental market for capital, local wages, input quantities, total factor productivity and the extent to which agglomeration economies are biased toward each factor can all vary over time. All other parameters are assumed to be fixed.

The constant returns to scale assumption opens up the reasonable observation that it is just as good to estimate parameters of this production technology using data aggregated to the metropolitan area or national levels. Because of the likely existence of agglomeration economies, we think it important to at least use data aggregated to the metro area level. Doing so distinguishes this research from most existing studies, most notably Krusell et al. (2000), which do not distinguish between local labor markets. In addition, there are many different ways of specifying the agglomeration force D_j, which includes linkages within and across industries. Greenstone, Hornbeck and Moretti (2010) demonstrate that such cross-industry linkages are likely important to firms' TFP, though they do not evaluate the extent to which agglomeration forces are biased toward a particular factor of production.

Combining the two first order conditions with respect to labor for cost minimization results in an inverse relative labor demand equation that relates the relative wages of skilled versus unskilled workers to relative input quantities. This is a generalization of the primary estimation equations used in Ciccone & Peri (2005), Autor, Katz & Kearney (2008) and others, as our specification of the production technology nests their two-factor models. This equation is particularly useful because it lays out a natural linear decomposition of the sources of the change in wage inequality over some time period.

$$d\ln\frac{w_j^s}{w_j^u} = \sigma d(\mu_s-\mu_u)\ln D_j+(\sigma-1)d\ln\left(\frac{S_j}{U_j}\right)+(\sigma-\rho)\omega_j^c d\ln\left(\frac{K_j}{S_j}\right)+(\sigma-\rho)\omega_j^c d(\mu_k-\mu_s)\ln D_j \quad (2)$$

This equation contains all three of the possible channels incorporated in the model to explain changes in inequality in each local labor market over time, plus an interaction. In particular, inequality can increase as a result of an increase in skill biased agglomeration forces, a decrease in the relative supply of skilled workers, because of an increase in the supply of capital relative to skilled workers or relative increases in the complementarity of city size and capital, assuming capital-skill complementarity ($\sigma > \rho$). In the third and fourth terms, ω_j^c denotes the share of capital in the theoretical factor of production that combines capital and skilled labor, and is specified more carefully below. Much of the labor literature focuses only on the second term in this equation, while the literature investigating capital-skill complementarity additionally investigates the third term, though typically in a time-series rather than cross-sectional context. This is the first paper to additionally consider the components of (2) that capture changes in the factor bias of agglomeration economies.

It is instructive to consider each term in (2) carefully, as this equation forms the basis for decompositions performed at the end of this paper. Because our empirical implementation below is better suited to decomposing variation across local labor markets in trends in inequality, rather than the overall secular trend, our discussion focuses on such cross-sectional variation. First, if skilled workers have become relatively more productive in larger cities, higher skill prices ensue in these cities assuming sufficient substitutability between skilled and unskilled labor. Note that with a Cobb-Douglas production technology, which is often assumed but has only rarely been empirically supported, this agglomeration channel does not matter for wage inequality. In the Cobb-Douglas environment, increases in the relative productivity of skilled labor are balanced by offsetting increases in the relative demand for the sufficiently complementary unskilled labor given fixed input quantities. Second, the relative price of skill increases in locations in which the relative quantity (supply) of skill decreases. Third, inequality increases more in locations where capital intensity increases if $\sigma > \rho$. Increases in the relative supply of capital raise the relative productivity of skilled workers, feeding through into greater demand for their services. Of course, understanding reasons for changes in the endogenous object $d\ln(K_j/S_j)$ must be part of the analysis of this third effect. Finally, holding factor quantities constant, inequality increases given capital-skill complementarity if the capital bias of agglomeration economies increases more rapidly than their skill bias. This interaction effect captures the increase in demand for skilled labor that comes with the relative increases in the productivity of the complementary input.

In practice, decompositions using (2) to understand why wage inequality has increased more rapidly in larger cities will come down to evaluating the relative importance of changes in the factor-bias of agglomeration economies and capital-skill complementarity coupled with more rapid increases in the relative supply of capital in larger cities. Evidence in Baum-Snow & Pavan (2013) indicates that changes in the relative supply of skills had a negligible impact on variation in changes in wage inequality across local labor markets of different sizes. Indeed, Baum-Snow and Pavan (2013) demonstrates that the relative quantity of skilled labor in large relative to small cities has changed very little since 1980, evidence which is echoed below in this paper. Therefore the narrative in this paper primarily examines the importance of various elements of capital skill complementarity relative to a residual explanation to which we affix a label of changes in the skill bias of agglomeration economies. We leave the development of an understanding of the particular micro-foundations through which such changes have occurred to future research.

It is crucial to account for the endogeneity of $d\ln(K_j/S_j)$ in (2).[1] One way of handling this endogeneity is to express $d\ln(K_j/S_j)$ in terms of exogenous objects and substitute in for it in (2). In doing so, we also see how changes in the factor bias of agglomeration economies interacts with capital-skill complementarity to generate greater increases in wage inequality in larger cities. We also treat log CBSA population $\ln(D_j)$ as exogenous.[2] Fully differentiating the first order condition from profit maximization with respect to capital yields the following expression, which can be used

[1] Of course it is also crucial to account for the endogeneity of $d\ln(S_j/U_j)$. But this is done solely through our empirical implementation, detailed in Section 4.

[2] While it is important to consider the likelihood that larger cities may have different unobserved attributes like workforce composition than smaller cities, the large empirical literature on agglomeration economies almost universally treats city size as exogenous. The few attempts in the literature to account for potentially endogenous city size with geological and historical instruments typically yield results that are almost identical to those in which such endogeneity concerns are not considered. See Combes et al. (2010) for a review.

to resolve this endogeneity problem:

$$\begin{aligned} d\ln\frac{K_j}{S_j} &= \frac{d\ln v - d\ln A_j}{(\sigma-\rho)\omega_j^c(1-\omega_j^{cs}) - (1-\rho)(1-\omega_j^{cs}\omega_j^c)} \\ &+ \frac{(1-\sigma)(1-\omega_j^{cs})d(\mu_s-\mu_u) + ((1-\sigma)\omega_j^{cs}\omega_j^c + (\sigma-\rho)\omega_j^c+\rho)d(\mu_s-\mu_k) - d\mu_s}{(\sigma-\rho)\,\omega_j^c(1-\omega_j^{cs}) - (1-\rho)(1-\omega_j^c\omega_j^{cs})}\ln D_j \\ &+ \frac{(1-\sigma)(1-\omega_j^{cs})}{(\sigma-\rho)\omega_j^c(1-\omega_j^{cs}) - (1-\rho)(1-\omega_j^{cs}\omega_j^c)} d\ln\frac{S_j}{U_j} \end{aligned} \tag{3}$$

In this equation, v denotes the rental price of capital. Our assumption of a national capital market means that v is not indexed by location. This assumption of perfectly elastic capital supply to each local labor market is crucial to pin down an expression for the equilibrium quantity of capital.[3] ω_j^{cs} and ω_j^c are output shares that can be calculated with the data. $\omega_j^{cs} = \frac{(1-c)\left(\lambda D_j^{\rho\mu_k}K_j^\rho+(1-\lambda)D_j^{\rho\mu_s}S_j^\rho\right)^{\frac{\sigma}{\rho}}}{cD_j^{\sigma\mu_u}U_j^\sigma+(1-c)\left(\lambda D_j^{\rho\mu_k}K_j^\rho+(1-\lambda)D_j^{\rho\mu_s}S_j^\rho\right)^{\frac{\sigma}{\rho}}}$ is the share of the theoretical capital-skill composite factor in production and $\omega_j^c = \frac{\lambda D_j^{\rho\mu_k}K_{jk}^\rho}{\lambda D_j^{\rho\mu_k}K_{jk}^\rho+(1-\lambda)D_j^{\rho\mu_s}S_{jk}^\rho}$ is the share of capital in this capital-skill composite. We will recover these objects empirically by using the facts that the capital share $\frac{vK_j}{Y_j} = \omega_j^c\omega_j^{cs}$ and the unskilled labor share $\frac{w_j^uU_j}{Y_j} = 1-\omega_j^{cs}$.

Given that σ and ρ are both always less than 1, the coefficients on $d\ln v - d\ln A_j$ and $d\ln\frac{S_j}{U_j}$ in (3) are always negative. In the first term, reductions in the price of capital promote capital intensity, as do positive TFP shocks.[4] Therefore, the gradient of $\ln\frac{K_j}{S_j}$ with respect to city size would increase with relatively positive TFP shocks in larger cities or (trivially) if the relative number of skilled workers decreases. The coefficient on $\ln D_j$ may be positive or negative, but it is equal to zero if the factor biases of agglomeration forces do not change. This agglomeration effect tends to be positive when $d\mu_s$ is positive and larger than $d\mu_k$ and $d\mu_u$. That is, increases in skill biased agglomeration forces increase wage inequality through two channels if there exist capital skill complementarities. In addition to the direct effect seen in (2), this indirect effect operates through enhancing capital intensity in larger locations, thereby further increasing the price of skill in such locations because of capital-skill complementarity.

While (3) is the basis for deriving the estimating equations discussed in Section 4, it is instructive to consider the following alternative representation of (3).

$$d\ln\left(\frac{K_j}{S_j}\right) = \frac{1}{1-\rho}d\ln\left(\frac{w_j^s}{v}\right) + \frac{\rho}{1-\rho}d\left(\mu_k-\mu_s\right)\ln D_j$$

The first term simply reflects the relative price effects, where $\frac{1}{1-\rho}$ is the elasticity of substitution

[3] It is true that structures capital is not supplied at the same price in all locations. However, Albouy (2014) determines that land, which is a large component of capital structures, only accounts for about 2.5 percent of input costs among firms producing tradeable goods. Krusell et al. (2000) estimate that capital structures account for 11.7 percent of input costs among all industries.

[4] In the empirical implementation, the variation across local labor markets in capital intensity because of this channel ends up as part of an error term since TFP is unobserved.

between capital and skilled labor. Potential reasons for which skilled labor may have become relatively more costly, or its marginal product has increased, in larger cities can be seen in (3). These locations may have experienced more rapid increases in factor unbiased agglomeration economies, skill-biased agglomeration economies, or declines in the relative supply of skilled labor. Once these price effects are held constant, a more direct agglomeration mechanism becomes clearer. Holding factor prices constant, an increase in the capital bias of agglomeration forces increases relative capital intensity whereas an increase in their skill bias decreases relative capital intensity, as is intuitive, provided that $0 < \rho < 1$, or capital and skill are sufficiently substitutable.

This section has laid out the equilibrium equations generated by firm optimization given the production technology in (1). We note that examinations of the evolution of cross-sectional relationships between relative factor prices $\ln(w_j^s/w_j^u)$ and quantities $\ln(S_j/U_j)$, $\ln(K_j/S_j)$ and $\ln D_j$ are informative about mechanisms driving the strengthening equilibrium relationship between wage gaps and city size.

3 Data and Descriptive Evidence

3.1 Data

To estimate the model's parameters, we require information about capital stocks, skilled and unskilled labor and the input price per unit of skilled and unskilled labor for each CBSA nationwide in multiple time periods.

To construct information about skilled and unskilled worker quantities and wages, we use the national 5 percent public use micro data samples for the 1980, 1990 and 2000 Censuses of Population and the 2005, 2006 and 2007 American Community Surveys (ACS) pooled into a national 3 percent sample (Ruggles et al., 2010). We select 2007 as the terminal year for worker data in order to match the timing of the available capital and output data, as is described below. We combine both 1% metro public use micro data samples from the 1970 census into a 2% sample in order to help build instruments, as is explained in Section 4.2 below. We require large sample sizes in order to build data for individual CBSAs, the smallest of which have under 50,000 residents. We use information for all individuals who report having positive wage and salary income, who usually worked at least one hour per week, and worked at least one week in the year prior to the survey.[5] Most of our analysis uses only those who report working in manufacturing.

We use the 922 Core Based Statistical Areas (CBSA) as of year 2003. These collections of counties replace Metropolitan Statistical Areas as the primary measure of local labor markets used by the U.S. government after 2001. They include both "micropolitan" and "metropolitan" areas, of which 380 had fewer than 50,000 residents in 1980 and 234 had 50,000-100,000 residents in 1980. One challenge with using census micro data for this analysis is that its geographic units rarely line up to CBSA definitions. The 1970 and 1980 censuses include county group (CG) identifiers whereas

[5] Decennial censuses ask about the prior calendar year whereas the American Community Surveys ask about the prior 12 months.

later censuses and the ACS report public use microdata areas (PUMAs).[6] Each CG and PUMA has a population of at least one hundred thousand and a geography that typically does not correspond to county boundaries. To assign sampled individuals in each decennial census to CBSAs, we make use of population allocation factors between CGs or PUMAs and counties published by the Census Bureau. For CGs and PUMAs that straddle a CBSA boundary, we allocate the fraction of each individual in the CG or PUMA given by the reported allocation factor to each CBSA unit. This means that some individuals are counted multiple times in our data, but with overall weights that still add to their contributions to the U.S. population.

For the majority of our analysis, we assign those with more than 12 years of education to the skilled group (S) and those with 12 years of education or less to the unskilled group (U). We drop individuals with imputed education. Each hour worked is considered one raw unit of labor, for which we calculate average wages in each CBSA, w_j^s and w_j^u. We also build an efficiency units measure which attempts to control for changes in the composition of the workforce within the skilled and unskilled categories. To calculate the number of efficiency units each worker contributes to the stock of skilled or unskilled labor, we regress the log hourly wage on a series of indicator variables for age, sex, race, years of education, occupation, CG of residential location, and country of birth in 1980 separately for each skill group. We include location because, as discussed above, differences in agglomeration economies and natural advantages generate variation in worker productivity across locations. We interpret the regression coefficients on worker attributes as the productivity of each element of observed skill within the broader skill classes. We use the coefficients on observed individual characteristics from these regressions, β_{1980}^U and β_{1980}^S, in all later years to predict the number of labor efficiency units associated with each worker. In particular, we assign $\exp(X_{it}\widehat{\beta}_{1980}^M)$ efficiency units of labor to each hour worked by individual i in year t in broad skill group M.[7] We maintain the 1980 weights β_{1980}^U and β_{1980}^S for later years to prevent these weights from changing endogenously in response to changes in labor market conditions. This amounts to assuming that the quantity of efficiency units of labor provided by each observed skill group within each broader skill classification does not change over time. We measure the prices of one efficiency unit of skilled and unskilled labor in each CBSA directly as means in the data.[8] Wage calculations exclude observations with imputed labor supply or income information. Implied hourly wages below 75% of the national minimum wage are also not incorporated. As is discussed further in Section 4.2 below, we also use population census data to build information about immigration flows to each CBSA by skill level. These flows are used as a basis for constructing instruments.

We use data from the semi-decadal Census of Manufacturers to construct information on capital stocks and total output in manufacturing. The Census of Manufacturing reports capital investment,

[6] The 1990 and 2000 census use different PUMA geographic definitions. The 2005-2007 ACS data sets use the 2000 census PUMA definitions.

[7] Technically we should also take into account the Jacobian transformation component from the prediction uncertainty. However, since our analysis is in logs, this component gets subsumed into a constant term.

[8] Another way to measure labor inputs would be to use information directly from the Census of Manufacturers, treating non-production workers as "skilled" and production workers as "unskilled." Unfortunately, reported hours are not broken out for these two worker types in the aggregate data in all years.

the wage bill, total value added and various other aggregate manufacturing statistics by county in 1982, 1987, 1992 and 1997. In 2002 and 2007, it reports these objects for each CBSA.[9] The information about capital combines equipment and structures capital. Using these data together with national capital price indices and depreciation rates reported by the Bureau of Labor Statistics (BLS), we construct CBSA-specific measures of the capital stock by year using the perpetual inventory method. To begin, we construct a time series of capital investments from 1948 to 2007 by interpolating reported investments for intercensal years and assuming constant investment at 1982 levels in prior years. We adjust using deflators and depreciation rates reported by the BLS by sector within manufacturing aggregated using sectoral shares, following the methodology laid out in Harper (1999). Annual capital investments are combined with deflators to construct the real CBSA capital stock in each year. Although the resulting capital shares already closely resemble national averages, we normalize the stocks in each survey year in order to have exactly the same shares as the national data on average across CBSAs. Shares are calculated as the rental price of capital multiplied by the stock of capital divided by the same quantity plus the wage bill.[10] Due to data suppression in counties or CBSAs with only a few manufacturing firms, we do not have capital or factor share information for 150 CBSAs in 1980, 182 CBSAs in 1990, 190 CBSAs in 2000 and 263 CBSAs in 2007. If capital data is unavailable in 1982, we impute backwards from 1987 instead. We set capital information to missing for all CBSAs with capital stocks first reported after 1987 or with only one year of capital data.

Table A1 presents summary statistics.

3.2 Basic Empirical Patterns

The results in Table 1 provide a broad motivation for this analysis. Each entry in the first four columns of Table 1 is the average wage gap for the average hour of work among "skilled" versus "unskilled" workers living in a 2003 definition CBSA in various years. Each column uses a different definition of skilled and unskilled workers, indicated in column headers. Panel A shows wage gaps for all workers whereas Panel B shows wage gaps for manufacturing workers only.

Table 1 shows that the well known rise in wage gaps between skilled and unskilled workers is a remarkably robust phenomenon. This rise has happened over every decade since 1980, does not depend on how skill groups are defined and appears within manufacturing as well as among all workers. While the levels of wage gaps differ across skill definitions, the increases in wage gaps between 1980 and 2007 are between 0.15 and 0.22 for all workers and 0.14 to 0.20 for manufacturing workers. Indeed, while manufacturing workers always have greater wage gaps than the full working population, for no definition of skill does the 1980-2007 increase in these gaps differ by more than 0.02 when comparing across these two groups.[11] Because trends in wage gaps are similar across

[9] The 2012 Census of Manufacturers CBSA data is not yet publicly available.

[10] Factor shares at the national level also incorporate materials, energy and services. As such, we first renormalize to include only capital and labor.

[11] Manufacturing made up 25 percent of urban hours worked in 1980, 20 percent in 1990, 17 percent in 2000 and 14 percent in 2005-7.

skill definitions, we focus on the definition in Column 1 for the remainder of this analysis. This definition best balances the data in 1980, when 42 percent of working hours amongst all workers and 31 percent amongst manufacturing workers were in the skilled group, while maintaining inclusion of workers with all levels of education in the sample. By 2007, 62 percent of all working hours and 53 percent of manufacturing hours were skilled by this definition.

The final column of Table 1 shows elasticities of wages with respect to 1980 CBSA population in each study year. These results indicate that the city size wage premium increased during the 1980s but remained relatively stable thereafter for manufacturing and all workers alike. Interpreted in the context of a Rosen (1979) & Roback (1982) type model, as in Albouy (2014), this is evidence of an increase in the magnitude of agglomeration economies among firms producing tradeable goods during the 1980s. The evidence presented below of the rising complementarity between skill and city size thus justifies both this overall rise in agglomeration economies during the 1980s and a decline in the importance of other mechanisms generating agglomeration economies since 1990.

Figures 1 and 2 show that a positive relationship between skilled-unskilled wage gaps and city size has largely developed since 1980. These figures are constructed using average wages by skill in each of the 922 CBSAs in our primary sample. Each plot is of predicted values from a local polynomial regression of the variable listed in the panel header on log 1980 CBSA population. Because the distribution of city sizes has a thin right tail, note that the density of the data declines moving from left to right in these plots.

Figure 1 Panel C shows that among all workers, wage gaps increased on average in CBSAs of all sizes in each decade since 1980. However, this increase was much greater in larger cities. Though no relationship exists between city size and wage gaps among cities with populations of less than $e^{11} = 60,000$ in any year, a clear positive relationship between these two variable among larger CBSAs strengthens in each year since 1980. (Dots at the left of the graph are for rural areas.). In 1980, the log wage gap in the largest city (New York) was about 0.10 more than in cities of 60,000 people. By 2005-7, this relative gap increased to 0.28.

Evidence in Panels A and B of Figure 1 show that this increasingly strong relationship between wage gaps and city size was driven both by increases in the gradient among skilled workers and declines in the gradient among unskilled workers. Panel A shows that skilled workers always enjoyed higher wages in larger cities, but that this relationship strengthened in each decade since 1980. This is *prima facie* evidence of increases in the complementarity between agglomeration economies and skill over time, or $d\mu_s > 0$ in the context of our model. Panel B shows the well documented general deterioration of wages for unskilled workers. At the same time, especially during the 1990s, the wage profile for this group gets much flatter with respect to city size. As is documented in Baum-Snow & Pavan (2013), this fed through to little change in the bottom part of the wage distribution during the 1990s. It also potentially indicates evidence of declines in the strength of agglomeration economies among unskilled workers, or $d\mu_u < 0$.

Figure 2 provides exactly the same information as in Figure 1 but for manufacturing workers only. It exhibits all of the same patterns, though stronger. Wage gaps diverge more over each decade in larger cities than in smaller cities across almost the entire city size distribution. Indeed in 1980, the relative log wage gap in the largest CBSA compared to CBSAs with a population of $e^{10} = 22,000$ was 0.17. By 2005-7, this relative gap had grown to 0.43. As with all workers, this strengthening relationship was driven both by increases in the gradient among skilled workers and declines in the gradient among unskilled workers.

Table 2 quantifies the changes in the relationships between relative skill prices or relative factor quantities and city size over time. Given that plots in Figures 1 and 2 Panel C are close to linear and that the model in the previous section implies linear relationships, we focus on average elasticities with respect to city size.[12] To relate our results in Table 2 to those in Table 1, all elasticities are estimated using 1980 CBSA population weights.

The first column of Table 2 quantifies the fact that the elasticity of relative wages with respect to city size faced by the average urban resident has increased in each decade since 1980 among all workers and manufacturing workers alike. Among all workers this elasticity increased from 0.019 to 0.051, whereas among manufacturing workers it increased from 0.030 to 0.072. Some of these increases are because of observed shifts in the compositions of the skilled and unskilled groups. The fourth column, under the "Efficiency Units" header, shows that accounting for shifts in the observed composition of skill groups over time reduces these increases by about 0.01 for all workers and manufacturing workers alike.

Results in the second and fifth columns of Table 2 show that the relationship between relative skill quantities and city size hardly changed since 1980. Indeed, when considering efficiency units, any such changes are negligible, both for all workers and manufacturing workers. This evidence echoes that in Baum-Snow & Pavan (2013). These robust changes in relative prices but small changes in relative quantities indicates that relative labor demand shifts must be central for understanding the increasingly positive relationship between wage inequality and city size over time.[13]

The third column of Table 2 Panel B shows that during three of the four periods studied, larger cities became more capital intensive relative to small cities (Because S_j/U_j hardly changed, we can conclude that increases in K_j/S_j also meant increases in K_j/U_j.). However, large cities still have smaller capital-skilled worker ratios than small cities.[14]

[12] Of course it would be possible to additionally incorporate second order equilibrium relationships into the model. However, we are skeptical that doing so would be instructive because quadratic terms in empirical elasticities of relative factor prices and quantities with respect to city size are not statistically significant in most cases.

[13] Diamond (2013) provides evidence that the 1980 to 2000 change in the fraction of the population with a college degree is positively correlated with 1980 college fraction using metropolitan area level data. Because skill intensive locations tend to have higher populations, this result may seem to be at odds with evidence in Table 2. This relationhip is weaker in our full sample of CBSAs relative to her MSA sample. In addition, the magnitude and sign of the relationship depend on whether it is estimated using shares or log shares and whether population weights are applied. Moreover, Diamond's result does not hold for CBSAs if those with some college education are included in the skilled group.

[14] We are hesitant to compare K_j/S_j in 2005-7 to that in 2000 for two reasons. First, the timing of data collection is different. K_{2005-7} is actually from 2007 and K_{2000} applies to 2002. However, S_{2005-7} actually applies to the 2004-7 period and S_{2000} applies to 1999. Second, sampling for the 2005-7 ACS data sets is based on the 2000 census, so absolute labor quantities are artificially similar to the 2000 data. Our use of K_j/Y_j and S_j/U_j instead of K_j/S_j in most of the empirical work below avoids these measurement problems.

The final column of Table 2 shows a similar pattern when S_j is measured as efficiency units These increases in the elasticity of capital intensity with respect to city size can be interpreted in the context of (3). Given little change in S_j/U_j, these results must either reflect more rapid increases in TFP in larger cities or increases in the skill bias of agglomeration economies. The following section shows how we disentangle the importance of these two mechanisms.

An additional way to summarize factor intensities is to examine how the shares ω_j^c and ω_j^{cs} vary with city size and over time, which we can only calculate for manufacturing. ω_j^c describes the fraction of the composite capital-skill factor made up by capital. As with K_j/S_j, this object is negatively correlated with city size in each year, though over each decade the correlation becomes weaker, from -0.041 in 1980 to -0.018 in 2005-7 (unreported). That is, the capital share of the capital-skill composite factor of production rose more rapidly in larger cities than in smaller cities even during the 2000-2007 period, when $\ln(K_j/S_j)$ became more negatively correlated with city size. ω_j^{cs} describes the fraction of the capital-skill composite in production. This object is positively related with city size in each year with a correlation that remains in the range of 0.020 to 0.033 with no systematic pattern over time.

Table 3 presents regressions of decadal changes in relative factor prices or quantities on city size and decadal dummy variables. These results are intended to capture the average decadal change in the elasticities of these objects with respect to city size. Commensurate with evidence in Table 2, results in Table 3 show that the elasticity of the skilled-unskilled wage ratio with respect to city size significantly increased by about 0.011 each decade, whether for all workers, manufacturing workers, raw units or efficiency units. Amongst all workers, this log wage ratio also experienced secular increases in each study period, with the greatest increase during the 1980s. Among manufacturing workers, the secular increase was more balanced across decades. The elasticity of the relative quantity of skilled labor with respect to city size did not significantly change, except for a small decline in the raw units measure of all workers, though it did experience secular increases in the 1980s and 1990s. Finally, the elasticity of capital intensity with respect to city size significantly increased by about 0.015 over each decade since 1980.

In summary, patterns in the data are consistent with the claim that some combination of capital-skill complementarity and increases in the skill bias of agglomeration economies have been central for generating changes in inequality across local labor markets since 1980. Quantification of these impacts requires estimation of σ and ρ for capital-skill complementarity and $d\mu_s, d\mu_u$ and $d\mu_k$ for changes in the factor bias of agglomeration economies. The following section shows how we recover these parameters.

4 Estimation

In this section, we show how we estimate parameters of the model developed in Section 2. The first step is to derive the structural equations of the model that are feasibly estimated. The second step is to establish econometric identification through isolation of exogenous variation in $d\ln(S_j/U_j)$

through immigration shocks, as in Lewis (2011). Throughout our treatment, we take $\ln(D_j)$ to be exogenous.

4.1 Estimating Equations

Assuming exogeneity of $d\ln(S_j/U_j)$, (3) substituted into (2) forms the first estimable structural equation of interest from the model. In estimation, we account for $d\ln v$ with a separate time fixed effect for each decade and think of $d\ln A_j$ as an i.i.d. stochastic component whose mean gets subsumed into these time fixed effects. In order to estimate all of the model parameters, we need two additional equations.

Manipulating the first order condition with respect to capital and the totally differentiated production function, we derive the second estimating equation that relates capital intensity to skill intensity and market scale.

$$
\begin{aligned}
d\ln\frac{K_j}{Y_j} \quad = \quad & \frac{(1-\omega_j^{cs}\omega_j^c)(d\ln v - d\ln A_j)}{(\sigma-\rho)\omega_j^c(1-\omega_j^{cs})-(1-\rho)(1-\omega_j^{cs}\omega_j^c)} - d\ln A_j \\
& + \frac{(\rho-\sigma)(1-\omega_j^c)(1-\omega_j^{cs})d(\mu_s-\mu_u)+(\sigma\omega_j^{cs}\omega_j^c+(\rho-\sigma)\omega_j^c-\rho)d\mu_k}{(1-\sigma)\omega_j^{cs}\omega_j^c+(\sigma-\rho)\omega_j^c+(\rho-1)}\ln D_j \\
& + \frac{(\rho-\sigma)(1-\omega_j^c)(1-\omega_j^{cs})}{(\sigma-\rho)\omega_j^c(1-\omega_j^{cs})-(1-\rho)(1-\omega_j^{cs}\omega_j^c)} d\ln\frac{S_j}{U_j}
\end{aligned} \tag{4}
$$

This equation is quite similar to (3). The first and third terms in this expression are positive if and only if $\sigma > \rho$, meaning that there is capital-skill complementarity. The sign of the coefficient on $\ln D_j$ is not easily determined. In principle, we could instead use (3) for estimation. However, the formulation given in (4) is more convenient as it will allow us to directly empirically verify that the coefficient on $d\ln(S_j/U_j)$ is positive using a simple linear IV estimator, providing direct evidence of capital-skill complementarity. This is similar to the setup and procedure used in Lewis (2011). Moreover, since capital data is not available in the same years as data on the labor force, absolute labor force quantities from 2005-7 are probably suspect as they depend on the sampling frame from 2000, and we prefer to construct the outcome using objects measured at the same point in time.

Manipulation of first order conditions for cost minimization with respect to capital and skilled labor yields the third estimating equation.

$$
d\ln w_j^s = d\ln v + \rho d\left(\mu_s - \mu_k\right)\ln D_j + (1-\rho)\, d\ln\left(\frac{K_j}{S_j}\right) \tag{5}
$$

This equation is quite intuitive. It says that skilled labor receives higher wage increases when capital gets more expensive, in larger cities if the agglomeration economies become more skill biased, or when capital intensity increases, as regulated by the elasticity of substitution $\frac{1}{1-\rho}$. When using this expression for estimation below, we substitute for $d\ln\left(\frac{K_j}{S_j}\right)$ with (3).

All three estimating equations have the same structure. Their right hand sides can be decomposed into three components. The first component contains elements that are not observed like the change in price of capital and the change in location specific TFP. These variables enter linearly in the function and they are multiplied by a coefficient that depend on the parameters of the model and on a combination of factor shares. The second component is a linear function of the log agglomeration measure. This variable is multiplied by a coefficient that depends on σ, ρ, the factor shares, and the parameters that describe changes in the factor biases of agglomeration forces, $d\mu$. The third component is linear in $d\ln\frac{S_j}{U_j}$, with a coefficient that is function of σ, ρ and the factor shares. It is already apparent from these equations that the likely correlation between the unobservables, like TFP, and the change in the skill ratio $d\ln\frac{S_j}{U_j}$ makes the identification of the parameters more difficult. The following sub-section explains our identification strategy.

4.2 Identification

Assuming exogenous variation across CBSAs in $d\ln(S_j/U_j)$ and $\ln D_j$, the following types of comparisons in the data allow us to identify the parameters of interest. Conditional on city size, the responses of relative wages, skilled wages and capital intensity to variation across CBSAs in relative labor supply shocks provides information about σ and ρ, which are related to the elasticities of substitution in the production technology described in (1). Note that these parameters are over-identified because one source of variation from each of the three estimating equations is being used to identify these two parameters. However, information about capital must be used to help identify ρ. That is, (4) is a central estimation equation. The parameters which govern changes in the factor bias of agglomeration economies, $d\mu_s$, $d\mu_u$ and $d\mu_k$, are identified through comparisons between CBSAs of different sizes. Comparison of two CBSAs of different sizes with the same labor supply shock recovers two of these parameters. The third is separately identified from a TFP time trend because of the pre-determined variation across CBSAs in factor intensities.

The clearest difficulty for successfully recovering model parameters is that the stocks of skilled and unskilled workers in each city at each point in time are equilibrium outcomes, yet the model does not specify a supply side of the labor or capital markets. For capital markets, we adopt the standard assumption that supply to any given local labor market is perfectly elastic. Therefore, identification of parameters requires exogenous variation in changes in the relative supply of skill across metropolitan areas. We achieve such variation through immigration shocks. The idea is that for reasons that are orthogonal to labor market conditions, immigrants are more likely to settle in locations with relativity high numbers of immigrants from their countries of origin. For example, the fact that the Los Angeles CBSA had a relatively large number of Mexican immigrants in 1970 is a good predictor that Mexicans arriving after 1980 are more likely to settle in the Los Angeles metropolitan area than in most other locations. Moreover, Mexican immigrants are more likely to have low skill levels. Especially among less skilled immigrants, historical enclaves are sources of job referrals and general support upon arrival. Therefore, most cities with such enclaves likely have higher amenity values for these immigrants than do other cities. Using this principle, we build a simulated instrument for $d\ln\frac{S_j}{U_j}$, which appears in our three estimation equations (2), (4) and (5), in some cases after substitution for $d \ln \frac{K_j}{S_j}$ using (3).

For our identification strategy to be valid, it must be that productivity shocks experienced by CBSAs after 1980 are not correlated with contemporaneous predicted immigration flows conditional on city size. For such correlations to exist, it would have to be true that the relative size of immigrant enclaves in 1970 predicts such post-1980 CBSA productivity shocks. For example, we must assume that high skilled immigrants settling in the United States prior to 1970 could not anticipate that the CBSAs in which they settled were more likely to have more rapid productivity growth after 1980. This identification assumption would be violated if some unobserved factor predicted both the locations and skill compositions of immigrant enclaves prior to 1970 and post-1980 productivity growth. While both such direct and indirect endogeneity problems seem unlikely, we perform robustness checks below in which we show that the estimates are not influenced by estimating a more flexible version of the model which conditions on more variables but allows us to estimate fewer parameters.

There is considerable debate in the literature about the wisdom of using immigration shocks as a source of exogenous variation in the supply of labor across local labor markets. Borjas (2003) argues that many direct estimates of the effects of immigrants on natives' wages using variation across local labor markets are expected to be small because the more footloose natives move in response to the negative wage pressure brought by immigration induced labor supply shifts. Using national data, Borjas (2003) finds sizable and significant negative effects of immigration on wages of native workers, providing indirect evidence of such induced migration, and that native and immigrant labor are close substitutes. However, Card & DiNardo (2000) and Card (2001) find little direct evidence of such migration responses. In any case, any potential native outflows in response to immigrant inflows would only weaken our first stage, and does not influence the validity of our instrument. As in Card (2001), we show below that our constructed instrument is indeed a strong predictor of changes in relative skill intensity among manufacturing workers.

A more subtle issue worth considering is that immigrant labor may not be a perfect replacement for native labor. This could be because immigrant and native labor are not perfect substitutes or because immigrants are simply less productive than natives given perfect substitutability. Dustmann et al. (2013) and Ottaviano & Peri (2012) provide evidence that immigrants and natives are likely not perfect substitutes and this is why immigration increases native wages in some parts of the wage distribution. This potential lack of substitutability is not a threat to identification for our purposes as long as skilled immigrants are better substitutes for skilled natives than they are for unskilled natives and vice-versa, as seem likely. If this degree of substitutability is different for skilled and unskilled workers, however, immigration induced variation in raw observed S_j/U_j would not accurately reflect the change in efficiency units. This is one reason that we carry out our analysis using efficiency units that incorporate country of origin in the set of observables for which we account, and the efficiency units calculation is done using different weights on country of origin for skilled and unskilled workers. Because our results, presented in the next section, are insensitive to using raw or efficiency units, we are not too concerned that differential differences in unobservables between immigrants and natives in the skilled versus unskilled groups have much influence on our results.

To implement our identification strategy, we begin by calculating shares of immigrants at least 25 years old living in each CBSA separately by education and 19 regions of origin in 1970. Using these shares, we build the predicted number of new arrivals from each region of the world by education to each CBSA during the 1980s, 1990s and between 2000 and 2007. To calculate these new arrivals, we first multiply the stock of immigrants nationwide ages 16-65 in each education group interacted with region of origin and year by the 1970 shares in each CBSA. We then calculate the differences over time in the logs of these predicted CBSA stocks. We use the same variable to instrument for both raw counts and efficiency units.

Parameter estimates are recovered by adding what can be thought of as either a first stage or a relative labor supply equation to the system of three estimating equations specified above. Incorporating variation in immigrant locations across CBSAs yields the following additional equation:

$$\Delta_t \ln \frac{S_j}{U_j} = \delta_t + \alpha_1 \Delta_t \ln \frac{\widehat{S}_j}{\widehat{U}_j} + \alpha_2 \ln \frac{S^{imm}_{jt-1}}{U^{imm}_{jt-1}} + u_{jt} \tag{6}$$

In this equation, Δ_t denotes the difference between periods t and $t-1$, variables with a hat are predicted using 1970 immigrant shares across CBSAs as described above, and the superscript imm indicates that these variables are for stocks of all immigrants. To be consistent with the estimating equations, the outcome variable uses manufacturing workers only, whereas the calculation of $\Delta_t \ln \frac{\widehat{S}_j}{\widehat{U}_j}$ uses all immigrants. Following Lewis (2011), we include the regressor $\ln \frac{S^{imm}_{jt-1}}{U^{imm}_{jt-1}}$ so as to remove any potential spurious correlation between period $t-1$ immigrant stocks and changes in labor factor ratios, thereby isolating only the contribution of predicted new immigrant flows to changes in these factor ratios. Our estimates of α_1 and structural parameters of interest reported below are not sensitive to the inclusion of this control.

Using four cross-sections of data starting in 1980, we estimate parameters of this equation with a panel that has three observations per CBSA. Our estimate of the key parameter α_1 in (6) is 0.21 (or 0.17 when predicting efficiency units) when weighted by 1980 CBSA population, and statistically significant with t-values of more than 3.5 when standard errors are clustered by CBSA.[15]

Table 4 presents regression results indicating that while most identifying variation comes from labor supply shocks for those with high school or less, some exogenous variation among college graduates also exists. Each column in this table is a separate regression of the change in log manufacturing employment with the indicated education on the change in log predicted number of immigrant workers in the same education group and other controls. Coefficients on $\Delta \ln$(Predicted Quantity) roughly decline in magnitude with education, as expected. A 10 percent increase in predicted immigrant quantity leads to an estimated 3.8 percent more manufacturing workers among

[15] Because the other equations in the system include $\ln D_j$ as a regressor, this variable should technically also appear in the first stage equation. However, it is almost orthogonal to $\Delta_t \ln \frac{\widehat{S}_{jt}}{\widehat{U}_{jt}}$ so its inclusion does not influence estimates of α_1.

high school dropouts, 2.5 percent more among high school graduates, and 1.5 percent more among college graduates. Coefficients for the two remaining education groups are not statistically significant. Coefficients on city size are consistently negative, indicating less rapid manufacturing employment growth in larger cities among all types of workers. As we demonstrate below, these coefficients do not significantly differ across education categories.

4.3 Estimation Procedure

For the purposes of estimation, we augment the three structural equations from the model with a linear control for $\ln \frac{S^{imm}_{jt-1}}{U^{imm}_{jt-1}}$ and time fixed effects, since our instrument may only be exogenous conditional on these variables. We think of $d\ln A_j$as a constant time effect plus an idiosyncratic component that is uncorrelated with our immigration shock instruments for $d\ln \frac{S_j}{U_j}$ and city size, and therefore ends up in error terms. Because we take levels and changes in the price of capital to be the same in all locations, time fixed effects also subsume $d\ln v$.

We estimate two versions of the model. In the "sparse" version, we focus on recovering accurate estimates of σ and ρ from coefficients on $d\ln \frac{S_j}{U_j}$, making use only of the exogenous variation available through relative labor supply shifts. In the sparse model, we account for the other terms of the structural equations with time fixed effects fully interacted with $\ln D_j$. To make this empirical model more flexible, we allow these time effects to be different across equations.[16] Controlling for time effects is important both because they control for differences in $d\ln v$ and average TFP growth in different decades and because the instrument is correlated with decade. In estimating the "full" version of the model, we retain differing time fixed effects across equations but no longer interact them with $\ln D_j$. Instead, we specify the coefficients on $\ln D_j$ that are indicated in the structural equations. With the full model, we can additionally identify estimates of changes in the three agglomeration parameters, $d\mu_s$, $d\mu_k$ and $d\mu_u$. Exact specifications of the two sets of structural equations are listed in the Appendix.

We estimate the structural parameters using feasible generalized simultaneous nonlinear least squares (FGNLS) as a simultaneous partially linear system of equations. Each of the structural equations whose parameters we estimate has the following form, with parameter vector ψ.

$$Y = f(\psi, W)\, X + \epsilon$$

We treat W as exogenous and X (or $d\ln(S_j/U_j)$) as endogenous. Because this equation is linear in the endogenous variable X, incorporating first stage Equation (6) which has exogenous predictors Z with coefficients δ yields

$$E(Y|W,Z) = f(\psi, W)\, \delta Z$$

[16] Because $\ln D_j$ is uncorrelated with the instrument, these controls only serve to reduce the standard errors on estimated parameters. When making use of variation in $d\ln(S_j/U_j)$ only to recover estimates of σ and ρ, it is not necessary to account for the heterogeneous coefficients on $\ln D_j$ that are predicted by the model.

and no remaining correlation between W or Z with the error term. Therefore, the parameters ψ are identified and can be consistently estimated (Cai et al., 2006).[17] All reported standard errors are clustered by CBSA.

5 Results

This section presents estimates of σ, ρ, $d\mu_s$, $d\mu_u$ and $d\mu_k$. Using these parameter estimates, we simulate the model and utilize (2) to back out the relative importance of capital-skill complementarity, changes in the factor bias of agglomeration economies, relative labor supply shocks and various interactions of these mechanisms for generating the observed increase in the relationship between city size and wage inequality since 1980.

5.1 Estimates of Linear Coefficients

To provide an intuitive sense of the causal effects on input prices of shocks to the relative supply of skilled labor among manufacturing workers, Table 5 reports IV estimates of the reduced form elasticities of the capital share, the skilled wage, and the skilled-unskilled wage ratio with respect to the skilled-unskilled labor factor ratio, log CBSA population, and appropriate control variables. For these regressions, the predicted change in the immigrant skill mix between periods $t-1$ and t enters as an instrument for $\Delta_t \ln(S_j/U_j)$. Though we present estimates of structural parameters below, these coefficients can be interpreted in the contexts of (2) and (5) after substituting for $d\ln\left(\frac{K_j}{S_j}\right)$, and (4) directly. The signs of estimated coefficients are more informative than their magnitudes since coefficients are predicted by the model to be heterogeneous. Given the structural equations, this simple linear exercise only allows us to estimate the sign of $\sigma - \rho$, indicating whether there is capital-skill complementarity. In the context of the two-factor model commonly estimated in the labor literature, or generalizations with additional factors that enter separably into the production function, the coefficient on $\Delta \ln(S_j/U_j)$ in Columns 4 and 8 of Table 5 is the elasticity of substitution between skilled and unskilled labor. In the following subsection, we discuss structural parameter estimates from the full model. The first four columns of Table 5 use raw counts of hours worked to measure labor quantities and the remaining four columns use labor efficiency units, constructed as described in Section 3.1. The first and fourth columns show first stage results, which are line with the discussion in Section 4.2 above.

Results in Table 5 are consistent with evidence in the literature showing the existence of capital-skill complementarity and an elasticity of substitution between skilled and unskilled labor that is greater than 1. Results do not depend on whether labor is measured in efficiency units. Columns 2 and 5 indicate that exogenous increases in the relative supply of skilled workers led to more capital intensity. As is seen in (4), this positive sign indicates the existence of capital-skill complementarity and mirrors Lewis' (2011) central result. However, we find no evidence of an additional force causing

[17] In our model, the shares ω_j^C and ω_j^{CS} appear in W and may be endogenous as we measure them. We address this potential endogeneity problem in robustness checks below.

firms to employ more capital in larger agglomerations conditional on this response to exogenous relative labor supply shocks. Note that the model has no prediction about the sign of the coefficient on $\ln D_j$ in (4) given the existence of capital-skill complementarity.

Results in the remaining columns of Table 5 indicate that exogenous increases in the relative supply of skilled labor led to declines in both the absolute and relative wages of skilled workers. These negative signs are consistent with theory and do not depend on the potential existence of capital-skill complementarity. Holding the skill mix constant, we see that both wage levels and gaps are higher in more populous metropolitan areas, which reiterates evidence in Baum-Snow & Pavan (2012 & 2013). Based on (5) and (2), these positive coefficients are evidence of increases in the skill bias of agglomeration economies.

A large literature going back to Katz & Murphy (1992) estimates the elasticity of substitution between skilled and unskilled labor using equations like those estimated in Columns 4 and 8 of Table 5. Equation (2), absent the final term, is the structural equation underlying this regression in a two factor model with agglomeration economies. As summarized in Ciccone & Peri (2005), evidence using both time series and cross-sectional variation typically yields elasticity of substitution estimates between skilled and unskilled labor of between 1.3 and 2.0. Our estimates in Table 5 augment the standard empirical specification with a control for city size. Interpreting our estimate of -0.43 in Table 5 Column 4 in the context of a two-factor model, it implies an elasticity of substitution of 2.3, though standard error bands put it between 1.6 and 4.2. Excluding the control for $\ln D_j$ changes the coefficient slightly to -0.52, implying an elasticity of substitution of 1.9. These estimates, which are on the high end of those in the literature, are heavily influenced by data since 1990. Estimating this traditional two-factor specification by decade yields implied substitution elasticities of 1.4 during the 1980s, 2.6 during the 1990s and 3.6 since 2000. Because most of the existing literature uses data prior to 2000, our estimates are thus in line with past estimates and may indicate rising substitutability in production between skill groups.

It is less straightforward to use other coefficients reported in Table 4 to learn about magnitudes of structural parameters. As such, we leave a discussion of remaining parameter estimates to the following sub-section.

5.2 Structural Parameter Estimates

There are clear limitations to the analysis in Table 5. The model tells us that coefficients on $\Delta_t \ln \frac{S_j}{U_j}$ and $\ln D_j$ are not constants. They are heterogeneous across CBSAs because ω_j^C and ω_j^{CS} differ across CBSAs. This makes reduced form coefficients difficult to interpret, and obviates the possibility of recovering estimates of structural parameters of interest from reduced form coefficients.
Table 6 reports estimates of selected production function parameters. The first two columns present estimates from the system of structural equations in which we only estimate heterogeneous coefficients on $d \ln \frac{S_j}{U_j}$. Time dummies fully interacted with city size that are allowed to differ across equations control for the other terms. This model allows us to recover estimates of σ and ρ only, as it only identifies parameters from variation in $d \ln \frac{S_j}{U_j}$. We estimate σ to be 0.85 using raw

labor hours and 0.88 using efficiency units. We estimate ρ to be 0.20 for raw hours and 0.27 for efficiency units, though with standard errors of about 0.15. This is strong evidence of capital-skill complementarity, since $\widehat{\sigma} > \widehat{\rho}$. The formal significance test that $\widehat{\sigma} > \widehat{\rho}$ has a p-value of less than 0.01.

The remaining two columns of Table 6 show estimates from the full model, which additionally identifies the agglomeration parameters $d\mu_u$, $d\mu_s$ and $d\mu_k$. Estimates of σ for the full model are within 0.04 smaller relative to sparse model estimates. Estimates of ρ are -0.33 for raw units and -0.38 for efficiency units in the full model, though they also do not significantly differ from sparse model estimates. We find that agglomeration economies have become significantly more biased toward skilled labor while the agglomeration multiplier on unskilled labor has perhaps slightly declined over time. The agglomeration bias of capital has significantly decreased over recent decades. These estimates provide strong evidence of the rising complementarity between city size and skilled labor, which we demonstrate in the next sub-section has been the necessary condition for generating more rapidly increasing returns to skill in larger cities. The other parameters of potential interest in the full model are the time fixed effects which capture $d\ln v - E[d\ln A_j]$ in the first term of (3). This term is negative for all decades and largest in magnitude during the 1980s.

There are only a few other papers which simultaneously structurally estimate σ and ρ or transformations thereof. Krusell et al. (2000) find evidence of capital-skill complementarity with $\sigma = 0.401$ (se=0.234) and $\rho = -0.495$ (se=0.048) using time-series data from 1963 to 1992 in the United States. Duffy et al. (2004) use data from a panel of countries and find some weak evidence that $\sigma > \rho$, though their estimates are quite imprecise and unstable with respect to model specification and estimation procedure. Both of these studies use lagged values as instruments for quantities. Important innovations of this study over past research are its explicit handling of potentially heterogeneous factor neutral and factor biased productivities across local labor markets and the use of exogenous variation in relative labor supply shocks for econometric identification.

5.3 Decomposing the Growth in Urban Inequality

We use these estimated parameters to investigate the extent to which each of the components of the growth in log wage gaps between skilled and unskilled workers captured in each of the four terms in (2) can explain the increasingly positive relationship between wage inequality and city size. Table 7 Panel A reports coefficients from regressions of actual and predicted $\Delta_t \ln \frac{w_j^s}{w_j^u}$, and components thereof, on $\ln D_j$ and time fixed effects, weighting by CBSA population. The first two rows, which report these coefficients for actual and predicted $\Delta_t \ln \frac{w_j^s}{w_j^u}$ respectively, are intended as a baseline. Subsequent rows report the relationships between each of the terms in (2) on $\ln D_j$.[18] They use actual data to measure $\Delta_t \ln \frac{S_j}{U_j}$ and $\Delta_t \ln \frac{K_j}{S_j}$ so as to replicate actual variation in $\Delta_t \ln \frac{w_j^s}{w_j^u}$ rather than just the portion of the variation with clean identification. If $\rho = \sigma$, or there is no capital-skill complementarity, the final two terms in (2) are 0 and the full change in the distribution of wage

[18]That is, they show results of regressing $\widehat{\sigma}(\widehat{d\mu_s} - \widehat{d\mu_u}) \ln D_j$, $(\widehat{\sigma} - 1)\, \Delta_t \ln\left(\frac{S_j}{U_j}\right)$, $(\widehat{\sigma} - \widehat{\rho})\, \omega_j^c \Delta_t \ln\left(\frac{K_j}{S_j}\right)$ and $(\widehat{\sigma} - \widehat{\rho})\, \omega_j^c (\widehat{d\mu_k} - \widehat{d\mu_s}) \ln D_j$ respectively on $\ln D_j$ and decade fixed effects.

gaps must be generated by changes in the skill bias of agglomeration economies and/or relative supply shifts.

Results in the first two rows of Table 7 Panel A show that the growth in the wage gap predicted using estimated parameters has the very similar strongly positive relationship with CBSA population as seen in the actual data. For both raw and efficiency units, the elasticities of the predicted decadal change in wage gaps with respect to city size are about 0.003 greater than the actual elasticities of 0.013 and 0.011 observed in the data. Using actual data on factors to build components of (2) increases the positive elasticity with respect to city size to 0.021 for both raw and efficiency units. Given that the model does a good job of predicting $\Delta_t \ln \frac{K_j}{S_j}$, as we demonstrate next, this discrepancy between predicted $\Delta_t \ln \frac{w_j^s}{w_j^u}$ fully from the four equation system versus from using actual factor quantity data must come from the discrepancy between variation in actual $\Delta_t \ln \frac{S_j}{U_j}$ and the exogenous portion of that variation that is predicted by the first stage equation (6).

Results in the final four rows of Panel A provide strong evidence that shifts in the factor bias of agglomeration economies toward skilled labor were central for generating the increasingly positive relationship between $\Delta_t \ln \frac{w_j^s}{w_j^u}$ and $\ln D_j$. Of the four terms in (2), the agglomeration term is the largest, accounting for 79-91 percent of the total elasticity. We find no evidence that relative labor supply shocks had any appreciable effect on the distribution of trends in wage inequality. However, more rapid capital accumulation in larger cities is an important part of the story. While skilled labor demand was boosted more in larger cities through its complementarity with this intensifying capital usage, these increases were almost exactly offset by the declines in skilled labor demand associated with the interaction between capital-skill complementarity and shifts in the factor bias of agglomeration economies. Holding factor quantities constant, the increased productivity of skilled labor in larger cities and declining productivity of capital in these locations caused skill demand to fall, since not as many skilled workers were required per unit of capital.

Through examination of estimates of each term in (3), results in Table 7 Panel B explore why capital intensity increased more rapidly in larger cities. The first two rows of Panel B show that the actual and predicted elasticities of $\Delta_t \ln \frac{K_j}{S_j}$ with respect to city size are very similar at about 0.015, meaning that the model predicts this aspect of the data quite well. Rows 3, 4 and 8 show the elasticities of the three terms in (3) with respect to city size. While the first (constant) term's influence is negligible, changes in the factor bias of agglomeration economies and relative labor supply shifts both play a role. Coefficients of 0.004-0.005 in Row 8 reveal that the slight flattening in the supply relationship between skill intensity and city size over time is magnified to promote increases in capital intensity in larger cities and greater wage inequality in these locations as a consequence. Results in Rows 5-7 decompose the impact of the agglomeration term on city size of 0.011-0.013 into components driven by the change in the agglomeration bias toward each factor separately. They show that the agglomeration component of $\Delta_t \ln \frac{K_j}{S_j}$ is entirely driven by increases in the skill bias of agglomeration economies. That is, as agglomeration economies have become more skill biased, the complementarity of capital with skill has driven more rapid capital accumulation in larger cities than smaller cities.

This is the first direct evidence showing how changes in the bias of agglomeration economies toward skilled workers has led to more rapid increases in between group wage inequality in larger cities. The evidence presented here is consistent with the indirect evidence provided in Baum-Snow & Pavan (2013), Bacold, Blum & Strange (2009) and Glaeser & Saiz (2003).

5.4 Robustness

This sub-section reports on several robustness checks. Each of these involves estimating the same model using a different version of the data that is used to generate results in Table 6. Results for each of these robustness checks are in Table 8.

Table 8 Columns 1 and 2 report results excluding the final year of data. We do this both because the data in 2005-2007 has a different timing from the data in other years, and in order to get a sense of the extent to which data from the most recent time period drive the main results. In the final cross-section, capital data is for 2007 and labor data is for the 2004-2007 period whereas in earlier years, capital data is for 1982, 1992 or 2002 and labor data is for 1979, 1989 or 1999. Results in the third and fourth columns exclude the initial year of data. Results in the final two columns both exclude the initial year of data and make use of ω_j^c and ω_j^{cs} that are predicted using shares from 1987 and 1997, rather than data from 1992 and 2002 directly, to calibrate input shares. This use of predicted shares using lagged values is intended to allay potential endogeneity concerns about using shares from the initial period in a difference equation.

None of the results in Table 8 differ significantly from our main results in Table 6. Results in Table 8 still show strong evidence of capital-skill complementarity. They also show increases in the skill-bias of agglomeration economies and declines in the capital-bias of agglomeration economies. There is also strong evidence of stability of parameters across sub-sample periods used to estimate the model.

We have also investigated the potential bias introduced by excluding materials from the analysis. To parsimoniously evaluate the potential importance of materials, we assume perfect substitutability with capital and estimate the model after constructing K_j as capital plus materials. We deflate expenditures on materials using the same investment price deflator used to deflate capital expenditures. Results from estimating the model including materials in capital yields very similar results to those reported in Table 6, though with point estimates indicating slightly stronger capital-skill complementarity. This result is consistent with the fact that these two capital measures are almost perfectly correlated.

As a final robustness check, we investigate the fit of the same production technology as in (1), but with the opposite nesting. We find that each equilibrium equation in this alternative model fits the data less well than each of their analogs among our three primary estimating equations.

6 Conclusions

We use Economic Census data and the micro data from several Decennial Censuses to evaluate mechanisms through which the gaps between average wages of more and less educated workers have become more positively related with city size since 1980. Evidence indicates that a secular increase in the bias of agglomeration economies toward skilled labor has been central for directly generating this trend in the data. However, positive shifts in demand for skill that come with increases in capital intensity driven by this increased bias coupled with the existence of capital-skill complementarity are almost offset by decreases in skill demand because of greater increases in skill productivity in larger locations holding relative factor quantities constant. The declining capital bias of agglomeration economies is an important element generating these greater inward relative demand shifts for skill in larger cities. Given that city size accounts for about one-third of the increase in wage inequality nationwide since 1980 (Baum-Snow & Pavan, 2013), an important fraction of the nationwide increase in the skill premium since 1980 can thus be traced back to increases in the skill bias of agglomeration economies.

While some existing research empirically examines the relative importance of various mechanisms through which agglomeration economies operate in the cross-section, this is among the first papers to examine how the relative importance of these mechanisms has changed in recent decades. Our results indicate that the increase in the complementarity in production between human capital and market scale points to increases in the importance of knowledge spillovers across workers and/or learning for generating agglomeration economies. While the magnitude of this human capital spillover mechanism is sufficiently large to account for most of the increase in the relationship between average wages and city size during the 1980s, the fact that skill and city size continued to become more complementary after 1990, during a period of relative stability in the elasticity of wages with respect to city size, is evidence of this mechanism's increasing importance for generating agglomeration economies. Other proposed broad explanations for agglomeration economies, such as labor market pooling and input sharing (Duranton & Puga, 2004), seem unlikely to have an inherent skill bias. Other explanations that compete with agglomeration economies for generating productivity differences across cities of different sizes, like differences in natural endowments and market access, also seem unlikely to interact with skill.

There remains much to be learned about why cities and skills have become more complementary in production. As such, we hope that this study sparks additional research into more microfounded mechanisms driving these changes in the nature of agglomeration economies.

References

Acemoglu, Daron. 1998. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." Quarterly Journal of Economics, 113:3, 1055-1089.

Albouy, David. 2014. "What Are Cities Worth? Land Rents, Local Productivity, and the Capitalization of Amenity Values." NBER Working Paper No. 14981.

Autor, David H., Lawrence F. Katz and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." Review of Economics and Statistics, 90:2, 300-323.

Autor, David H., Lawrence F. Katz and Alan B. Krueger. 1998. "Have Computers Changed the Labor Market?" Quarterly Journal of Economics, 113:3, 1169-1213.

Autor, David H., Frank Levy and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." Quarterly Journal of Economics, 118:4, 1279-1334.

Baum-Snow, Nathaniel and Ronni Pavan. 2012. "Understanding the City Size Wage Gap" Review of Economic Studies, 79:1, 88-127.

Baum-Snow, Nathaniel and Ronni Pavan. 2013. "Inequality and City Size" Review of Economics and Statistics, 95:5, 1535-1548.

Bacold, Marigee, Bernardo S. Blum and William S. Strange. 2009. "Skills in the City." Journal of Urban Economics, 65:2, 136-153.

Borjas, George J. 2003. "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market." Quarterly Journal of Economics,

Bound, John and George Johnson. 1992. "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations." American Economic Review, 82:3, 371-392.

Cai, Zongwu, Mitali Das, Huaiyu Xiong and Xizhi Wu. 2006. "Function Coefficient Instrumental Variable Models." Journal of Econometrics, 133:1, 207-241.

Card, David. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." Journal of Labor Economics, 19:1, 22-64.

Card, David and John E. DiNardo. 2000. "Do Immigrant Inflows Lead to Native Outflows?" American Economic Review, 90:2, 360-367.

Ciccone, Antonio & Giovanni Peri. 2005. "Long-Run Substitutability between More and Less Educated Workers: Evidence from U.S. States 1950-1990" Review of Economics and Statistics, 87:4, 652-663.

Combes, Pierre-Philippe, Gilles Duranton and Sebastien Roux. 2010. "Estimating Agglomeration Economies With History, Geology and Worker Effects." In The Economics of Agglomeration, edited by Edward Glaeser, University of Chicago Press.

Diamond, Rebecca. 2013. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000" manuscript.

Duffy, John, Chris Papageorgiou and Fidel Perez-Sebastian. 2004. "Capital-Skill Complementarity? Evidence from a Panel of Countries." Review of Economics and Statistics, 86:1, 327-344.

Dunne, Timothy, Lucia Foster, John Haltiwanger and Kenneth R. Troske. 2004. “Wage and Productivity Dispersion in U.S. Manufacturing: The Role of Computer Investment.” Journal of Labor Economics, 22:2, 397-429.

Duranton, Gilles & Diego Puga. 2004. “Micro-Foundations of Urban Agglomeration Economies,” in Handbook of Urban and Regional Economics Vol. 4, J.V. Henderson & J-F Thisse eds.North Holland-Elsevier.

Dustmann, Christian, Tommaso Frattini and Ian P. Preston. 2013. “The Effect of Immigration Along the Distribution of Wages.” Review of Economic Studies, 80:1, 145-173.

Glaeser, Edward L. and Albert Saiz. 2003. “The Rise of the Skilled City.” Harvard Institute of Economic Research Discussion Paper Number 2025.

Goldin, Claudia and Lawrence F. Katz. 1998. “The Origins of Technology-Skill Complementarity.” Quarterly Journal of Economics, 113:3, 693-732.

Griliches, Zvi. 1969. “Capital-Skill Complementarity.” Review of Economics and Statistics, LI, 465-468.

Greenwood, Jeremy, Zvi Hercowitz & Per Krusell. 1997. “Long-Run Implications of Investment-Specific Technological Change,” American Economic Review, 87:3, 342-362.

Harper, Michael J. 1999. “Estimating Capital Inputs for Productivity Measurement: An Overview of U.S. Concepts and Methods,” International Statistical Review, 67:3, 327-337.

Juhn, Chinhui, Kevin Murphy & Brooks Pierce. 1993. “Wage Inequality and the Rise in Returns to Skill.” Journal of Political Economy, 101:3, 410-442.

Katz, Lawrence and Kevin Murphy. 1992. “Changes in Relative Wages 1963-1987: Supply and Demand Factors.” Quarterly Journal of Economics, 107:1, 35-78.

Krusell, Per, Lee E. Ohanian, Jose-Victor Rios-Rull, and Giovanni L. Violante. 2000. “Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis.” Econometrica, 68:5, 1029-1053.

Lewis, Ethan. 2011. “Immigration, Skill Mix, and Capital-Skill Complementarity.” Quarterly Journal of Economics, 126:2, 1029-1069.

Moretti, Enrico. 2013. “Real Wage Inequality.” American Economic Journal: Applied Economics, 5:1, 65-103.

Ottaviano, Gianmarco I.P. & Giovanni Peri. 2012. “Rethinking the Effects of Immigration on Wages.” Journal of the European Economic Association, 10:1, 152-197.

Ruggles, Steven J., Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.

A Structural Model Estimation Specifications

The full structural model has the three equations (2), (4) and (5) plus the additional relative supply or "first stage" equation (6). For the purpose of estimation, the "sparse" model simply includes flexible controls for local labor market scale whereas the "full" model is a flexible specification of the four main structural equations. w^s, w^u, S, U, K and Y are observed in the data for each CBSA in each study year. ω^c and ω^s are also observed in each study year but used in initial years only. In the notation below, $\Delta_t(x) = x_t - x_{t-1}$.

A.1 "Sparse" Empirical Model

For the sparse empirical model, we estimate the following 28 parameters in the estimation equations below: $\delta_t, \lambda_t^1, \lambda_t^2, \lambda_t^3, \kappa_t^1, \kappa_t^2, \kappa_t^3, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \sigma, \rho$

$$\Delta_t \ln \frac{S_j}{U_j} = \delta_t + \alpha_1 \Delta_t \ln \frac{\widehat{S}_j}{\widehat{U}_j} + \alpha_2 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + u_{jt}$$

$$\Delta_t \ln \frac{w_j^s}{w_j^u} = \lambda_t^1 + \beta_1 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \kappa_t^1 \ln D_j + (\sigma - 1)\Delta_t \ln \left(\frac{S_j}{U_j}\right) + (\sigma - \rho)\omega_{jt-1}^c Q_{jt}^s + \varepsilon_{jt}^1$$

$$\Delta_t \ln w_j^s = \lambda_t^2 + \beta_2 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \kappa_t^2 \ln D_j + (1 - \rho)\, Q_{jt}^s + \varepsilon_{jt}^2$$

$$\begin{aligned} \Delta_t \ln \frac{K_j}{Y_j} = \lambda_t^3 \qquad & + \beta_3 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \kappa_t^3 \ln D_j \\ & + \frac{(\rho - \sigma)(1 - \omega_{jt-1}^c)(1 - \omega_{jt-1}^{cs})}{(\sigma - \rho)\omega_{jt-1}^c (1 - \omega_{jt-1}^{cs}) - (1 - \rho)(1 - \omega_{jt-1}^{cs}\omega_{jt-1}^c)} \Delta_t \ln \frac{S_j}{U_j} + \varepsilon_{jt}^3 \end{aligned}$$

In the expressions above,

$$Q_{jt}^s = \frac{(1 - \sigma)(1 - \omega_{jt-1}^{cs})}{(\sigma - \rho)\omega_{jt-1}^c (1 - \omega_{jt-1}^{cs}) - (1 - \rho)(1 - \omega_{jt-1}^{cs}\omega_{jt-1}^c)} \Delta_t \ln \frac{S_j}{U_j}$$

A.2 "Full" Empirical Model

For the full empirical model, we estimate the following 25 parameters in the estimation equations below: $\delta_t, \lambda_t^1, \lambda_t^2, \lambda_t^3, \pi_t, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \sigma, \rho, d\mu_s, d\mu_u, d\mu_k$

$$\Delta_t \ln \frac{S_j}{U_j} = \delta_t + \alpha_1 \Delta_t \ln \frac{\widehat{S}_j}{\widehat{U}_j} + \alpha_2 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + u_{jt}$$

$$\begin{aligned}\Delta_t \ln \frac{w_j^s}{w_j^u} &= \lambda_t^1 + \beta_1 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \frac{(\sigma-\rho)\omega_{jt-1}^c}{(\sigma-\rho)\omega_{jt-1}^c(1-\omega_{jt-1}^{cs}) - (1-\rho)(1-\omega_{jt-1}^{cs}\omega_{jt-1}^c)}\pi_t \\ &+\sigma(d\mu_s - d\mu_u)\ln D_j + (\sigma-1)\Delta_t \ln\left(\frac{S_j}{U_j}\right) + (\sigma-\rho)\omega_{jt-1}^c Q_{jt} + (\sigma-\rho)\omega_{jt-1}^c(d\mu_k - d\mu_s)\ln D_j + \varepsilon_{jt}^1\end{aligned}$$

$$\begin{aligned}\Delta_t \ln w_j^s &= \lambda_t^2 + \beta_2 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \frac{1-\rho}{(\sigma-\rho)\omega_{jt-1}^c(1-\omega_{jt-1}^{cs}) - (1-\rho)(1-\omega_{jt-1}^{cs}\omega_{jt-1}^c)}\pi_t \\ &+\rho\,(d\mu_s - d\mu_k)\ln D_j + (1-\rho)\,Q_{jt} + \varepsilon_{jt}^2\end{aligned}$$

$$\begin{aligned}\Delta_t \ln \frac{K_j}{Y_j} = \lambda_t^3 \quad &+\beta_3 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \frac{1-\omega_{jt-1}^{cs}\omega_{jt-1}^c}{(\sigma-\rho)\omega_{jt-1}^c(1-\omega_{jt-1}^{cs}) - (1-\rho)(1-\omega_{jt-1}^{cs}\omega_{jt-1}^c)}\pi_t \\ &+\frac{(\rho-\sigma)(1-\omega_{jt-1}^c)(1-\omega_{jt-1}^{cs})(d\mu_s - d\mu_u) + (\sigma\omega_{jt-1}^{cs}\omega_{jt-1}^c + (\rho-\sigma)\omega_{jt-1}^c - \rho)d\mu_k}{(1-\sigma)\omega_{jt-1}^{cs}\omega_{jt-1}^c + (\sigma-\rho)\omega_{jt-1}^c + (\rho-1)}\ln D_j \\ &+\frac{(\rho-\sigma)(1-\omega_{jt-1}^c)(1-\omega_{jt-1}^{cs})}{(\sigma-\rho)\omega_{jt-1}^c(1-\omega_{jt-1}^{cs}) - (1-\rho)(1-\omega_{jt-1}^{cs}\omega_{jt-1}^c)}\Delta_t \ln \frac{S_j}{U_j} + \varepsilon_{jt}^3\end{aligned}$$

In the expressions above,

$$\begin{aligned}Q_{jt} = \quad &\frac{(1-\sigma)(1-\omega_{jt-1}^{cs})(d\mu_s - d\mu_u) + ((1-\sigma)\omega_{jt-1}^{cs}\omega_{jt-1}^c + (\sigma-\rho)\omega_{jt-1}^c + \rho)(d\mu_s - d\mu_k) - d\mu_s}{(\sigma-\rho)\,\omega_{jt-1}^c(1-\omega_{jt-1}^{cs}) - (1-\rho)(1-\omega_{jt-1}^c\omega_{jt-1}^{cs})}\ln D_j \\ &+\frac{(1-\sigma)(1-\omega_{jt-1}^{cs})}{(\sigma-\rho)\omega_{jt-1}^c(1-\omega_{jt-1}^{cs}) - (1-\rho)(1-\omega_{jt-1}^{cs}\omega_{jt-1}^c)}\Delta_t \ln \frac{S_j}{U_j}\end{aligned}$$

Table 1: Patterns in Log Wage Premia by Education and Location
CBSA Residents

	Wage Premium For Indicated Skill Definitions				Elasticity of Wages
Skilled Worker Definition	Some College+	College+	College+	College Only	With Respect to
Unskilled Worker Definition	High School-	Some College-	High School-	High School Only	1980 CBSA Pop.
Panel A: All Workers					
1980	0.30	0.47	0.50	0.38	0.056
1990	0.38	0.54	0.62	0.46	0.074
2000	0.43	0.57	0.66	0.50	0.067
2005-7	0.48	0.62	0.72	0.55	0.067
Panel B: Manufacturing Workers Only					
1980	0.36	0.56	0.60	0.48	0.045
1990	0.42	0.59	0.67	0.53	0.060
2000	0.46	0.63	0.72	0.57	0.056
2005-7	0.53	0.70	0.80	0.64	0.061

Notes: Skilled wage premia for those living in all locations are within 0.01 of those reported for urban locations. All reported premia and elasticities are statistically significant. Calculations incorporate census weights interacted with labor supply.

Table 2: Elasticities of Skill Price Gaps and Relative Factor Intensities With Respect to City Size

S = Some College or More
U = High School or Less

	Raw Counts			Efficiency Units		
	w^S/w^U	S/U	K/S	w^S/w^U	S/U	K/S
	Panel A: All Workers					
1980	0.019	0.103	No	0.008	0.114	No
1990	0.029	0.107	Capital	0.014	0.121	Capital
2000	0.044	0.102	Data	0.024	0.121	Data
2005-7	0.051	0.089		0.029	0.110	
	Panel B: Manufacturing Workers					
1980	0.030	0.135	-0.070	0.008	0.156	-0.091
1990	0.045	0.136	-0.047	0.020	0.160	-0.062
2000	0.064	0.124	-0.025	0.033	0.154	-0.039
2005-7	0.072	0.128	-0.042	0.042	0.158	-0.055

each indicated year. Regressions are weighted by 1980 CBSA population. All estimated coefficients are statistically significant.

Table 3: Changes in Relative Factor Prices and Quantities With Respect to City Size

	All Workers				Manufacturing Workers					
	Raw Counts		Efficiency Units		Raw Counts			Efficiency Units		
	$\Delta\ln(w^S/w^U)$	$\Delta\ln(S/U)$	$\Delta\ln(w^S/w^U)$	$\Delta\ln(S/U)$	$\Delta\ln(w^S/w^U)$	$\Delta\ln(S/U)$	$\Delta\ln(K/S)$	$\Delta\ln(w^S/w^U)$	$\Delta\ln(S/U)$	$\Delta\ln(K/S)$
log(1980 CBSA Pop)	0.011***	-0.005***	0.007***	-0.001	0.014***	-0.002	0.014***	0.011***	0.001	0.017***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
1990-2000	-0.048***	-0.289***	-0.054***	-0.283***	0.000	-0.335***	0.219***	-0.025***	-0.308***	0.197***
Indicator	(0.002)	(0.005)	(0.001)	(0.005)	(0.003)	(0.008)	(0.011)	(0.003)	(0.008)	(0.011)
2000-2007	-0.042***	-0.516***	-0.027***	-0.526***	0.003	-0.524***	0.270***	-0.000	-0.516***	0.255***
Indicator	(0.002)	(0.005)	(0.001)	(0.005)	(0.003)	(0.008)	(0.011)	(0.003)	(0.008)	(0.011)
Constant	0.072***	0.552***	0.061***	0.558***	0.031***	0.598***	-0.247***	0.033***	0.591***	-0.262***
	(0.002)	(0.004)	(0.001)	(0.004)	(0.003)	(0.007)	(0.009)	(0.002)	(0.007)	(0.009)
Observations	2,766	2,766	2,766	2,766	2,766	2,766	2,202	2,766	2,766	2,202
R-Squared	0.356	0.817	0.397	0.827	0.113	0.639	0.260	0.144	0.624	0.233

Notes: Each column reports coefficients and standard errors from a separate regression of the the variables listed at top (in decadal changes) on the variables listed at left. Skilled workers are defined as those with at least some college and unskilled workers are defined as those with high school or less. Regressions are weighted by 1980 CBSA population.

Table 4: Supply Shock Regressions by Education
Manufacturing Workers

	Δln(Quantity of Workers With Indicated Education)				
	< HS	HS	Some Coll.	College	>College
Δln(Predicted	0.38***	0.23***	0.080	0.15***	-0.040
Quantity)	(0.044)	(0.037)	(0.050)	(0.055)	(0.097)
ln(CBSA Population)	-0.11***	-0.041***	-0.032**	-0.075***	-0.0038
	(0.012)	(0.016)	(0.013)	(0.016)	(0.022)
ln(Immigrants of	0.059***	-0.0043	-0.015	0.034***	0.0070
indicated educ.$_{t-1}$)	(0.0078)	(0.012)	(0.0092)	(0.010)	(0.015)
Observations	2,752	2,765	2,707	2,374	2,424
R-Squared	0.25	0.22	0.66	0.48	0.21
Year FE	Yes	Yes	Yes	Yes	Yes

Notes: We construct Δln(Predicted Quantity) using 1970 immigrant settlement patterns across CBSAs by region of origin interacted with national immigration trends from each region of origin over subsequent decades. See the text for more details. While the full sample includes 922 CBSAs over three time periods, we must drop those CBSAs which had 0 sampled immigrants in a given education group in 1970. Regressions are weighted by 1980 CBSA population and standard errors are clustered by CBSA.

Table 5: IV Regression Results Incorporating Agglomeration Economies
Manufacturing Workers

	Raw Counts				Efficiency Units			
	Δln(S/U) F.S.	Δln(K/Y)	Δln(w^s)	Δln(w^s/w^u)	Δln(S/U) F.S.	Δln(K/Y)	Δln(w^s)	Δln(w^s/w^u)
Δln(Predicted S	0.21***				0.17***			
/Predicted U)	(0.045)				(0.043)			
Δln(Skilled Labor		0.63***	-0.25**	-0.43***		0.77***	-0.26**	-0.31***
/Unskilled Labor)		(0.20)	(0.11)	(0.099)		(0.27)	(0.12)	(0.089)
ln(CBSA Population)	0.0012	0.0061	0.0094***	0.013***	0.0033	0.0047	0.012***	0.012***
	(0.0065)	(0.0081)	(0.0026)	(0.0020)	(0.0062)	(0.0096)	(0.0027)	(0.0014)
ln(Skilled Imm. /	0.044***	-0.029**	0.021**	0.026***	0.046***	-0.039**	0.016	0.019**
Unskilled Imm.)$_{t-1}$	(0.012)	(0.013)	(0.0100)	(0.0088)	(0.012)	(0.017)	(0.010)	(0.0072)
Year = 2000	-0.31***	-0.19	0.12**	0.24***	-0.29***	-0.27*	0.13*	0.17***
	(0.017)	(0.12)	(0.060)	(0.059)	(0.018)	(0.15)	(0.067)	(0.052)
Year = 2005-2007	-0.45***	0.051	0.12***	0.081***	-0.46***	0.013	0.12***	0.040*
	(0.030)	(0.050)	(0.025)	(0.022)	(0.029)	(0.067)	(0.029)	(0.021)
Constant	0.51***	-0.18***	-0.027***	0.064***	0.52***	-0.19***	-0.057***	0.054***
	(0.026)	(0.016)	(0.0077)	(0.0067)	(0.026)	(0.020)	(0.0084)	(0.0060)
Observations	2,751	2,047	2,751	2,751	2,751	2,047	2,751	2,751
First stage F		20.0	21.7	21.7		14.8	15.5	15.5

Notes: The first column in each block gives first stage results. Remaining columns show IV estimates in which the change in the log of predicted skilled vs. unskilled workers using historical immigration pathways and contemporaneous national immigration shocks instruments for the change in the log of actual skilled vs. unskilled workers. Observations are weighted by 1980 CBSA population and standard errors are clustered on CBSA.

Table 6: Parameter Estimates

Para-meter	Description	Sparse Model Counts	Sparse Model Eff Unit	Full Model Counts	Full Model Eff Unit
α_1	Coefficient on instrument in Equation (6)	0.28	0.24	0.26	0.21
		(0.05)	(0.04)	(0.04)	(0.04)
σ	1/(1-σ)=elast of sub btw K or S and U	0.85	0.88	0.81	0.85
		(0.03)	(0.03)	(0.03)	(0.03)
ρ	1/(1-ρ)=elast of sub btw K and S	0.20	0.27	-0.33	-0.38
		(0.15)	(0.16)	(0.33)	(0.47)
$d\mu_u$	change in unskilled labor biased agglom.			-0.008	-0.003
				(0.003)	(0.002)
$d\mu_s$	change in skilled labor biased agglom.			0.015	0.016
				(0.003)	(0.003)
$d\mu_k$	change in capital biased agglom.			-0.011	-0.009
				(0.002)	(0.002)

Entries list parameter estimates and standard errors from two specifications of the four equation structural model. The first two columns show parameter estimates identified entirely from the heterogeneous coefficients on Δln(S/U), with the remaining parameters of interest not identified. The final two columns show parameter estimates of the complete model as explained in the text. Exact estimation equations are in the Appendix. Observations are weighted by 1980 CBSA population and standard errors are clustered by CBSA. Totals of 28 and 25 parameters are estimated in the two models respectively.

Table 7: Relationships Between Components of $\Delta\ln(w^s/w^u)$ and $\Delta\ln(K/S)$ and ln(D)

	Object	Equation and Term	Raw Counts	Efficiency Units
		Panel A: $\Delta\ln(w^s/w^u)$		
1	Actual		0.013	0.011
2	Predicted		0.016	0.014
3	Predicted Using K,S,U from Data		0.021	0.021
4	Agglomeration	Eqn 2, Term 1	0.019	0.016
5	S-U Shifts	Eqn 2, Term 2	0.001	0.000
6	K-S Shifts	Eqn 2, Term 3	0.013	0.016
7	Agglom. K-S Shifts Interaction	Eqn 2, Term 4	-0.012	-0.012
		Panel B: $\Delta\ln(K/S)$		
1	Actual		0.014	0.016
2	Predicted		0.016	0.016
3	Constant	Eqn 3, Term 1	-0.001	-0.001
4	Agglomeration	Eqn 3, Term 2	0.011	0.013
5	Agglom, S Bias	Eqn 3, Term 2, $d\mu_s \neq 0$	0.014	0.015
6	Agglom, U Bias	Eqn 3, Term 2, $d\mu_u \neq 0$	-0.003	-0.002
7	Agglom, K Bias	Eqn 3, Term 2, $d\mu_k \neq 0$	-0.001	0.000
8	S-U Shifts	Eqn 3, Term 3	0.005	0.004

Each entry is the coefficient in the regression of the object listed at left on the log of 1980 CBSA population and year fixed effects weighted by 1980 CBSA population. All objects in Panel A except those in the first two rows are calculated using actual data for S, U and K. Rows 5-7 in Panel B use the second term in Equation (3) in the text but restrict two of the three factor biases of agglomeration economies to 0. Indented estimates sum to the estimate immediately above them. All estimates with an absolute value greater than 0.001 are strongly significant.

Table 8: Robustness Checks on Parameter Estimates

	Exclude 2005-7		Exclude 1980		Exclude 1980 & Pred. ω^c, ω^{cs}	
	Counts	Eff Unit	Counts	Eff Unit	Counts	Eff Unit
Panel A: Sparse Model						
α_1	0.27	0.23	0.29	0.26	0.31	0.29
	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
σ	0.84	0.91	0.73	0.80	0.76	0.82
	(0.03)	(0.02)	(0.06)	(0.06)	(0.06)	(0.05)
ρ	0.28	0.51	-0.52	0.15	-0.26	0.27
	(0.22)	(0.19)	(1.02)	(0.50)	(0.79)	(0.41)
Panel B: Full Model						
α_1	0.25	0.21	0.26	0.23	0.30	0.32
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)
σ	0.83	0.89	0.69	0.79	0.75	0.87
	(0.03)	(0.02)	(0.07)	(0.06)	(0.06)	(0.04)
ρ	-0.80	-3.06	-0.87	-0.91	-0.62	0.25
	(0.92)	(6.96)	(1.02)	(1.35)	(0.77)	(0.19)
$d\mu_u$	-0.011	-0.003	-0.013	-0.009	-0.011	-0.006
	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
$d\mu_s$	0.019	0.019	0.013	0.012	0.015	0.012
	(0.003)	(0.003)	(0.004)	(0.003)	(0.005)	(0.004)
$d\mu_k$	-0.013	-0.010	-0.011	-0.010	-0.014	-0.010
	(0.003)	(0.002)	(0.003)	(0.003)	(0.005)	(0.004)

Notes: Estimates and standard errors are reported for three alternative ways of setting up the data. The first two columns show results when the final period is excluded from the data. Columns 3-4 show results when the first period is excluded from the data. Columns 5-6 additionally use input shares that have been predicted using 1987 and 1997 shares.

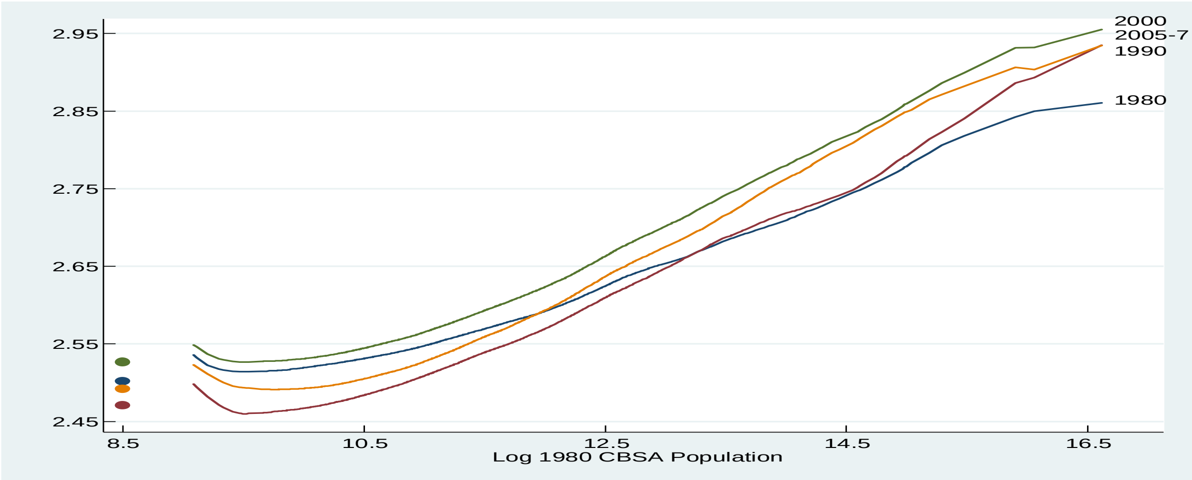
Table A1: Summary Statistics, Manufacturing

1980 CBSA Population (Obs.)	1980	2005-7	1980	2005-7	1980	2005-7	1980	2005-7	1980	2005-7
	0-50k (380)		50k-100k (234)		100k-250k (167)		250k-1m (103)		>1m (38)	
Panel A: Raw Units										
ln S	13.7	14.4	14.7	15.4	15.6	16.2	16.9	17.4	18.7	19.1
	(0.7)	(0.7)	(0.6)	(0.6)	(0.6)	(0.6)	(0.7)	(0.6)	(0.8)	(0.7)
ln U	15.0	14.8	16.0	15.7	16.6	16.3	17.7	17.3	19.3	18.7
	(0.9)	(0.9)	(0.7)	(0.7)	(0.6)	(0.6)	(0.7)	(0.6)	(0.8)	(0.7)
ln K	11.6	12.6	13.0	13.4	13.9	14.2	15.2	15.5	16.7	17.0
	(1.4)	(1.0)	(1.1)	(0.9)	(0.9)	(0.8)	(0.8)	(0.7)	(0.8)	(0.8)
ln w^s	2.66	2.60	2.72	2.67	2.78	2.75	2.85	2.87	2.93	3.01
	(0.16)	(0.16)	(0.16)	(0.14)	(0.16)	(0.16)	(0.14)	(0.14)	(0.10)	(0.13)
ln w^u	2.42	2.32	2.47	2.34	2.53	2.41	2.53	2.41	2.60	2.46
	(0.18)	(0.13)	(0.19)	(0.13)	(0.18)	(0.13)	(0.17)	(0.11)	(0.14)	(0.10)
Panel B: Efficiency Units										
ln S	16.4	17.1	17.4	18.2	18.4	19.0	19.7	20.3	21.5	22.0
	(0.7)	(0.7)	(0.6)	(0.6)	(0.6)	(0.6)	(0.7)	(0.6)	(0.8)	(0.7)
ln U	17.6	17.5	18.6	18.4	19.3	19.0	20.4	20.0	21.9	21.5
	(0.9)	(0.9)	(0.7)	(0.6)	(0.6)	(0.6)	(0.7)	(0.6)	(0.8)	(0.7)
ln w^s	-0.03	-0.14	0.01	-0.09	0.06	-0.03	0.09	0.05	0.15	0.17
	(0.14)	(0.14)	(0.12)	(0.12)	(0.12)	(0.13)	(0.11)	(0.12)	(0.08)	(0.11)
ln w^u	-0.20	-0.39	-0.16	-0.37	-0.12	-0.32	-0.11	-0.30	-0.04	-0.24
	(0.13)	(0.11)	(0.14)	(0.11)	(0.14)	(0.11)	(0.13)	(0.09)	(0.11)	(0.08)

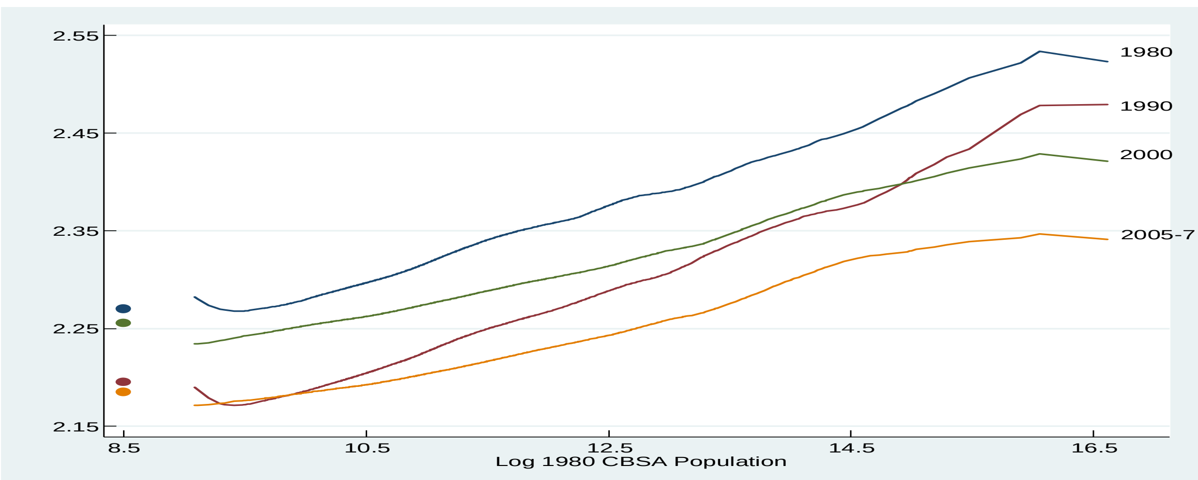
Notes: Entries give means with standard deviations in parentheses for each variable listed at left in the set of CBSAs in each population range listed at top.

Figure 1: The Relationship Between Labor Factor Prices and City Size - All Workers

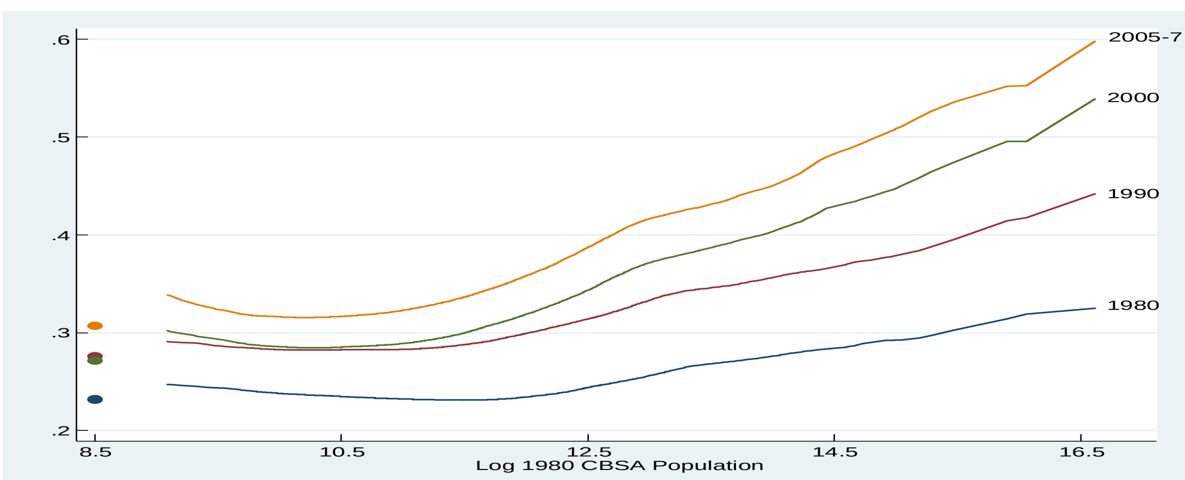
Panel A: Log Wage, Some College or More



Panel B: Log Wage, High School or Less



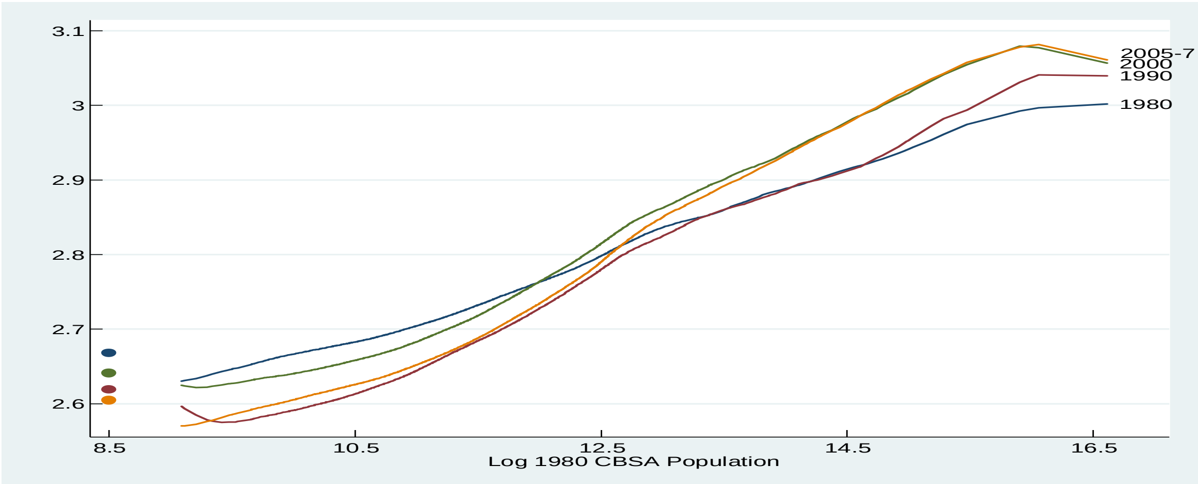
Panel C: Gap



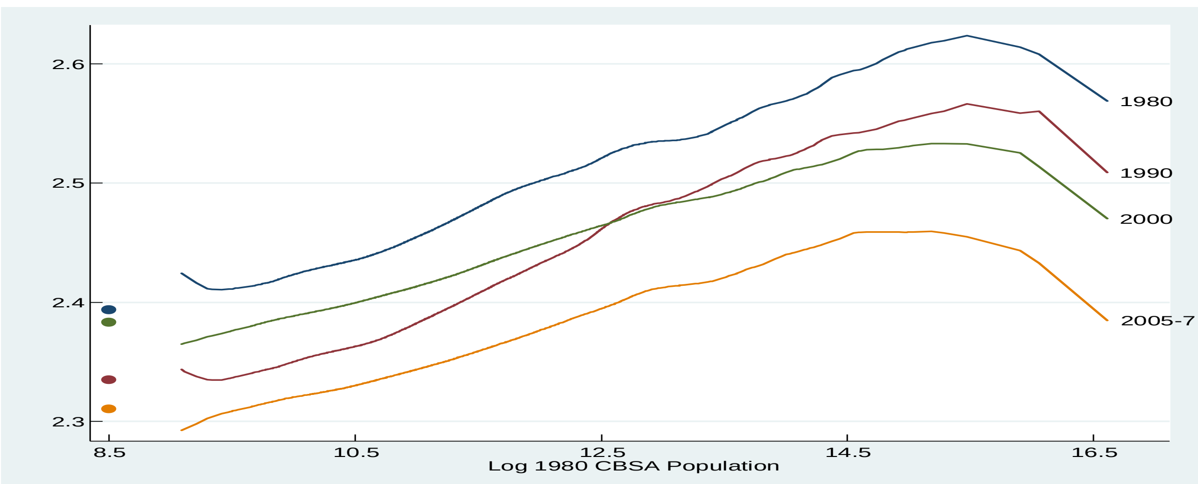
Notes: Each wage observation is weighted using census weights interacted with hours worked. Isolated dots at the left of the graphs are for rural areas.

Figure 2: The Relationship Between Labor Factor Prices and City Size - Manufacturing Workers

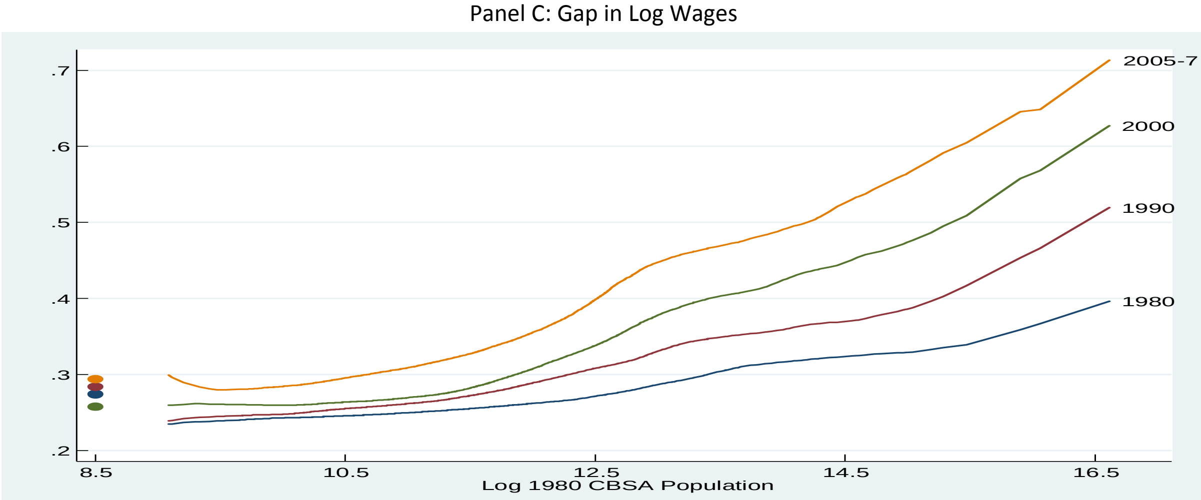
Panel A: Log Wage, Some College or More



Panel B: Log Wage, High School or Less



Panel C: Gap in Log Wages



Notes: Each wage observation is weighted using census weights interacted with hours worked. Isolated dots at the left of the graphs are for rural areas.