Why Has Urban Inequality Increased?*

Nathaniel Baum-Snow, University of Toronto
Matthew Freedman, University of California, Irvine
Ronni Pavan, University of Rochester

August, 2017

Abstract

This paper examines mechanisms driving the more rapid increases in wage inequality in larger cities between 1980 and 2007. Production function estimates indicate strong evidence of capital-skill complementarity and increases in the skill bias of agglomeration economies in the context of rapid skill-biased technical change. Immigration shocks are the source of identifying variation across cities in changes to the relative supply of skilled versus unskilled labor. Estimates indicate that changes in the factor biases of agglomeration economies rationalize at least 80 percent of the more rapid increases in wage inequality in larger cities.

*We thank Alex Mas, two referees, Edward Glaeser, Kurt Schmidheiny, and numerous seminar and conference participants for valuable comments. Camilo Acosta and Adrian Rubli provided excellent research assistance.

1 Introduction

Since the seminal work of Katz and Murphy (1992), Bound and Johnson (1992), and Juhn, Murphy, and Pierce (1993), economists have recognized that the structure of wages in the U.S. economy shifted markedly after 1980 toward greater inequality. These studies, as well as more recent work by Autor, Katz, and Kearney (2008), have highlighted the importance of changes in the relative demand for skill in explaining the rise in inequality nationwide over the past several decades. Only more recently, however, have studies uncovered the extent to which the rise in inequality varies across geography. For example, Moretti (2013) and Baum-Snow and Pavan (2013) provide evidence that relative demand shifts have occurred disproportionately in cities with higher costs of living and greater populations, respectively, with at least one-quarter of the increase in wage inequality nationwide since 1980 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 variation in skill price changes over time across local labor markets of different sizes. 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 secular factor-biased technical change in generating the more rapid 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 nested constant elasticity of substitution production functions similar to those employed in Griliches (1969), Krusell et al. (2000), and Autor and Dorn (2013) with capital, skilled labor, and unskilled labor as factors of production. To standard specifications, we add agglomeration economies that are allowed to be factor biased and allow for unrestricted factor-biased technical change.

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. 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). The use of both cross-sectional and time-series variation for identification allows us to overcome the well-known challenge, discussed in León-Ledesma et al. (2010), of how to separately identify elasticities of substitution from factor-biased technical change parameters in aggregate production function estimation. These parameter estimates facilitate decompositions of mechanisms driving secular changes in between-group wage inequality in the U.S. across local labor markets over time since 1980.

Our results strongly indicate the existence of capital-skill complementarity and an increase in the bias of agglomeration economies toward skilled labor between 1980 and 2007 in the context of rapid skill-biased technical change. 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. These agglomeration forces primarily operate through a direct effect, with a small additional increase coming from interactions between the 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. Given a national capital market, our high estimated elasticity of substitution between unskilled labor and capital explains why unskilled wages are much less variable across locations than are skilled wages.

Changes in the factor biases of agglomeration economies have been central for understanding trends in wage inequality across local labor markets. Holding factor quantities constant and imposing that the elasticity of substitution between unskilled labor and skilled labor is the same as that between capital and skilled labor, changes in the factor biases of agglomeration economies rationalize at least 80 percent of the more rapid increase in wage inequality in larger cities. However, forces that involve capital-skill complementarity have driven less than 16 percent of this phenomenon. Given evidence in Baum-Snow and Pavan (2013) that the more rapid growth in inequality in larger cities has fed through to explain over 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, including Goldin and Katz (1998), Autor, Katz, and Krueger (1998), Krusell et al. (2000), Autor, Levy, and 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. In addition, we recover estimates of specific production function parameters that govern capital-skill complementarity using panel data and exogenous shocks to local labor markets for econometric identification. Our structural estimates are used to perform a novel accounting 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. Moreover, in contrast to Krusell et al.'s (2000) conclusions, we find that skill-biased technical change is needed in addition to capital-skill complementarity to rationalize the evolution of wage gaps over time in the U.S.

In the context of our estimated model, increases in the skill bias of agglomeration economies means that relative factor demands for more versus less educated workers have been increasing more rapidly in larger cities even after accounting for capital-skill complementarity and secular changes in factor productivities. In practice, this result is consistent with the idea that skilled workers' rates of on-the-job human capital accumulation have increased more rapidly in larger cities over time. This interpretation is in line with Baum-Snow and Pavan's (2012) evidence that returns to experience increase more rapidly with workers' education in larger cities than smaller cities and Wang's (2016) evidence that this phenomenon has become more pronounced since 1980. Coupled with Baum-Snow and Pavan's (2012) scant evidence that differences in search frictions or firm-worker match qualities across cities of different sizes generate differences in wages for workers with the same education levels, our findings point to the increasing importance of "learning" spillovers for understanding worker wage and productivity gaps across cities of different sizes (Duranton and Puga 2004).

Our results also have important implications for macroeconomic analyses of factor markets. First, we provide additional evidence using an identification approach that is novel in this literature that capital-skill complementarities are present and strong. Second, even with constant returns to scale, agglomeration economies render aggregate production technologies to be different across local markets. As such, failure to consider local labor markets and local heterogeneity in production technologies can lead to model mis-specification in some contexts. This observation has become increasingly important with the growing wage disparity of educated workers across cities of different sizes over time. More broadly, our evidence highlights the importance of considering the operations of local labor markets for understanding nationwide trends in wage inequality. Our results paint a nuanced picture of the rich and spatially varied dynamics underlying these trends.

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

2 Data and Basic Empirical Facts

2.1 Data

Our analysis is motivated by recent trends in wage inequality across local labor markets. To provide a descriptive picture of these trends, as well as to estimate the parameters of the model described below, we require information about quantities and prices of skilled and unskilled labor and capital stocks 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, which we describe below. We additionally combine both 1 percent metro public-use micro data samples from the 1970 census into a 2 percent sample in order to help build instruments, as we explain in Section 4.2. We require large sample sizes in order to build data for individual CBSAs. 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.[1] Most of our analysis uses only those who report working in manufacturing.

We use the 922 CBSAs 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 and 234 had 50,000-100,000 residents in 1980. One challenge associated 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

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

group (CG) identifiers, whereas later censuses and the ACS report public-use microdata areas (PUMAs).[2] 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 most of our analysis, we assign those individuals 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).[3] We construct average skilled and unskilled wages in each CBSA, w_j^s and w_j^u, per hour worked and per efficiency unit of labor provided. Our efficiency units adjustment attempts to control for post-1980 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 residence, and country of birth separately for each skill group in 1980. We include residential location because, as discussed below, differences in agglomeration economies and natural advantages could generate variation in worker productivity across locations, and we wish to normalize everything to be relative to a 1980 reference location. We interpret the regression coefficients on worker attributes as the productivity of each element of observed skill within the broader skill classes in a reference location. We use the coefficients on observed individual characteristics from these regressions, β_{1980}^S and β_{1980}^U, in all later years to predict the number of labor efficiency units associated with each worker if they were to work in the reference location. In particular, we assign $\exp(X_{it}\widehat{\beta}_{1980}^g)$ efficiency units of labor to each hour worked by individual i in year t in broad skill group g.[4] We maintain the 1980 weights β_{1980}^S and β_{1980}^U for later years to prevent these weights from changing endogenously in response to changes in labor market conditions and to facilitate

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

[3] We only keep those with imputed education for the purpose of constructing aggregate quantities, not skill prices. This exclusion has a negligible effect on results.

[4] 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.

estimation of changes in productivities. 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. As with raw hours, we measure the price of one efficiency unit of skilled and unskilled labor in each CBSA directly as means in the data.[5] Wage calculations exclude observations with imputed labor supply, education, or income information, though we rescale weights to maintain accurate aggregate population by education. 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, 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. 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. Following the methodology laid out in Harper (1999), we adjust using deflators and depreciation rates reported by the BLS by sector within manufacturing aggregated using sectoral shares. 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.[6] 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

[5] 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.

[6] Factor shares at the national level also incorporate materials, energy, and services. As such, we first renormalize to include only capital and labor. We explore the potential bias introduced by excluding materials in robustness checks.

1980, 182 CBSAs in 1990, 190 CBSAs in 2000, and 263 CBSAs in 2007. If capital data are 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.

While there are some necessary approximations and assumptions in this construction of CBSA capital stocks, the fact that we perform the empirical work in decadal changes gives us more confidence that these issues do not affect the validity of our estimates. The differenced setup of the empirical work means that variation in post-1982 capital investment across CBSAs is the primary way we measure variation in changes in capital stocks. Moreover, as we discuss further below, any classical measurement error in such changes ends up in the error term of one of the estimation equations, thereby not affecting parameter point estimates.

Table A1 presents summary statistics.

2.2 Basic Empirical Patterns

The results in Table 1 provide a broad motivation for our 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 0.18 and 0.26 for all workers and 0.18 to 0.24 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.[7] Because trends in wage gaps are similar across skill definitions, we use some college or more as the skilled group and high school or less as the unskilled

[7] 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.

group (the definition in the first column) for the remainder of this analysis. This definition best balances the data in 1980, when 42 percent of working hours among all workers and 31 percent among 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 all workers and manufacturing workers alike. Interpreted in the context of a Rosen (1979) and Roback (1982) type model, as in Albouy (2016), this is evidence of an increase in the magnitude of agglomeration economies among firms producing tradeable goods during the 1980s.

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, although for completeness we also show results for rural areas (represented by dots at the left of each graph). Each plot is of predicted values from a local polynomial regression of the variable listed in the panel header on log 1980 CBSA population. Sample sizes decline from left to right in these plots because the city size distribution has a thin right tail.

Panel C of Figure 1 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 variables strengthens among larger CBSAs in each year since 1980. 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.[8] Panel B shows the well-documented general deterioration of wages for unskilled workers. At the same time, especially during

[8]In the context of the model presented in the next section, this suggests that $d\mu_s > 0$. Our structural estimates confirm this sign.

the 1990s, the wage profile for this group gets much flatter with respect to city size. This fed through to little change in the bottom part of the wage distribution. It also potentially indicates evidence of declines in the strength of agglomeration economies among unskilled workers.[9]

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 Panel C of Figures 1 and 2 are close to linear (and that the model presented in the next section also implies linear relationships), we focus on average elasticities with respect to city size.[10] 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 of skilled and unskilled workers (w^s/w^u) 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 0.020 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 0.010 for all workers and 0.007 for manufacturing workers.

Results in the second and fifth columns of Table 2 show that the relationship between relative skill quantities (S/U) and city size has changed little since 1980. Indeed, when considering efficiency units, any such changes are negligible, both for all workers and for

[9]In the context of the model presented in the next section, this is consistent with $d\mu_u < 0$. Our structural estimates confirm this sign.

[10]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.

manufacturing workers. These robust changes in relative prices of skilled versus unskilled workers in large vs. small cities 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.[11]

The third column of Panel B in Table 2 shows that during three of the four periods studied, larger cities became more capital intensive relative to small cities. Because S/U changed little, we can conclude that increases in K/U also meant increases in K/S.[12] The final column of Table 2 shows a similar pattern when U is measured as efficiency units. As we demonstrate in the following section, in the context of our model, given the stable relationship between S/U and city size, these results must reflect either more rapid increases in total factor productivity (TFP) in larger cities or decreases in the unskilled labor bias of agglomeration economies.

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. Consistent 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 0.01 over each decade, both for all workers and manufacturing workers, whether measured in raw units or efficiency units. Among all workers, this log wage ratio also experienced secular increases in each study period, with the greatest increase during the 1980s. This is evidence that the other mechanisms we consider have operated against the backdrop of skill-biased technical change. Among manufacturing workers, the secular increases were 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

[11] Diamond (2016) 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. Diamond's result does not hold for CBSAs or MSAs if those with some college education are included in the skilled group.

[12] We are hesitant to compare K/U 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, U_{2005-7} actually applies to the 2004-7 period and U_{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/Y and S/U instead of K/U in most of the empirical work below avoids these measurement problems.

decade since 1980.

3 Theoretical Framework

Patterns in the data discussed in the previous section are consistent with the idea that some combination of capital-skill complementarity and increases in the skill bias of agglomeration economies have been central drivers behind variation in changes in wage inequality across local labor markets since 1980. Moreover, we see evidence of skill-biased technical change operating in addition to these other two mechanisms. Here we lay out a theoretical framework that, given parameter estimates, allows for quantification of the relative importance of these mechanisms, along with relative labor supply shifts across local labor markets, for generating shifts in the wage structure over time and across CBSAs since 1980.

We begin with a standard nested constant elasticity of substitution production technology that incorporates capital-skill complementarity. We augment this specification to additionally incorporate agglomeration economies that may be factor biased. The following resulting specification is a generalization of the national technology estimated in Krusell et al. (2000), though with unskilled labor rather than skilled labor nested with capital, and variants explored in Antras (2004) and León-Ledesma et al. (2010):

$$Y_{jt} = A_{jt}\left[cA_{st}^{\sigma}D_j^{\sigma\mu_{st}}S_{jt}^{\sigma} + (1-c)\left(\lambda A_{kt}^{\rho}D_j^{\rho\mu_{kt}}K_{jt}^{\rho} + (1-\lambda)A_{ut}^{\rho}D_j^{\rho\mu_{ut}}U_{jt}^{\rho}\right)^{\frac{\sigma}{\rho}}\right]^{\frac{1}{\sigma}} \quad (1)$$

In (1), S_{jt} is skilled labor efficiency units, U_{jt} is unskilled labor efficiency units, and K_{jt} is capital, all as chosen by firms in location j. These inputs are combined to produce output Y_{jt}. A_{ut}, A_{kt}, and A_{st} capture factor-specific productivities, and A_{jt} captures location-specific TFP. These productivities are allowed to change over time. D_j denotes the CBSA population level or density. We capture differential changes in agglomeration forces across factors by introducing the exponents μ_{kt}, μ_{st}, and μ_{ut} on D_j. These parameters are allowed to change over time, but do not differ across CBSAs. If the μs are equal, agglomeration is factor neutral. Skill-biased agglomeration requires that $\mu_{st} > \mu_{ut}$ and $\mu_{st} > \mu_{kt}$. 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), that is distinct from skill-biased technical change captured by $d\ln(A_s/A_u)$. More generally, we can think of these changes as shifts in rates of learning on the job by workers of different skill across cities of different sizes. Parameters σ and ρ are related to elasticities of substitution, and c and λ are share parameters. The elasticity of substitution between capital or unskilled labor and skilled labor is $\frac{1}{1-\sigma}$, while that between capital and unskilled labor is $\frac{1}{1-\rho}$, with $\sigma < 1$ and $\rho < 1$. If capital-skill complementarity exists, then $\sigma < \rho$. If either σ or ρ is equal to zero, the corresponding nesting is Cobb-Douglas. Because agglomeration forces differ across CBSAs of different sizes, it would not be possible to represent a national aggregate version of this production technology concisely.[13]

Incorporating two factors in the same nest of the production function imposes that two of the three possible elasticities of substitution are identical. Our nesting choice in (1) is similar to Autor and Dorn's (2013) model for considering labor market polarization, but with a few generalizations. They conceptualize goods production as depending on a CES composite of capital and "routine" labor that is combined with "abstract" labor, with an elasticity of substitution that is constrained to unity (Cobb-Douglas). Rather than nesting capital and unskilled labor together, an alternative logical choice would be to nest capital with skilled labor, as in Krusell et al. (2000), which would constrain the elasticities of substitution between unskilled labor and the other two factors to be the same. The two alternative nesting choices result in similar structural and estimation equations, with a swapping of S and U in the factor demand equations we derive below. Our primary analysis nests unskilled labor with capital primarily because this choice delivers a slightly better fit of the data in our empirical analysis. However, we present results using the alternative nesting choice in Section 5.5. Estimated magnitudes of capital-skill complementarity and shifts in the factor bias of agglomeration economies are similar for both nesting choices.[14] Fully freeing up the production technology to incorporate all three elasticities of substitution makes it intractable for estimation.

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 a higher level than CBSAs. Because of the likely existence of agglomeration economies, we think it important to at least use data disaggregated to the CBSA level. Doing so distinguishes this research from many existing studies, most notably Krusell et al. (2000), that only use national data. In addition, there are many different ways of specifying the

[13]Some of the mechanisms in our model have also been considered in Holmes and Mitchell (2008), which theoretically relates increases in aggregate wage inequality to the positive elasticity of skill intensity with respect to plant size in the context of expanding markets.

[14]Of course, the other possibility would be to nest skilled and unskilled labor together. This specification would impose no capital-skill complementarity by assumption.

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 explore the extent to which agglomeration forces are biased toward a particular factor of production.

In order to utilize and eventually estimate (1), we assume that firms cost minimize over all factors and profit maximize over capital. As described below, we use the first-order conditions to analyze changes over time. In doing so, we assume that the rental market for capital, local wages, input quantities, productivities, and the extent to which agglomeration economies are biased toward each factor can all vary over time. Other parameters are assumed to be fixed, though we explore the implications of allowing σ and ρ to change over time in robustness checks.

Our notation references two key output shares that can be calculated with our data. Here,

$$\omega_j^{cu} = \frac{(1-c)\left(\lambda A_k^\rho D_j^{\rho\mu_k} K_j^\rho + (1-\lambda) A_u^\rho D_j^{\rho\mu_u} U_j^\rho\right)^{\frac{\sigma}{\rho}}}{cA_s^\sigma D_j^{\sigma\mu_s} S_j^\sigma + (1-c)\left(\lambda A_k^\rho D_j^{\rho\mu_k} K_j^\rho + (1-\lambda) A_u^\rho D_j^{\rho\mu_u} U_j^\rho\right)^{\frac{\sigma}{\rho}}}$$

is the share of the theoretical capital-skill composite factor in production and

$$\omega_j^c = \frac{\lambda A_k^\rho D_j^{\rho\mu_k} K_j^\rho}{\lambda A_k^\rho D_j^{\rho\mu_k} K_j^\rho + (1-\lambda) A_u^\rho D_j^{\rho\mu_u} U_j^\rho}$$

is the share of capital in this capital-skill composite. We recover these objects empirically by using the facts that the capital share $\frac{vK_j}{Y_j} = \omega_j^c \omega_j^{cu}$ and the skilled labor share $\frac{w_j^s S_j}{Y_j} = 1 - \omega_j^{cu}$. In these expressions, v denotes the rental price of capital in real terms. We assume national capital and output markets, which 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.[15]

We first derive a central equation that describes the evolution of skilled-unskilled wage gaps over time in each CBSA j. This equation forms the basis of one of three structural estimation equations. Combining the two first-order conditions from cost minimization

[15] Our treatment of capital is most consistent with K capturing capital equipment rather than capital structures. It is true that structures capital is not supplied at the same price in all locations. However, Albouy (2016) 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 in all industries.

with respect to skilled and unskilled labor and differencing over time yields the following inverse relative labor demand equation that relates the relative wages of skilled versus unskilled workers to relative input quantities and parameters:

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

We omit time subscripts, as all time-indexed objects from (1) appear after differential operators in (2).[16] In the third, fourth, and final terms, ω_j^c denotes the share of capital in the theoretical factor of production that combines capital and unskilled labor, as specified above. Derivations of this and all other equations in the text can be found in the Online Appendix. Equation (2) is a generalization of the primary estimation equations used in Ciccone and Peri (2005), Autor, Katz, and 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 location-specific changes in wage inequality.

Equation (2) incorporates four channels that may drive changes in between-group wage inequality in each local labor market over time plus interactions. 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, an increase in the supply of capital relative to unskilled workers, or relative increases in the complementarity of city size and capital, assuming capital-skill complementarity ($\sigma < \rho$). The final two terms capture effects of factor-biased technical change. Much of the labor economics literature focuses only on the second term in this equation. The literature investigating capital-skill complementarity additionally investigates the third term. The literature investigating technical change typically focuses on the final two terms. This is the first paper to additionally consider the components of (2) that capture changes in the factor bias of agglomeration economies. Moreover, this is one of the few investigations to use cross-sectional and time-series variation to aid in identification of parameters other than σ.

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. First, if skilled workers have become

[16] We re-introduce the time subscripts in our estimating equations.

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 reduce the relative productivity of unskilled workers, feeding through into lower demand for their services. Of course, understanding reasons for changes in the endogenous object $d\ln(K_j/U_j)$ must be part of the analysis of this third effect. Next, holding factor quantities constant, inequality increases given capital-skill complementarity if the capital bias of agglomeration economies increases more rapidly than their unskilled labor 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. Finally, changes in factor productivities influence their relative prices through secular demand shifts, as regulated by the capital share in the final term. The final two terms are the secular analogs to the shifts in agglomeration-biased forces that appear in the first and fourth terms.

In practice, decompositions using (2) to understand why wage inequality has increased more rapidly in larger cities 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. As previously discussed, the relative quantity of skilled labor in large versus small cities has changed very little since 1980, consistent with Baum-Snow and Pavan's (2013) finding 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. 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.

To derive an expression for $d\ln(K_j/U_j)$, which appears as an endogenous object in (2), we use the first-order condition from profit maximization with respect to capital and fully differentiate in logs. This yields

$$
\begin{aligned}
d\ln\left(\frac{K_j}{U_j}\right) &= \psi_1^K\left(\omega_j^{cu},\omega_j^c,\rho,\sigma,d\mu_k,d\mu_s,d\mu_u\right)\ln D_j \\
&\quad +\psi_2^K\left(\omega_j^{cu},\omega_j^c,\rho,\sigma\right)[d\ln\frac{S_j}{U_j}+d\ln\left(\frac{A_s}{A_u}\right)] \\
&\quad +\psi_3^K\left(\omega_j^{cu},\omega_j^c,\rho,\sigma\right)[d\ln v-d\ln A_j-d\ln A_u] \\
&\quad +\psi_4^K\left(\omega_j^{cu},\omega_j^c,\rho,\sigma\right)d\ln\frac{A_k}{A_u}
\end{aligned} \tag{3}
$$

where

$$
\begin{aligned}
\psi_1^K &= \frac{(1-\sigma)(1-\omega_j^{cu})d(\mu_u-\mu_s)+[(1-\sigma)\omega_j^{cu}\omega_j^c+(\sigma-\rho)\omega_j^c+\rho]d(\mu_u-\mu_k)-d\mu_u}{(\sigma-\rho)\,\omega_j^c(1-\omega_j^{cu})-(1-\rho)(1-\omega_j^c\omega_j^{cu})} \\
\psi_2^K &= -\frac{(1-\sigma)(1-\omega_j^{cu})}{(\sigma-\rho)\omega_j^c(1-\omega_j^{cu})-(1-\rho)(1-\omega_j^{cu}\omega_j^c)} \\
\psi_3^K &= \frac{1}{(\sigma-\rho)\omega_j^c(1-\omega_j^{cu})-(1-\rho)(1-\omega_j^{cu}\omega_j^c)} \\
\psi_4^K &= -\frac{(1-\sigma)\omega_j^{cu}\omega_j^c+(\sigma-\rho)\omega_j^c+\rho}{(\sigma-\rho)\omega_j^c(1-\omega_j^{cu})-(1-\rho)(1-\omega_j^{cu}\omega_j^c)}
\end{aligned}
$$

Given that σ and ρ are both always less than 1, ψ_2^K is always positive and ψ_3^K is always negative. Meanwhile, ψ_1^K may be positive or negative, but it is equal to zero if the factor biases of agglomeration forces do not change over time. This agglomeration effect tends to be positive when $d\mu_u$ is less than $d\mu_k$ and $d\mu_s$. In this case, firms in larger cities substitute away from the relatively less productive unskilled labor and toward capital. Because ψ_3^K is negative, reductions in the price of capital promote capital adoption, as do positive TFP shocks.[17] Therefore, the gradient of $\ln\frac{K_j}{U_j}$ with respect to city size would increase with greater TFP shocks in larger cities or (trivially) if the relative number of unskilled workers decreases. In addition to the direct effects of shifts in relative skill supply and production function parameter values seen in (2), these indirect effects operate through enhancing capital intensity in larger locations, thereby increasing the productivity of skilled workers and the price of skill in such locations given capital-skill complementarity.

[17]In the empirical implementation, the variation across local labor markets in capital intensity through this channel ends up as part of an error term.

It is also instructive to consider an alternative representation for $d\ln(K_j/U_j)$, which can be obtained from the ratio of first-order conditions from cost minimization with respect to capital and unskilled labor:

$$d\ln\left(\frac{K_j}{U_j}\right) = \frac{1}{1-\rho}d\ln\left(\frac{w_j^u}{v}\right) + \frac{\rho}{1-\rho}d\left(\mu_k - \mu_u\right)\ln D_j + \frac{\rho}{1-\rho}d\ln\left(\frac{A_k}{A_u}\right) \tag{4}$$

The first term simply reflects the relative price effects, where $\frac{1}{1-\rho}$ is the elasticity of substitution between capital and unskilled labor. Potential reasons for which unskilled labor may have become relatively more costly, or its relative marginal product may have increased, in larger cities can be seen in (3). These locations may have experienced more rapid increases in factor-unbiased agglomeration economies, unskilled-biased agglomeration economies, or declines in the relative supply of unskilled labor. Once these price effects are held constant, a more direct agglomeration mechanism becomes clearer, as can be seen in (4). Holding factor prices constant and provided that $0 < \rho < 1$ (i.e., that capital and unskilled labor are sufficiently substitutable), an increase in the capital bias of agglomeration forces increases relative capital intensity, whereas an increase in their unskilled bias decreases relative capital intensity, as is intuitive. Additionally, as is also intuitive, increases in the productivity of capital relative to unskilled labor increases capital accumulation.

Returning to (2) and substituting in for $d\ln(K_j/U_j)$ using (3) yields

$$\begin{aligned} d\ln\left(\frac{w_j^s}{w_j^u}\right) &= \psi_1^{SU}\left(d\mu_s, d\mu_k, d\mu_u, \sigma, \rho, \omega_j^c, \omega_j^{cu}\right)\ln D_j \\ &\quad +\psi_2^{SU}\left(\sigma, \rho, \omega_j^c, \omega_j^{cu}\right)\left[d\ln\left(\frac{S_j}{U_j}\right) + d\ln\left(\frac{A_s}{A_u}\right)\right] + d\ln\left(\frac{A_s}{A_u}\right) \\ &\quad +\psi_3^{SU}\left(\sigma, \rho, \omega_j^c, \omega_j^{cu}\right)\left[d\ln v - d\ln A_j - d\ln A_k\right] \end{aligned} \tag{5}$$

In (5), ψ_3^{SU} is negative if $\sigma < \rho$, ψ_2^{SU} is always negative, and ψ_1^{SU} is a more complicated object that depends on the relative strengths of the factor-biased agglomeration forces but tends to be positive when $d\mu_s$ is positive and $d\mu_u$ is negative. These coefficients are written out explicitly in the Online Appendix.[18]

Estimation of the empirical counterpart to (5) alone provides some information about model parameters. In order to recover information about capital-unskilled labor substitutibility, however, we need an additional equation that relates capital intensity to skill

[18] We address the endogeneity of $d\ln(S_j/U_j)$ in the empirical implementation using immigration shocks.

intensity and market scale. Using the first-order condition for profit maximization with respect to capital and the totally differentiated production function, we derive

$$
\begin{aligned}
d\ln\left(\frac{K_j}{Y_j}\right) \quad = \quad & \frac{(\rho-\sigma)(1-\omega_j^c)(1-\omega_j^{cu})d(\mu_u-\mu_s)+(\sigma\omega_j^{cu}\omega_j^c+(\rho-\sigma)\omega_j^c-\rho)d\mu_k}{(1-\sigma)\omega_j^{cu}\omega_j^c+(\sigma-\rho)\omega_j^c+\rho-1}\ln D_j \\
& +\frac{(\rho-\sigma)(1-\omega_j^c)}{1-\sigma}\psi_2^K[d\ln\frac{S_j}{U_j}+d\ln\left(\frac{A_s}{A_u}\right)] \qquad (6) \\
& +(1-\omega_j^{cu}\omega_j^c)\psi_3^K[d\ln v-d\ln A_j-d\ln A_k]-d\ln A_k-d\ln A_j
\end{aligned}
$$

This equation is quite similar to (3), but in this case the coefficient on $d\ln\frac{S_j}{U_j}$ is positive if and only if $\sigma<\rho$; that is, if there is capital-skill complementarity. The formulation given in (6) is more convenient than (3) for the empirical analysis, as it will allow us to directly empirically verify this sign, providing direct evidence of capital-skill complementarity. This is a generalization of the setup and procedure used in Lewis (2011).

Finally, manipulation of first-order conditions for cost minimization with respect to capital and unskilled labor yields

$$
d\ln w_j^u=\rho d\left(\mu_u-\mu_k\right)\ln D_j+(1-\rho)\,d\ln\left(\frac{K_j}{U_j}\right)+d\ln v-\rho d\ln\left(\frac{A_k}{A_u}\right) \qquad (7)
$$

This equation indicates that unskilled labor receives higher wage increases in larger cities if the agglomeration economies become more unskilled biased, when capital intensity increases as regulated by the elasticity of substitution $\frac{1}{1-\rho}$, or when capital gets more expensive. If capital and skill are sufficiently substitutable ($\rho>0$), capital-biased technical change pushes unskilled wages down while unskilled-biased technical change pushes skilled wages up, holding capital intensity constant. Substituting for $d\ln\left(\frac{K_j}{U_j}\right)$ with (3) we obtain

$$
\begin{aligned}
d\ln w_j^u \quad = \quad & \left[\rho d\left(\mu_u-\mu_k\right)+(1-\rho)\psi_1^K\right]\ln D_j \\
& +(1-\rho)\psi_2^K\left[d\ln\left(\frac{S_j}{U_j}\right)+d\ln\left(\frac{A_s}{A_u}\right)\right] \qquad (8) \\
& +(1-\rho)\psi_3^K\left[d\ln v-d\ln A_j-d\ln A_k\right] \\
& +d\ln v-\rho d\ln\left(\frac{A_k}{A_u}\right)
\end{aligned}
$$

Equations (5), (6), and (8) are the three fundamental theoretical factor demand equations that we use in empirical implementation.

4 Estimation

In this section, we show how we estimate parameters of the model. First, we discuss the empirical counterparts to the three key structural equations of the model. Next, we show how we isolate exogenous variation from labor supply conditions. Finally, we discuss econometric identification more generally.

4.1 Estimating Equations

The three structural equations that we implement empirically are all linear in the same three right-hand side variables and nonlinear in parameters and factor shares. One term is a linear function of the log agglomeration measure, a second term is linear in $d\ln\frac{S_j}{U_j}$, and a third term contains elements that are not observed, including changes in the price of capital, factor productivities, and location-specific TFP. For empirical implementation, we account for $d\ln v$, $d\ln A_s$, $d\ln A_u$, and $d\ln A_k$ with decade fixed effects, and we decompose $d\ln A_j$ into an additional decade fixed effect plus a mean zero error term. Each of these objects is multiplied by coefficients that depend on the parameters of the model and factor shares. To be consistent with Lewis (2011) and for identification reasons discussed below, we also control for the initial gap between log skilled and unskilled hours worked among all CBSA immigrants. We also generalize from the model to allow errors to have an arbitrary covariance structure across equations and over time within CBSAs.

We empirically implement the model in two ways. 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 empirical 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 all time effects to differ across equations.[19] In the "full" version of the empirical model, we retain different time effects in each equation, as they have distinct structural interpretations, and estimate a second set of common time effects across equations (structurally interpreted as $d\ln v - E[d\ln A_j] - d\ln A_k$) interacted with a heterogeneous coefficient laid out in the model. In addition, we specify the coefficients on $\ln D_j$ from the structural equations. With

[19] Because, as shown below, $\ln D_j$ is uncorrelated with our identifying variation in $d\ln(S_j/U_j)$, 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.

the full model, we can additionally identify estimates of changes in the three agglomeration parameters, $d\mu_s, d\mu_k$, and $d\mu_u$ and glean information about skill-biased technical change $d\ln(A_s/A_u)$. Exact specifications of the two sets of structural equations and structural interpretations of all estimated fixed effects in the full model are listed in Appendix A.

Assuming exogenous variation across CBSAs in $d\ln(S_j/U_j)$ and $\ln D_j$, which we establish below, the following types of comparisons in the data allow us to identify the parameters of interest. Conditional on city size, responses of relative wages, skilled wages, and capital intensity to variation across CBSAs in relative labor supply shocks provide 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, (6) is a central estimation equation. The parameters that 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 that receive the same labor supply shock. In theory, we are also able to identify various linear combinations of changes in productivity and the real price of capital, including notably $d\ln(A_s/A_u)$. However, identification of these objects comes off decade fixed effects rather than clean exogenous variation in labor supply shocks across CBSAs. Therefore, we treat these estimates with caution.[20]

The clearest difficulty for successful recovery of model parameters is that the stocks of skilled and unskilled workers in each city at each point in time are equilibrium outcomes. Likely correlations between unobservables like TFP and the change in the skill ratio $d\ln\frac{S_j}{U_j}$ makes credible identification difficult. For example, locations with growing TFP may have other attributes, such as better consumer amenities, that disproportionately draw in skilled workers. Therefore, identification of parameters requires exogenous variation in changes in the relative supply of skill across metropolitan areas. One can view our use of exogenous shocks to $d\ln(S_j/U_j)$ as part of a "first-stage" equation or, equivalently, as incorporating labor supply conditions into the model. The following sub-section develops such a labor supply environment that justifies our specification of this additional estimation equation for relative factor supply.

[20] Appendix A describes the parameter clusters that are identified in each estimation equation.

4.2 Labor Supply Environment

We achieve exogenous variation in the relative supply of skill 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 in the U.S. after 1980 are more likely to settle in the Los Angeles CBSA 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. We now develop a model of labor supply that incorporates these ideas to justify the use of a simulated instrument for $d\ln\frac{S_j}{U_j}$ constructed using immigration shocks. Our discussion is primarily intended to fix ideas about conditions under which immigration shocks can be made into useful sources of exogenous variation in local labor supply.

4.2.1 Theory

We think of the supply of labor to each local labor market as generated from a standard Rosen (1979) and Roback (1982) type condition that workers are indifferent across locations in long run equilibrium. Consumers, which are heterogeneous in skill g and country of origin o, have preferences over consumption x, housing h, and local quality of life q.[21] Some immigrants from each country of origin are assumed to live in every CBSA j. We differentiate countries of origin because variation in labor supply as a function of country of origin will be the ultimate source of identifying variation. Commensurate with the discussion in Glaeser (2008) and applications in Albouy (2016) and Diamond (2016), the resulting long-run equilibrium condition used as a basis for understanding labor supply shocks for each worker type go is:

$$V(p_j, w_j^g; q_j^{go}) = \overline{v}^{go} \tag{9}$$

Variation in amenity values across locations for each group q_j^{go} manifests itself as static labor supply shifters to local labor markets for each go. Preferences imply labor supply

[21] U.S. natives have their own o index value.

and housing demand functions for each group that depend on wages and house prices in each CBSA.

The process for deriving equilibrium prices and allocations is similar to that carried out in Albouy and Stuart (2014). Equilibrium vectors of wages w^s and w^u and house prices p solve the system of equations which describe two labor supply conditions, one housing supply condition, an aggregate housing demand condition and two labor demand conditions given by (5) and (8) for each CBSA, while imposing national adding-up constraints for skilled and unskilled populations. Conditional on housing supply and labor demand functions, higher amenity places have higher equilibrium populations, thereby bidding housing prices up and wages down to equalize indirect utilities across locations.

We assume that exogenous masses of immigrants $dN^{so} \equiv dS^o$ and $dN^{uo} \equiv dU^o$ arrive from abroad to the U.S. over the course of each decade after 1980. These new arrivals have the choice of which local labor market in which to settle. If migrants have the same amenity valuation of locations as incumbents from their countries of origin, (9) indicates that these new arrivals must be indifferent between all locations on the margin. To break this indifference, we introduce a cost of moving to CBSA j, which can capture the availability of job referrals, home referrals, and general knowledge about how to get settled. Individuals choose the destination that maximizes utility, anticipating new arrivals' location choices and endogenous location responses of natives. Incumbents may move because immigration changes real wages in each location.

To formalize what (9) implies about changes over time in local labor supply functions with migration costs, we follow Notowidigdo (2013) and introduce the cost distributions $m^g(d \ln N_j^{go}, N_{j0}^{go})$ of moving to j. Migration costs exhibit $m_1^g > 0$, $m_2^g < 0$, and $m^g(0, N_{j0}^o) = 0$, where subscripts indicate partial derivatives and N_{j0}^{go} is the stock of immigrants from o of skill group g in CBSA j in some pre-analysis time period. That is, each additional migrant finds it more costly to move to local labor market j, but this cost is declining in the number of incumbents of type go. This is a formalization of the mechanism proposed in Card (2001). People of each type flow in or out of each CBSA until the marginal person has a change in utility that is the same in every CBSA, though migration frictions mean that positive local shocks result in relative utility gains and negative local shocks result in relative utility losses for incumbents. This delivers an equilibrium population change of each type to each CBSA that comes about because of shocks to the quantity of immigrants entering the U.S. from each country in addition to a rearrangement of natives across local labor markets.

Differentiating (9) and applying Roy's identity yields the following labor supply equation expressed in first differences:[22]

$$d\ln w_j^g - \beta_g d\ln p_j - M^g(d\ln N_j^{go}, \ln N_{j0}^{go}) = \zeta d\ln \overline{v}^{go} \tag{10}$$

In (10), the migration cost acts as a wedge between trends in local real wages and national trends in real wages, generating an equilibrium change in population of each group go. Specifying $M = \alpha_g^1 d\ln N_j^{go} - \alpha_g^2 N_{j0}^{go}$ (without loss of generality), we derive the labor supply expression

$$\begin{aligned} d\ln N_j^g \quad = \quad & \frac{1}{\alpha_g^1}\left[d\ln w_j^g - \beta_g d\ln p_j\right] \\ & -\frac{1}{\alpha_g^1}\sum_o \frac{N_j^{go}}{N_j^g}\sum_{j'\neq j}\frac{N_{j'}^{go}}{N^{go}}\left[d\ln w_{j'}^g - \beta_g d\ln p_{j'} + \alpha_g^2 N_{j'0}^{go}\right] \\ & +\sum_o \frac{N_j^{go}}{N_j^g}\left[d\ln N^{go} + \frac{\alpha_g^2}{\alpha_g^1}N_{j0}^{go}\right] \end{aligned} \tag{11}$$

This expression has three intuitive components. First, we see that population change of type g in local labor market j is increasing in the change in the real wage, as regulated by the marginal migration cost. Second, population increases less in j if competing local labor markets have high increases in real wages and/or lower migration costs. Third, population increases more in CBSAs with a greater prevalence of incumbents from countries of origin receiving greater inflows from abroad to the U.S. This final term does not include any (endogenous) prices and is not codetermined through interactions with labor demand or the housing market.

We emphasize that because (11) includes wages and house prices, it is not appropriate to use it for estimation in this form. Instead, we think of wages and house prices as being functions of exogenous objects in the reduced-form system of equations for equilibrium wages and prices. Our aim is merely to highlight the fact that the final term in (11) shifts labor supply without shifting labor demand, and can thus be used to construct an instrument for the change in the relative supply of skill across CBSAs.

[22]In (10), $M^g(\cdot) \equiv \zeta\frac{m^g(\cdot)}{\overline{v}^{go}}$. We assume $\zeta = 1/\frac{d\ln V}{d\ln w}$ is a constant. The national change in utility for group go $d\ln\overline{v}^{go}$ is an endogenous element which can be solved for given the constraint that all new arrivals must settle in a location.

4.2.2 Empirical Implementation

We consider the following empirical counterpart to the theoretical labor supply equation (11), which can be thought of as a reduced-form expression for employment growth of skill group g in CBSA j:[23]

$$\Delta_t \ln N_j^g = \delta_t + \alpha_1 \Delta_t \ln \widehat{N}_j^g + \alpha_2 \ln N_{jt-1}^{g,imm} + \alpha_3 \ln D_j + u_{jt} \tag{12}$$

In this equation, Δ_t denotes the difference between periods t and $t-1$ and the superscript imm indicates that these variables are for stocks of all immigrants. $\Delta_t \ln \widehat{N}_j^g$ is the empirical counterpart of $\sum_o \left(\frac{N_j^{go}}{N_j^g} d \ln N^{go} \right)$ in (11) and indicates the change in log number of immigrants in CBSA j predicted using initial year CBSA composition and flows into the U.S. Following Lewis (2011), we include the regressor $\ln N_{jt-1}^{g,imm}$ 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. Assuming truly exogenous variation in $\ln \widehat{N}_j^g$, it should not be necessary to control for anything. Indeed, this exogenous variation is what allows us to subsume most terms on the right hand side of (11) into the time effects and the error term and still recover consistent estimates of α_1.

To be consistent with the estimating equations, the outcome variable uses manufacturing workers only, whereas the calculation of $\Delta_t \ln \widehat{N}_j^g$ uses all immigrants. To minimize the possibility that country of origin shares $\frac{N_j^{go}}{N_j^g}$ are related to unobservables driving shifts in the wage structure, we calculate these using data from 1970. In particular, we calculate $\Delta_t \ln \widehat{N}_j^g$ as

$$\Delta_t \ln \widehat{N}_j^g = \sum_o \left(\frac{N_{j,1970}^{go}}{N_{j,1970}^g} (\ln N_t^{go} - \ln N_{t-1}^{go}) \right)$$

where $N_{j,1970}^{go}$ is the year 1970 number of residents in skill group g from country of origin o living in CBSA j, $N_{j,1970}^g$ is the total number of year 1970 residents of CBSA j in skill group g, and N_t^{go} is the total number of U.S. workers from country of origin o in skill group g. 1970 shares are calculated using all immigrant residents at least 25 years old. N_t^{go} is constructed using workers aged 16 to 65 who were born outside of the U.S. and were not

[23] It may also be appropriate to incorporate housing supply elasticities in the estimation equations. The best estimates of this object are in Saiz (2010), but only for 289 metropolitan areas. Therefore, to maintain our sample size, we do not incorporate these elasticities.

living in group quarters. We use the same variable to instrument for both raw counts and efficiency units.

For our identification strategy to be valid, it must be the case 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 U.S. 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 as well as post-1980 productivity growth.

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 about by immigration-induced labor supply shifts. Using national data, Borjas (2003) finds sizable and significant negative effects of immigration on the wages of native workers, providing indirect evidence of such induced migration, and implying that native and immigrant labor are close substitutes. However, Card and 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 less productive than natives given perfect substitutability. Dustmann et al. (2013) and Ottaviano and Peri (2012) provide evidence that immigrants and natives are likely not perfect substitutes, and contend that 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 seems 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 that we calculate efficiency units using different weights on country of origin for skilled and unskilled workers. Because 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.

Table 4 presents OLS estimates of coefficients in (12). These results indicate 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. The regressions are all weighted by 1980 CBSA population, and standard errors are clustered by CBSA. 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 3.7 percent more manufacturing workers among high school dropouts, 2.3 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 CBSA 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.

In the context of our full structural model, parameter estimates of factor demand equations 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 main estimating equations specified above. Differencing (12) across our two main aggregated skill groups yields the following fourth estimating equation of our empirical model:

$$\Delta_t \ln \frac{S_j}{U_j} = \delta_t + \alpha_1 \Delta_t \ln \frac{\widehat{S}_j}{\widehat{U}_j} + \alpha_2 \ln \frac{S^{imm}_{jt-1}}{U^{imm}_{jt-1}} + \alpha_3 \ln D_j + u_{jt} \tag{13}$$

Using four cross-sections of data starting in 1980, we estimate parameters of this equation with a panel that has three observations per CBSA. As shown in the first and fifth columns of Table 5, in regressions weighted by 1980 CBSA population, our estimates of the key parameter α_1 in (13) is 0.21 when predicting raw counts and 0.17 when predicting efficiency units. These estimates are both statistically significant, with t-values of more than 3.5 when standard errors are clustered by CBSA.[24]

4.3 Estimation Procedure and Identification

We jointly estimate the structural parameters of the partially linear system of equations (5), (6), and (8) using feasible generalized simultaneous nonlinear least squares (FGNLS). Each of the structural equations whose parameters we estimate has the following form, with parameter vector θ:

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

These equations are specified explicitly in Appendix A. 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 (13) which has exogenous predictors Z with coefficients δ yields

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

and no remaining correlation between W or Z with the error term. Therefore, the parameters θ are identified and can be consistently estimated (Cai et al. 2006).[25]

Consistent identification of parameters formally requires that $E[Z\varepsilon] = 0$, where Z consists of $\Delta_t \ln \frac{\widehat{S}_j}{\widehat{U}_j}$, $\ln D_j$, $\ln \frac{S^{imm}_{jt-1}}{U^{imm}_{jt-1}}$, and time fixed effects. As is laid out in Appendix A, ϵ in each equation is $d\ln A_j - E[d\ln A_j]$, with a different heterogeneous coefficient for each equation. Within equation, we allow the errors to be correlated over time within CBSA and to be heteroskedastic. We also allow errors to be correlated across equations.

[24]It is possible to specify this equation with different coefficients on variables describing skilled and unskilled workers. Because the absolute values of coefficients on these two sets of predictors are not significantly different, disaggregating them in this way does not affect the results.

[25]In our model, the shares ω^c_j and ω^{cu}_j appear in W and may be endogenous as we measure them. We address this potential endogeneity problem in robustness checks below.

5 Results

This section presents estimates of σ, ρ, $d\mu_s$, $d\mu_u$, and $d\mu_k$. Using these parameter estimates, we then 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 (5), (6), and (8). 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 economics literature, or generalizations with additional factors that enter separably into the production function, the coefficient on $\Delta \ln(S_j/U_j)$ in the fourth and eighth columns of Table 5 is the elasticity of substitution between skilled and unskilled labor. 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 2.1. The first and fifth columns show first-stage results, as described 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 are not sensitive to whether labor is measured in efficiency units. The second and sixth columns indicate that exogenous increases in the relative supply of skilled workers led to capital intensification. As is seen in (6), 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 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 (6) 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 the 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 wage gaps are higher in more populous metropolitan areas. Based on (5), these positive coefficients are evidence of increases in the skill bias of agglomeration economies. Our theoretical framework predicts a positive impact of the relative supply of skilled labor on unskilled wages. Although this is true in Table 5, the relationship is only statistically significant for raw counts, hinting that ρ may be close to 1 (which we confirm below). All IV estimates in Table 5 exceed corresponding OLS estimates by at least a factor of two. This ordering is consistent with a positive correlation between $d\ln(S_j/U_j)$ and relative labor demand shocks, just the sort of endogeneity problem our empirical strategy is designed to overcome.

A large literature going back to Katz and Murphy (1992) estimates the elasticity of substitution between skilled and unskilled labor using equations like those estimated in the fourth and eighth columns of Table 5. The first two terms of (2) form the structural equation underlying this regression in a two-factor model with agglomeration economies. As summarized in Ciccone and 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 results in the fourth column in the context of a two-factor model, our coefficient estimate of -0.43 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 5 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^{cu} differ across CBSAs. This makes reduced-form coefficients difficult to interpret and obviates the possibility of recovering estimates of structural parameters of interest.

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 $d \ln \frac{S_j}{U_j}$ (the "sparse" empirical model). Decade fixed effects fully interacted with city size that are allowed to differ across equations control for the other terms. This specification allows us to recover estimates of σ and ρ only, as it identifies parameters strictly from variation in $d \ln \frac{S_j}{U_j}$. We estimate σ to be 0.65 using raw labor hours and 0.78 using efficiency units. We estimate ρ to be 0.95 for raw hours and 0.97 for efficiency units. 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 empirical model, which additionally identifies the agglomeration parameters $d\mu_u$, $d\mu_s$, and $d\mu_k$. Full model estimates of ρ are very close to those from the sparse model, though estimates of σ rise to 0.79 for raw units and 0.88 for efficiency units. We find that agglomeration economies have become significantly more biased toward skilled labor, while the agglomeration multiplier on unskilled labor has slightly declined over time. The agglomeration bias of capital has remained stable. 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 main driver of more rapidly increasing returns to skill in larger cities. This comes in the context of skill-biased technical change. In particular, our estimates of $d \ln(A_s/A_u)$ are about 0.3 and statistically significant. The decade fixed effect, interpreted as capturing $d \ln v - E[d \ln A_j] - d \ln A_k$, is marginally significantly negative in the 1980s and insignificantly positive in other periods (unreported). However, because fixed effect estimates are not as directly identified off exogenous variation in $\Delta \ln(S/U)$, they should be interpreted with caution.

There are only a few other papers that simultaneously structurally estimate σ and ρ or transformations thereof. Krusell et al. (2000) find evidence of capital-skill complementarity in a model in which skilled labor is nested with capital using time-series data from 1963 to 1992 in the U.S. They estimate an elasticity of substitution between skilled labor or capital and unskilled labor of 1.67 and between skilled labor and capital of 0.67. Duffy et al. (2004) use data from a panel of countries and find evidence of capital-skill complementarity, 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 now 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 described in (2) can explain the increasingly positive relationship between wage inequality and city size. Table 7 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 $\ln D_j$.[26] They use actual data to measure $\Delta_t \ln \frac{S_j}{U_j}$ and $\Delta_t \ln \frac{K_j}{U_j}$ 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 if there is no capital-skill complementarity, the third and fourth terms in (2) are 0 and the full change in the distribution of wage 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 show that the growth in the wage gap predicted using estimated parameters has a 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 approximately match the actual elasticities of 0.015 and 0.012 observed in the data. Using actual data on factors to build components of (2) changes the elasticity with respect to city size to 0.013 for

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

raw units and 0.011 for efficiency units. 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 comes 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 (13). That is, the small discrepancies between the numbers in Rows 2 and 3 of Table 7 indicate the extent to which more rapid increases in skill intensity in larger cities are endogenously driven.

Results in the remaining rows of Table 7 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 six terms in (2), the agglomeration term is the largest, accounting for at least 80 percent of the elasticity with respect to city size. We find that relative labor supply shocks account for less than 5 percent of this elasticity and that more rapid capital accumulation in larger cities accounts for 8-15 percent. Results in the remaining two rows indicate that changing relative factor productivities had essentially no effect on the relationship between wage inequality and city size.

We similarly investigate the reasons for which capital intensity increased more rapidly in larger cities. Table A2 reports elasticities of each term in (3) with respect to city size using estimated parameters reported in Table 6. This exercise is somewhat less successful than that for $\Delta_t \ln \frac{w_j^s}{w_j^u}$, as the model does a poorer job of matching $\Delta_t \ln \frac{K_j}{U_j}$. The predicted elasticities of $\Delta_t \ln \frac{K_j}{U_j}$ with respect to city size markedly exceed the actual values of about 0.005. Using the predicted elasticity of 0.035 for efficiency units, we find that the increasing skill bias of agglomeration economies, relative labor supply shifts, and skill-biased technical change were the three most important factors pushing toward more rapid capital accumulation in larger cities.

We replicate the same analysis after separately calibrating ρ, $d\mu_k$, and σ such that the predicted elasticity of $\Delta_t \ln \frac{K_j}{U_j}$ with respect to city size matches the data, as reported in Table A2 Panels B, C, and D respectively. Calibrating ρ or $d\mu_k$ in this way does not change the main mechanisms driving more rapid capital accumulation in larger cities, though with ρ calibrated to almost 1 the impact of skill-biased technical change is small. If instead σ is calibrated so that the predicted elasticity matches the data, the impacts of relative supply shifts and skill-biased technical change turn negative, with the gap made up for with an even stronger positive importance of increases in the skill bias of agglomeration economies. Therefore, we conclude that increases in the skill bias of agglomeration economies have been central for generating more rapid capital accumulation in larger cities. We do not have robust enough results to comment on the importance of other potential mechanisms.

It is tempting to use this framework to try to understand secular changes in wage inequality, not just its variation across local labor markets. Our estimates of $d\ln(A_s/A_u)$ of about 0.3 indicate that factor-biased technical change has been a centrally important driver of the growth of wage inequality. However, because in estimation secular changes in $\Delta_t \ln \frac{w_j^s}{w_j^u}$ mostly get subsumed into constants that capture multiple forces plus a control for base period immigrant stocks by skill, this framework is not well-suited to perform a unified statistical decomposition of this sort. Because our identifying variation is cross-sectional, it seems doubtful that any well-identified empirical study similar to ours would be able to credibly carry out such a decomposition. Instead, our results are best suited for understanding the drivers of relative changes in wage inequality across local labor markets.

5.4 Robustness

Table 8 reports results of robustness checks in which we estimate the model using sub-samples of the data. 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.[27] Results in columns (3) and (4) 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^{cu} that are predicted using shares from 1987 and 1997, rather than data from 1992 and 2002 directly. 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.

As with the main estimates in Table 6, results in Table 8 using data from each sub-period also show strong evidence of capital-skill complementarity and increases in the skill bias of agglomeration economies ($d\mu_s - d\mu_u > 0$) in the context of skill-biased technical change. For efficiency units, estimates of σ and ρ are stable over time in the full model specification. While estimates of $d\mu_s$ in particular bounce around some, and appear to be more positive in the 1980s, they remain statistically larger than estimates of $d\mu_u$ for all sub-periods.

[27] 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.

Rather than estimate the model using fewer time periods, we also considered the system of structural equations that comes about if σ and ρ are allowed to explicitly change over time. Unfortunately, estimation of this version of the model is not feasible given that it requires isolating separate exogenous variation in levels of and changes in relative skill intensity. Given the reasonable stability of σ and ρ over time in Table 8, we are not too concerned about this potential source of model mis-specification.

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 are not reported here, but are very similar to those reported in Table 6. This is consistent with the fact that these two capital measures are highly correlated.

5.5 Nesting Skilled Labor with Capital

Throughout this analysis, we have imposed that unskilled labor and capital are nested together in the production function. We have adopted this nesting structure mainly because it fits the data slightly better than under alternative nesting assumptions. Our primary specification can also be viewed as a generalization of that adopted by Autor and Dorn (2013). The literature has more commonly employed the specification in which skilled labor is nested with capital instead, as in Krusell et al. (2000). Estimating the model with this alternative nesting reduces the R-squared on the estimated version of Equation (3) from 0.159 to 0.157, though the R-squared rises slightly for the other two primary estimation equations. In this alternative nesting, the elasticity of substitution $\frac{1}{1-\sigma}$ is between unskilled labor and capital or skilled labor, and $\frac{1}{1-\rho}$ is between capital and skilled labor. The associated factor demand equations are identical to (2), (3), and (6) that we use for estimation above, except that S and U are swapped throughout.

Estimation results using this alternative nesting yield similar conclusions to those from Tables 6 and 7, subject to the different constraints imposed by the alternative production function specification. This is consistent with the IV regression results in Table 5 showing that, without explicitly imposing a particular production function nesting, capital and unskilled labor are more substitutible in production than capital and skilled labor. Table 9 reports parameter estimates for the alternative nesting. For the full model, σ is estimated to be 0.84 (raw units) or 0.87 (efficiency units), and ρ is estimated to be -0.54 (raw units) or -0.67 (efficiency units) with large standard errors. The fact that $\widehat{\sigma}$ is now greater than $\widehat{\rho}$ indicates, as before, that capital and skilled labor are more complementary than capital and unskilled labor. As with the results in Table 6, the skill bias of agglomeration economies is estimated to be significantly increasing over time, and the unskill bias of agglomeration economies significantly declining over time. Unlike the results in Table 6, however, we estimate significant declines in the capital bias of agglomeration economies.

Taken together, results in Tables 6 and 9 indicate that the order of elasticities of substitution from lowest to highest is capital and unskilled labor, skilled and unskilled labor, and capital and skilled labor. Since two of these elasticities are constrained to be identical in each model specification, each nesting choice has its advantages. With the elasticity of substitution between skilled labor and capital more similar to that between skilled and unskilled labor, parameter estimates from our primary specification are more stable.

Table 10 reports results of decomposing relationships between wage gaps and city size into components, as in Table 7, using the alternative model. The relevant structural equation with the alternative nesting is

$$\begin{aligned} d\ln\frac{w_j^s}{w_j^u} &= \sigma d(\mu_s-\mu_u)\ln D_j + (\sigma-1)d\ln\left(\frac{S_j}{U_j}\right) + (\sigma-\rho)\omega_j^c d\ln\left(\frac{K_j}{S_j}\right) \\ &\quad +(\sigma-\rho)\omega_j^c d(\mu_k-\mu_s)\ln D_j + \sigma d\ln\left(\frac{A_s}{A_u}\right) + (\sigma-\rho)\omega_j^c d\ln\left(\frac{A_k}{A_s}\right). \end{aligned} \tag{14}$$

As with the standard nesting, results in Table 10 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 six terms in this decomposition, the agglomeration term is still the largest, now accounting for more than the full elasticity of $d\ln\frac{w_j^s}{w_j^u}$ with respect to city size. However, more rapid capital accumulation in larger cities is now a more important part of the story. While skilled labor demand was boosted more in larger cities through its complementarity with this intensifying capital usage (third term in (14)), these increases were more than 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 (fourth term). 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. The net result is that capital-skill complementarity led to slightly less rapid increases in skilled-unskilled wage gaps in larger cities.

While some details are different when comparing results using our primary and the alternative nestings, the broad message is the same. Cities and skills have become more complementary over time, driving most of the more rapid increase in wage inequality in larger cities. While capital-skill complementarity exists, it has not interacted with city size to generate much variation in trends in wage gaps across local labor markets of different sizes. Moreover, skill-biased technical change and capital-skill complementarity have co-existed to influence secular changes in the wage structure.

6 Conclusions

This paper uses Economic and Population Census data to estimate a flexible aggregate production function that facilitates evaluating mechanisms through which the gaps between average wages of more and less educated workers have become more positively related with city size since 1980. Parameters governing elasticities of substitution between capital, skilled labor, and unskilled labor as well as shifts in the factor bias of agglomeration economies are recovered using exogenous variation in skill intensity across local labor markets from immigration shocks, while allowing for factor-biased technical change.

The results indicate that a secular increase in the bias of agglomeration economies toward skilled labor has been central for directly generating greater increases in wage inequality in larger cities. Increases in capital intensity in larger cities, driven in part by the greater complementarity between cities and skills in production, have also driven some relative increases in skill demand in larger cities given capital-skill complementarity. Given that city size accounts for about one-third of the increase in wage inequality nationwide since 1980 (Baum-Snow and 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 capital-skill complementarity is an important part of the context for understanding both secular increases in wage inequality and more rapid increases in wage inequality in larger cities, it is not sufficient to fully, or even mostly, rationalize patterns in the data across local labor markets. Skill-biased technical change is additionally needed to rationalize secular increases in wage gaps, and increases in the skill bias of agglomeration economies are additionally needed to rationalize shifts in the patterns of wage gaps across local labor markets.

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. The increase in the complementarity in production between human capital and market scale we find points to a growing 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 relative importance for generating agglomeration economies. Other proposed broad explanations for agglomeration economies, such as labor market pooling and input sharing (Duranton and Puga 2004), seem less likely 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 in a dynamic way.

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. 2016. "What Are Cities Worth? Land Rents, Local Productivity, and the Capitalization of Amenity Values." Review of Economics and Statistics, 98:3, 477-487.

Albouy, David, and Brian Stuart. 2014. "Urban Population and Amenities: The Neoclassical Model of Location." NBER Working Paper No. 19919.

Antras, Pol. 2004. "Is the US Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution." Contributions in Macroeconomics, 4:1.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." American Economic Review, 103:5, 1553-1597.

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.

Borjas, George J. 2003. "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market." Quarterly Journal of Economics, 118:4, 1335-1374.

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, and 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.

Diamond, Rebecca. 2016. “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980–2000.” The American Economic Review, 106:3, 479-524.

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, and Diego Puga. 2004. “Micro-Foundations of Urban Agglomeration Economies.” In Handbook of Urban and Regional Economics Vol. 4, edited by J.V. Henderson and J-F Thisse, 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. 2008. “Cities, Agglomeration, and Spatial Equilibrium (The Lindahl Lectures).” Oxford University Press.

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.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. “Identifying Agglomeration Spillovers: Evidence from Million Dollar Plants.” Journal of Political Economy, 118:3, 536-598.

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.

Holmes, Thomas, and Matthew Mitchell. 2008. "A Theory of Factor Allocation and Plant Size." RAND Journal of Economics, 39:2, 329-351.

Juhn, Chinhui, Kevin Murphy, and 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.

Leon-Ledesma, Miguel A., Peter McAdam, and Alpo Willman. 2010. "Identifying the Elasticity of Substitution with Biased Technical Change." American Economic Review, 100:4, 1330-1357.

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.

Notowidigdo, Matthew. 2013. "The Incidence of Local Labor Demand Shocks." manuscript.

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

Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban Quality of Life." In Current Issues in Urban Economics, edited by Peter Mieszkowski and Mahlon Straszheim, Johns Hopkins University Press.

Roback, Jennifer. 1982. "Wages, Rents and the Quality of Life." Journal of Political Economy, 90:6, 1257-1278.

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.

Saiz, Albert. 2010. "The Geographic Determinants of Housing Supply." Quarterly Journal of Economics, 125:3, 1253-1296.

Wang, Zhi. 2016. "Wage Growth, Ability Sorting, and Location Choice at Labor-Force Entry: New Evidence from U.S. Census Data." Journal of Urban Economics, 96, 112-120.

A Structural Model Estimation Specifications

The full structural model has the three equations (5), (6), and (8) plus the additional relative labor supply or "first-stage" equation (13). For the purpose of estimation, the "sparse" model 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^{cu} and ω^c are also observed in each study year, but are 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 29 parameters in the estimation equations below: $\delta_t, \lambda_t^1, \lambda_t^2, \lambda_t^3, \kappa_t^1, \kappa_t^2, \kappa_t^3, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \sigma$, and ρ . Only σ and ρ are structural parameters of the model. All other parameters are clusters of model parameters or exist only to strengthen identification. We allow the error terms $u_{jt}, \varepsilon_{jt}^1, \varepsilon_{jt}^2$, and ε_{jt}^3 to be heteroskedastic and correlated across equations and over time within CBSA. All objects that are not parameters or error terms are data.

Next to each equation, we indicate the relevant equation number in the text.

$$\Delta_t \ln \frac{S_j}{U_j} = \delta_t + \alpha_1 \Delta_t \ln \frac{\widehat{S}_j}{\widehat{U}_j} + \alpha_2 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \alpha_3 \ln D_j + u_{jt} \quad (13)$$

$$\begin{aligned}
\Delta_t \ln \frac{w_j^s}{w_j^u} &= \lambda_t^1 + \beta_1 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \kappa_t^1 \ln D_j + (\sigma - 1)\Delta_t \ln \left(\frac{S_j}{U_j}\right) + (\sigma - \rho)\omega_{jt-1}^c Q_{jt}^s + \varepsilon_{jt}^1 \quad (5) \\
\Delta_t \ln w_j^u &= \lambda_t^2 + \beta_2 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \kappa_t^2 \ln D_j - (1 - \rho)\, Q_{jt}^s + \varepsilon_{jt}^2 \quad (8) \\
\Delta_t \ln \frac{K_j}{Y_j} &= \lambda_t^3 + \beta_3 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \kappa_t^3 \ln D_j - \frac{(\rho - \sigma)(1 - \omega_{jt-1}^c)}{1 - \sigma} Q_{jt}^s + \varepsilon_{jt}^3 \quad (6)
\end{aligned}$$

In the expressions above,

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

A.2 "Full" Empirical Model

For the full empirical model, we estimate the following 27 parameters in the estimation equations below: $\delta_t, \lambda_t^1, \lambda_t^2, \lambda_t^3, \pi_t, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \sigma, \rho, d\mu_s, d\mu_u, d\mu_k$, and $d\ln\frac{A_s}{A_u}$. $\sigma, \rho, d\mu_s, d\mu_u, d\mu_k$, and $d\ln\frac{A_s}{A_u}$ are structural parameters of the model. Other estimated parameters are clusters of model parameters or exist only to strengthen identification. Structural interpretation of fixed effects and error terms are listed after each equation. We allow the error terms $u_{jt}, \varepsilon_{jt}^1, \varepsilon_{jt}^2$, and$\varepsilon_{jt}^3$ to be heteroskedastic and correlated across equations and over time within CBSA. All objects that are not parameters or error terms are data.

$$\Delta_t \ln\frac{S_j}{U_j} = \delta_t + \alpha_1 \Delta_t \ln\frac{\widehat{S}_j}{\widehat{U}_j} + \alpha_2 \ln\frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \alpha_3 \ln D_j + u_{jt} \quad (13)$$

$$\begin{aligned}\Delta_t \ln\frac{w_j^s}{w_j^u} \quad = \quad & \lambda_t^1 + \beta_1 \ln\frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \sigma(d\mu_s - d\mu_u)\ln D_j + (\sigma-1)\Delta_t \ln\left(\frac{S_j}{U_j}\right) \\ & -(\sigma-\rho)\omega_{jt-1}^c Q_{jt} - (\sigma-\rho)\omega_{jt-1}^c (d\mu_k - d\mu_u)\ln D_j + \varepsilon_{jt}^1 \quad (5)\end{aligned}$$

In the context of the model, $\lambda_t^1 = \sigma d\ln\left(\frac{A_s}{A_u}\right)$ and $\varepsilon_{jt}^1 = -\frac{(\sigma-\rho)\omega_{j,t-1}^c}{(1-\sigma)\omega_{jt-1}^{cu}\omega_{jt-1}^c + (\sigma-\rho)\omega_{jt-1}^c + \rho - 1}(d\ln A_j - E[d\ln A_j])$. We index λ^1 by time because it also plays the role of accounting for secular changes in immigration.

$$\Delta_t \ln w_j^u = \lambda_t^2 + \beta_2 \ln\frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \rho\,(d\mu_u - d\mu_k)\ln D_j + (1-\rho)\,Q_{jt} + \varepsilon_{jt}^2 \quad (8)$$

In the context of the model, $\lambda_t^2 = d\ln v - d\ln\left(\frac{A_k}{A_u}\right)$ and $\varepsilon_{jt}^2 = \frac{1-\rho}{(1-\sigma)\omega_{jt-1}^{cu}\omega_{jt-1}^c + (\sigma-\rho)\omega_{jt-1}^c + \rho - 1}(d\ln A_j - E[d\ln A_j])$.

$$
\begin{aligned}
\Delta_t \ln \frac{K_j}{Y_j} &= \lambda_t^3 + \beta_3 \ln \frac{S_{jt-1}^{imm}}{U_{jt-1}^{imm}} + \frac{1-\omega_{jt-1}^{cu}\omega_{jt-1}^{c}}{(1-\sigma)\omega_{jt-1}^{cu}\omega_{jt-1}^{c} + (\sigma-\rho)\omega_{jt-1}^{c} + \rho - 1}\pi_t \\
&+ \frac{(\rho-\sigma)(1-\omega_{jt-1}^{c})(1-\omega_{jt-1}^{cu})(d\mu_u - d\mu_s) + (\sigma\omega_{jt-1}^{cu}\omega_{jt-1}^{c} + (\rho-\sigma)\omega_{jt-1}^{c} - \rho)d\mu_k}{(1-\sigma)\omega_{jt-1}^{cu}\omega_{jt-1}^{c} + (\sigma-\rho)\omega_{jt-1}^{c} + \rho - 1} \ln D_j \\
&- \frac{(\rho-\sigma)(1-\omega_{jt-1}^{c})(1-\omega_{jt-1}^{cu})}{(1-\sigma)\omega_{jt-1}^{cs}\omega_{jt-1}^{c} + (\sigma-\rho)\omega_{jt-1}^{c} + \rho - 1}\Delta_t \ln \frac{S_j}{U_j} + \varepsilon_{jt}^3 \qquad (6)
\end{aligned}
$$

In the context of the model, $\lambda_t^3 = -E(d\ln A_j) - d\ln A_k$ and $\varepsilon_{jt}^3 = (\frac{1-\omega_j^{cu}\omega_j^c}{(1-\sigma)\omega_{jt-1}^{cu}\omega_{jt-1}^{c} + (\sigma-\rho)\omega_{jt-1}^{c} + \rho - 1} + 1)(d\ln A_j - E[d\ln A_j])$. In the expressions above,

$$
\begin{aligned}
Q_{jt} &= \frac{(1-\sigma)(1-\omega_{jt-1}^{cu})(d\mu_u - d\mu_s) + ((1-\sigma)\omega_{jt-1}^{cu}\omega_{jt-1}^{c} + (\sigma-\rho)\omega_{jt-1}^{c} + \rho)(d\mu_u - d\mu_k) - d\mu_u}{(1-\sigma)\omega_{jt-1}^{cu}\omega_{jt-1}^{c} + (\sigma-\rho)\omega_{jt-1}^{c} + \rho - 1} \ln D_j \\
&- \frac{(1-\sigma)(1-\omega_{jt-1}^{cu})}{(1-\sigma)\omega_{jt-1}^{cu}\omega_{jt-1}^{c} + (\sigma-\rho)\omega_{jt-1}^{c} + \rho - 1}[\Delta_t \ln \frac{S_j}{U_j} + d\ln \frac{A_s}{A_u}] \\
&+ \frac{\pi_t}{(1-\sigma)\omega_{jt-1}^{cu}\omega_{jt-1}^{c} + (\sigma-\rho)\omega_{jt-1}^{c} + \rho - 1}
\end{aligned}
$$

In the context of the model, $\pi_t = d\ln v - E[d\ln A_j] - d\ln A_k$.

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

Skilled Worker Defn. Unskilled Worker Defn.	Wage Premium For Indicated Skill Definitions Some College+ High School-	College+ Some College-	College+ High School-	College Only High School Only	Elasticity of Wages With Respect to 1980 CBSA Pop.
			Panel A: All Workers		
1980	0.25	0.41	0.43	0.34	0.048
1990	0.36	0.51	0.58	0.44	0.070
2000	0.40	0.53	0.62	0.47	0.067
2005-7	0.46	0.59	0.69	0.53	0.065
			Panel B: Manufacturing Workers Only		
1980	0.33	0.51	0.55	0.44	0.046
1990	0.40	0.57	0.65	0.52	0.062
2000	0.45	0.60	0.70	0.55	0.060
2005-7	0.52	0.69	0.79	0.63	0.062

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 have standard errors of less than 0.01. 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 w^s/w^u	S/U	K/U	Efficiency Units w^s/w^u	S/U	K/U
			Panel A: All Workers			
1980	0.020	0.103	No	0.008	0.113	No
1990	0.029	0.107	Capital	0.014	0.121	Capital
2000	0.044	0.101	Data	0.024	0.120	Data
2005-7	0.051	0.089		0.029	0.110	
			Panel B: Manufacturing Workers			
1980	0.030	0.133	0.074	0.007	0.155	0.077
1990	0.045	0.136	0.093	0.020	0.160	0.104
2000	0.064	0.122	0.102	0.033	0.152	0.119
2005-7	0.072	0.128	0.090	0.042	0.158	0.107

Notes: Each entry is the coefficient in a regression of the log of the variable listed at top on log 1980 CBSA population in each indicated year. Regressions are weighted by 1980 CBSA population. Standard errors on estimated coefficients are below 0.003 in Columns 1 and 4, below 0.008 in Columns 2 and 5 and below .013 in columns 3 and 6.

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		
	$\Delta\ln(w^S/w^U)$	$\Delta\ln(S/U)$	$\Delta\ln(w^S/w^U)$	$\Delta\ln(S/U)$	$\Delta\ln(w^S/w^U)$	$\Delta\ln(S/U)$	$\Delta\ln(K/U)$	$\Delta\ln(w^S/w^U)$	$\Delta\ln(S/U)$	$\Delta\ln(K/U)$
log(1980 CBSA Pop)	0.011***	-0.005***	0.007***	-0.001	0.014***	-0.002	0.011***	0.011***	0.001	0.015***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
1990-2000	-0.051***	-0.285***	-0.054***	-0.281***	-0.003	-0.341***	-0.189***	-0.025***	-0.318***	-0.187***
Indicator	(0.002)	(0.005)	(0.001)	(0.005)	(0.003)	(0.008)	(0.011)	(0.003)	(0.008)	(0.011)
2000-2007	-0.046***	-0.444***	-0.027***	-0.457***	-0.001	-0.465***	-0.199***	-0.000	-0.461***	-0.210***
Indicator	(0.002)	(0.005)	(0.001)	(0.005)	(0.003)	(0.008)	(0.011)	(0.003)	(0.008)	(0.011)
Constant	0.074***	0.521***	0.061***	0.529***	0.035***	0.574***	0.385***	0.035***	0.574***	0.365***
	(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	0.371	0.771	0.398	0.786	0.113	0.592	0.172	0.146	0.578	0.189

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	0.37***	0.23***	0.074	0.15***	-0.031
Quantity)	(0.044)	-0.036	(0.048)	(0.055)	(0.095)
ln(CBSA Population)	-0.11***	-0.039***	-0.029**	-0.073***	-0.0049
	(0.012)	(0.015)	(0.013)	(0.016)	(0.022)
ln(Immigrants of	0.058***	-0.0067	-0.019**	0.033***	0.0082
indicated educ.$_{,t-1}$)	-0.0076	(0.012)	(0.0091)	(0.010)	(0.015)
Observations	2,752	2,765	2,707	2,376	2,426
R-Squared	0.33	0.20	0.63	0.50	0.23
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)	$\Delta \ln(w^u)$	$\Delta \ln(w^s/w^u)$	Δln(S/U) F.S.	Δln(K/Y)	$\Delta \ln(w^u)$	$\Delta \ln(w^s/w^u)$
Δln(Predicted S	0.21***				0.17***			
/Predicted U)	(0.045)				(0.043)			
Δln(Skilled Labor		0.63***	0.18**	-0.43***		0.77***	0.047	-0.30***
/Unskilled Labor)		(0.19)	(0.092)	(0.096)		(0.26)	(0.107)	(0.089)
ln(CBSA Population)	0.002	0.0072	-0.004**	0.014***	0.004	0.0058	0.000	0.012***
	(0.0065)	(0.0078)	(0.002)	(0.002)	(0.006)	(0.0093)	(0.002)	(0.0014)
ln(Skilled Imm. /	0.044***	-0.032***	-0.005	0.026***	0.046***	-0.042***	-0.002	0.018**
Unskilled Imm.)$_{t-1}$	(0.012)	(0.012)	(0.006)	(0.009)	(0.012)	(0.016)	(0.007)	(0.007)
Year = 2000	-0.31***	0.25***	0.15***	-0.16***	-0.30***	0.29***	0.11***	-0.13***
	(0.016)	(0.078)	(0.040)	(0.039)	(0.018)	(0.088)	(0.042)	(0.033)
Year = 2005-2007	-0.39***	0.15	0.10**	-0.21***	-0.41***	0.22*	0.035	-0.15***
	(0.030)	(0.10)	(0.049)	(0.052)	(0.029)	(0.13)	(0.054)	(0.046)
Constant	0.49***	-0.36***	-0.20***	0.29***	0.50***	-0.44***	-0.15**	0.21***
	(0.026)	(0.11)	(0.057)	(0.060)	(0.026)	(0.15)	(0.064)	(0.053)
Observations	2,751	2,047	2,751	2,751	2,751	2,047	2,751	2,751
First stage F		20.3	21.9	21.9		15.2	15.9	15.9

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

Parameter	Description	Sparse Model		Full Model	
		Counts	Eff Unit	Counts	Eff Unit
α_1	Coefficient on instrument in Equation (6)	0.21	0.17	0.25	0.22
		(0.04)	(0.03)	(0.04)	(0.04)
σ	$1/(1-\sigma)$=elast of sub btw K or U and S	0.65	0.78	0.79	0.88
		(0.06)	(0.05)	(0.05)	(0.04)
ρ	$1/(1-\rho)$=elast of sub btw K and U	0.95	0.97	0.95	0.97
		(0.01)	(0.01)	(0.01)	(0.01)
$d\mu_s$	Change in skilled labor biased agglom.			0.0063	0.0090
				(0.0028)	(0.0020)
$d\mu_k$	Change in capital biased agglom.			-0.0009	-0.0004
				(0.0007)	(0.0003)
$d\mu_u$	Change in unskilled labor biased agglom.			-0.0067	-0.0021
				(0.0031)	(0.0024)
$d\ln(A_s/A_u)$	Skill-biased technical change			0.28	0.34
				(0.11)	(0.14)

Notes: 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. Estimation equations are written out in the Appendix. Observations are weighted by 1980 CBSA population and standard errors are clustered by CBSA. Totals of 29 and 27 parameters are estimated in the two models respectively.

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

	Object	Equation and Term	Raw Counts	Efficiency Units
1	Actual		0.0147	0.0123
2	Predicted		0.0148	0.0123
3	Predicted Using K,S,U from Data		0.0127	0.0108
4	Agglomeration	Eqn 2, Term 1	0.0102	0.0098
5	S-U Shifts	Eqn 2, Term 2	0.0006	0.0000
6	K-U Shifts	Eqn 2, Term 3	0.0014	0.0008
7	Agglom. K-U Complem. Interaction	Eqn 2, Term 4	0.0005	0.0001
8	Skill-Biased Technical Change	Eqn 2, Term 5	0.0000	0.0000
9	Capital-Biased Technical Change	Eqn 2, Term 6	0.0000	0.0000
	Portion Due to Shifts in Agglomeration Economies		80%	91%
	Portion Due to Capital-Skill Complementarity		15%	9%

Notes: 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 except those in the first two rows are calculated using actual data for S, K, and U. Estimated parameter values from the full model (Table 6) are used throughout. All estimates with an absolute value greater than 0.001 are strongly statistically significant.

Table 8: Robustness Checks on Parameter Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Exclude 2005-7		Exclude 1980		Exclude 1980 & Pred. ω^c, ω^{cu}	
	Counts	Eff Unit	Counts	Eff Unit	Counts	Eff Unit
			Panel A: Sparse Model			
α_1	0.21	0.17	0.21	0.18	0.25	0.23
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
σ	0.66	0.78	0.18	0.47	0.40	0.62
	(0.06)	(0.04)	(0.16)	(0.12)	(0.12)	(0.10)
ρ	0.95	0.97	0.92	0.95	0.92	0.95
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
			Panel B: Full Model			
α_1	0.24	0.21	0.32	0.30	0.19	0.32
	(0.04)	(0.04)	(0.05)	(0.06)	(0.03)	(0.06)
σ	0.76	0.88	0.80	0.87	-0.01	0.91
	(0.06)	(0.04)	(0.05)	(0.04)	(0.20)	(0.03)
ρ	0.95	0.97	0.93	0.95	0.90	0.96
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
$d\mu_s$	0.008	0.012	-0.004	-0.002	0.021	0.000
	(0.004)	(0.003)	(0.003)	(0.002)	(0.005)	(0.002)
$d\mu_k$	-0.001	0.000	0.001	0.000	-0.014	0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.003)	(0.000)
$d\mu_u$	-0.010	-0.002	-0.010	-0.008	-0.022	-0.008
	(0.004)	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)
$d\ln(A_s/A_u)$	0.165	0.214	0.760	0.951	0.048	1.292
	(0.128)	(0.164)	(0.155)	(0.223)	(0.044)	(0.371)

Notes: Estimates and standard errors are reported for three alternative ways of setting up the data. Columns 1 and 2 show results when the final period is excluded from the data. Columns 3 and 4 show results when the first period is excluded from the data. Columns 5 and 6 additionally use input shares that have been predicted using 1987 and 1997 shares.

Table 9: Parameter Estimates If Skilled Labor is Nested With Capital

Parameter	Description	Sparse Model		Full Model	
		Counts	Eff Unit	Counts	Eff Unit
α_1	Coefficient on instrument in Equation (6)	0.28	0.24	0.25	0.21
		(0.05)	(0.05)	(0.04)	(0.04)
σ	$1/(1-\sigma)$=elast of sub btw K or S and U	0.84	0.90	0.84	0.87
		(0.03)	(0.02)	(0.02)	(0.02)
ρ	$1/(1-\rho)$=elast of sub btw K and S	0.18	0.39	-0.54	-0.67
		(0.29)	(0.25)	(0.44)	(0.66)
$d\mu_s$	Change in skilled labor biased agglom.			0.0168	0.0169
				(0.0024)	(0.0021)
$d\mu_k$	Change in capital biased agglom.			-0.0144	-0.0110
				(0.0029)	(0.0022)
$d\mu_u$	Change in unskilled labor biased agglom.			-0.0088	-0.0039
				(0.0030)	(0.0022)
$d\ln(A_s/A_u)$	Skill-biased technical change			0.23	0.19
				(0.06)	(0.07)

Notes: Estimates are analogous to those in Table 7, except that the productiontechnology nests skilled labor with capital. The structural equations used for estimation match those in the Appendix after skilled and unskilled labor are swapped.

Table 10: Relationships Between Components of $\Delta\ln(w^s/w^u)$ and ln(D), Alternative Nesting

	Object	Equation and Term	Raw Counts	Efficiency Units
1	Actual		0.0147	0.0123
2	Predicted		0.0154	0.0128
3	Predicted Using K,S,U from Data		0.0132	0.0125
4	Agglomeration	Eqn 2, Term 1	0.0214	0.0182
5	S-U Shifts	Eqn 2, Term 2	0.0005	0.0000
6	K-S Shifts	Eqn 2, Term 3	0.0097	0.0127
7	Agglom. K-S Complem. Interaction	Eqn 2, Term 4	-0.0184	-0.0183
8	Skill-Biased Technical Change	Eqn 2, Term 5	0.0000	0.0000
9	Capital-Biased Technical Change	Eqn 2, Term 6	-0.0005	-0.0005

Notes: Entries are analogous to those in Table 7, except that skilled labor is nested with capital in the production technology. Relevant parameter estimates are reported in Table 9.

Table A1: Summary Statistics, Manufacturing

1980 CBSA Population	0-50k 1980	0-50k 2005-7	50k-100k 1980	50k-100k 2005-7	100k-250k 1980	100k-250k 2005-7	250k-1m 1980	250k-1m 2005-7	>1m 1980	>1m 2005-7
					Panel A: Raw Units					
ln S	13.8	14.4	14.8	15.4	15.7	16.2	17.0	17.4	18.8	19.1
	(0.7)	(0.7)	(0.6)	(0.6)	(0.6)	(0.6)	(0.6)	(0.6)	(0.8)	(0.7)
ln U	15.0	14.8	16.1	15.7	16.7	16.3	17.8	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	12.9	13.4	13.8	14.2	15.1	15.4	16.8	17.0
	(1.3)	(1.0)	(1.1)	(0.8)	(0.9)	(0.8)	(0.8)	(0.6)	(0.8)	(0.8)
ln w^s	2.74	2.64	2.79	2.70	2.85	2.77	2.91	2.89	2.99	3.03
	(0.15)	(0.15)	(0.15)	(0.14)	(0.15)	(0.17)	(0.13)	(0.14)	(0.09)	(0.13)
ln w^u	2.53	2.36	2.57	2.38	2.62	2.44	2.63	2.44	2.70	2.49
	(0.17)	(0.13)	(0.18)	(0.13)	(0.18)	(0.13)	(0.17)	(0.10)	(0.13)	(0.11)
					Panel B: Efficiency Units					
ln S	13.7	14.4	14.7	15.4	15.7	16.2	17.0	17.5	18.8	19.2
	(0.7)	(0.7)	(0.6)	(0.6)	(0.6)	(0.6)	(0.7)	(0.6)	(0.8)	(0.7)
ln U	15.0	14.9	16.0	15.7	16.7	16.3	17.8	17.3	19.3	18.8
	(0.9)	(0.9)	(0.7)	(0.6)	(0.6)	(0.6)	(0.7)	(0.6)	(0.8)	(0.7)
ln w^s	2.87	2.68	2.91	2.72	2.95	2.78	2.98	2.87	3.03	2.98
	(0.12)	(0.12)	(0.11)	(0.12)	(0.12)	(0.13)	(0.10)	(0.12)	(0.08)	(0.11)
ln w^u	2.57	2.33	2.60	2.35	2.65	2.40	2.66	2.41	2.73	2.48
	(0.13)	(0.11)	(0.13)	(0.11)	(0.14)	(0.11)	(0.13)	(0.09)	(0.10)	(0.08)
Full Sample Size	380	380	234	234	167	167	103	103	38	38
Sample Size for K	335	252	211	188	137	129	77	73	25	25

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.

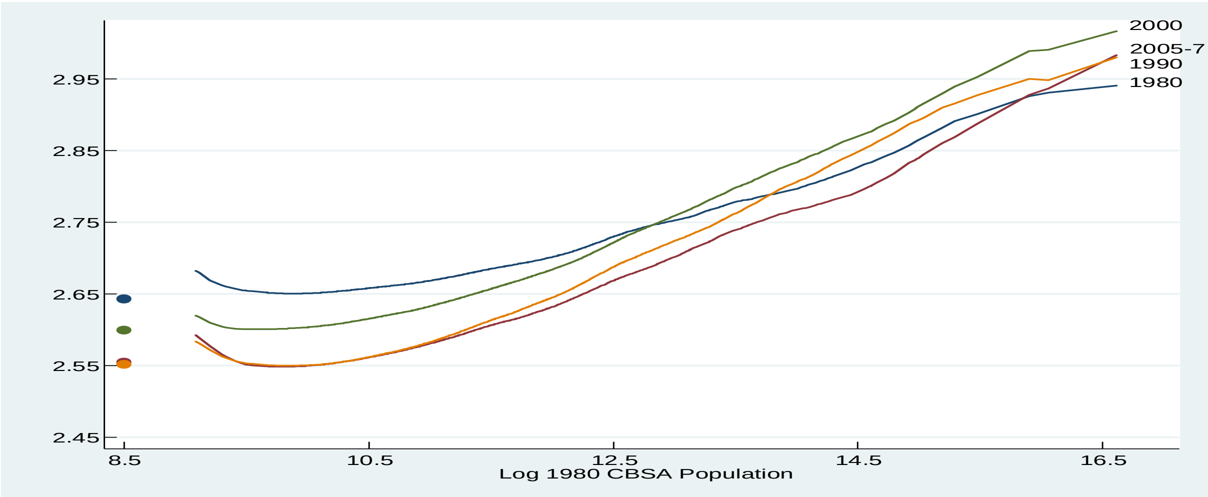
Table A2: Relationships Between Components of Δ ln(K/U) and ln(D)

	Panel A: Using Estimated Parameters			
1	Actual		0.003	0.007
2	Predicted Using K,S,U from Data		0.027	0.035
3	Agglomeration	Eqn 3, Term 1	-0.003	0.002
4	Agglom, S Bias	Eqn 3, Term 1, $d\mu_s \neq 0$	0.008	0.010
5	Agglom, K Bias	Eqn 3, Term 1, $d\mu_k \neq 0$	-0.012	-0.009
6	Agglom, U Bias	Eqn 3, Term 1, $d\mu_u \neq 0$	0.001	0.000
7	S-U Shifts	Eqn 3, Term 2 (1st pt)	0.017	0.020
8	Skill-Biased Technical Change	Eqn 3, Term 2 (2nd pt)	0.013	0.017
9	Prod. + Capital Price	Eqn 3, Term 3	0.000	-0.003
	Panel B: Using ρ Calibrated to Match Predicted with Actual			
Calibrated value of ρ			0.98	0.99
3	Agglomeration	Eqn 3, Term 1	-0.003	0.003
4	Agglom, S Bias	Eqn 3, Term 1, $d\mu_s \neq 0$	0.010	0.014
5	Agglom, K Bias	Eqn 3, Term 1, $d\mu_k \neq 0$	-0.016	-0.012
6	Agglom, U Bias	Eqn 3, Term 1, $d\mu_u \neq 0$	0.004	0.001
7	S-U Shifts	Eqn 3, Term 2 (1st pt)	0.007	0.012
8	Skill-Biased Technical Change	Eqn 3, Term 2 (2nd pt)	0.002	0.004
9	Prod. + Capital Price	Eqn 3, Term 3	-0.003	-0.012
	Panel C: Using $d\mu_K$ Calibrated to Match Predicted with Actual			
Calibrated value of $d\mu_K$			-0.0027	-0.0016
3	Agglomeration	Eqn 3, Term 1	-0.026	-0.027
4	Agglom, S Bias	Eqn 3, Term 1, $d\mu_s \neq 0$	0.008	0.010
5	Agglom, K Bias	Eqn 3, Term 1, $d\mu_k \neq 0$	-0.035	-0.038
6	Agglom, U Bias	Eqn 3, Term 1, $d\mu_u \neq 0$	-0.037	-0.023
7	S-U Shifts	Eqn 3, Term 2 (1st pt)	0.017	0.017
8	Skill-Biased Technical Change	Eqn 3, Term 2 (2nd pt)	0.013	0.017
9	Prod. + Capital Price	Eqn 3, Term 3	0.000	-0.003
	Panel D: Using σ Calibrated to Match Predicted with Actual			
Calibrated value of σ			0.11	0.50
3	Agglomeration	Eqn 3, Term 1	0.012	0.014
4	Agglom, S Bias	Eqn 3, Term 1, $d\mu_s \neq 0$	0.011	0.015
5	Agglom, K Bias	Eqn 3, Term 1, $d\mu_k \neq 0$	-0.004	-0.003
6	Agglom, U Bias	Eqn 3, Term 1, $d\mu_u \neq 0$	0.005	0.001
7	S-U Shifts	Eqn 3, Term 2 (1st pt)	-0.001	0.000
8	Skill-Biased Technical Change	Eqn 3, Term 2 (2nd pt)	-0.005	-0.005
9	Prod. + Capital Price	Eqn 3, Term 3	-0.001	-0.004

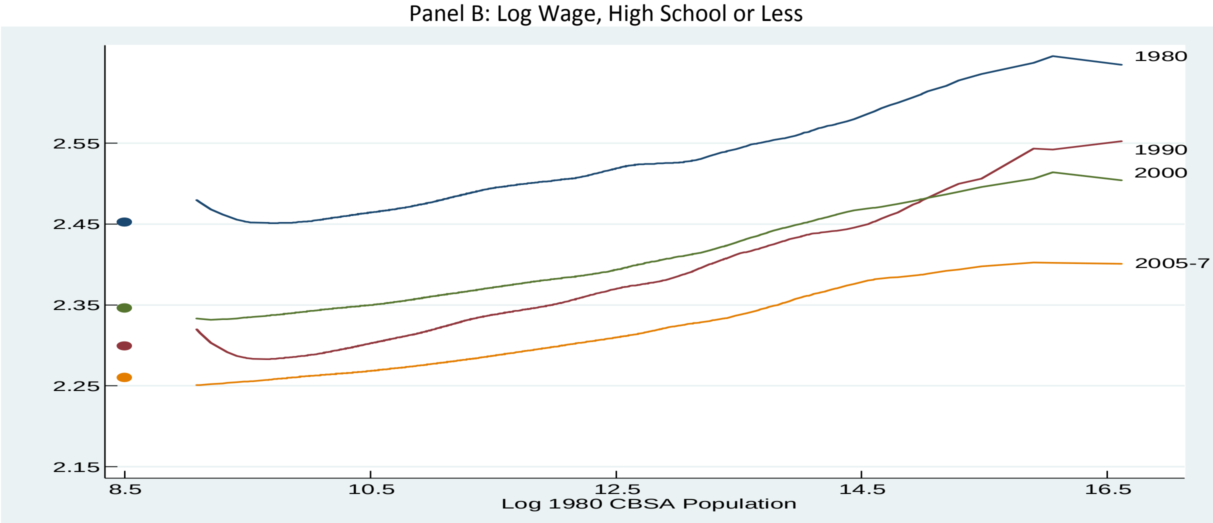
Notes: Each entry is the coefficient in the regression of the object listed at left on the log of 1980 CBSA population and year fixed effects weighted by 1980 CBSA population. All objects except those in the first row of Panel A are calculated using actual data for S, K and U. Estimated parameter values from Table 6 are used throughout except for the calibrated value indicated in each panel. Rows 5-7 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 statistically significant.

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

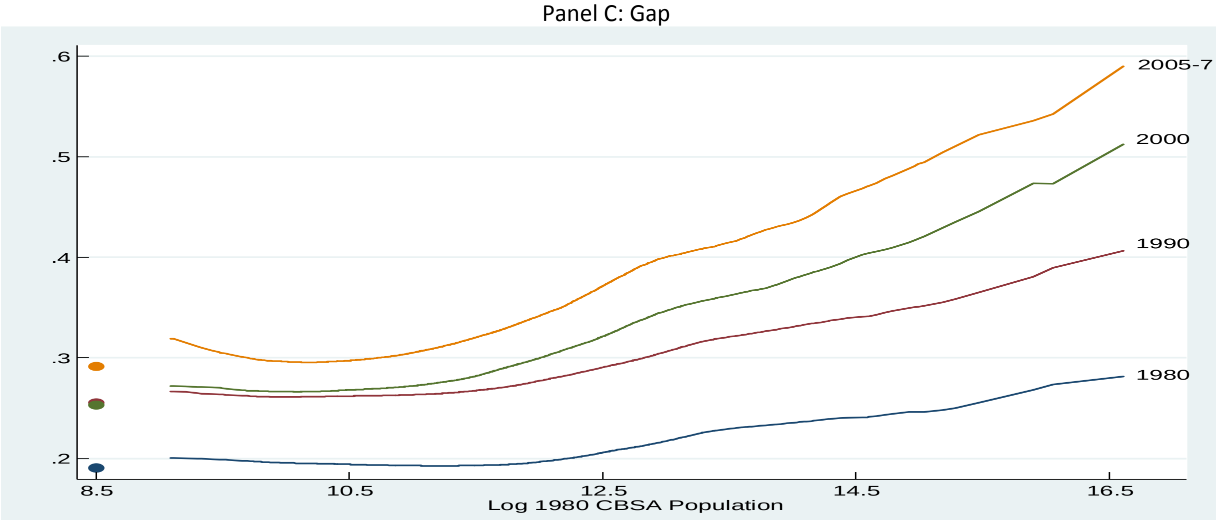
Panel A: Log Wage, Some College or More



Panel B: Log Wage, High School or Less



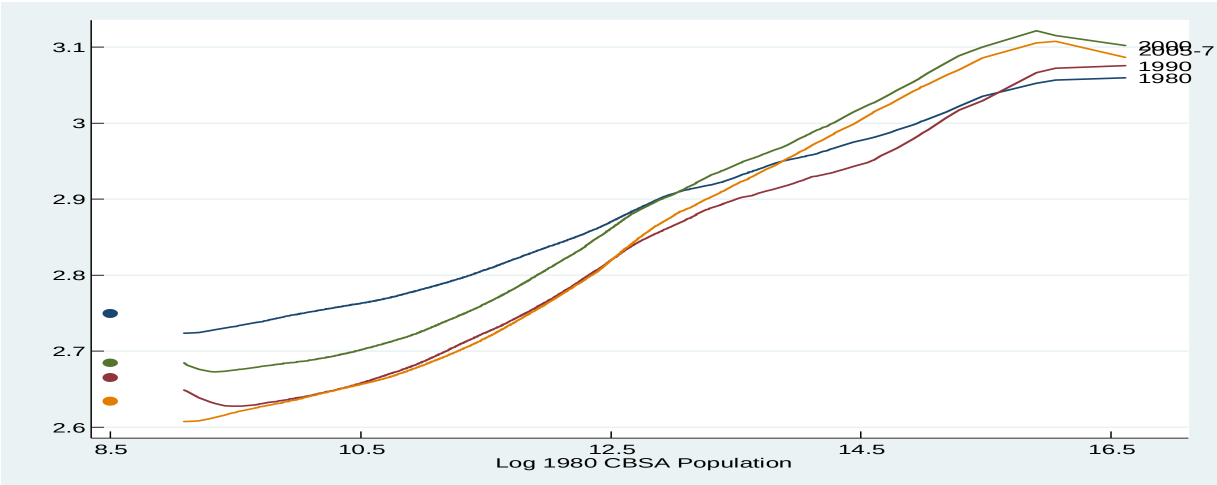
Panel C: Gap



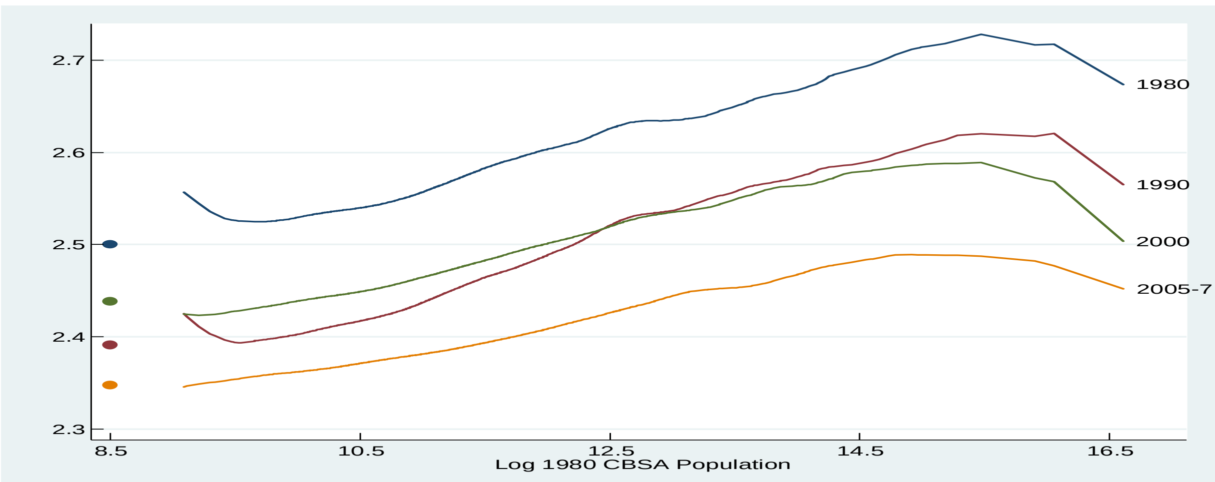
Notes: Each wage observation is weighted using census weights interacted with hours worked. Isolated dots at the left of the graphs are for rural areas.

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

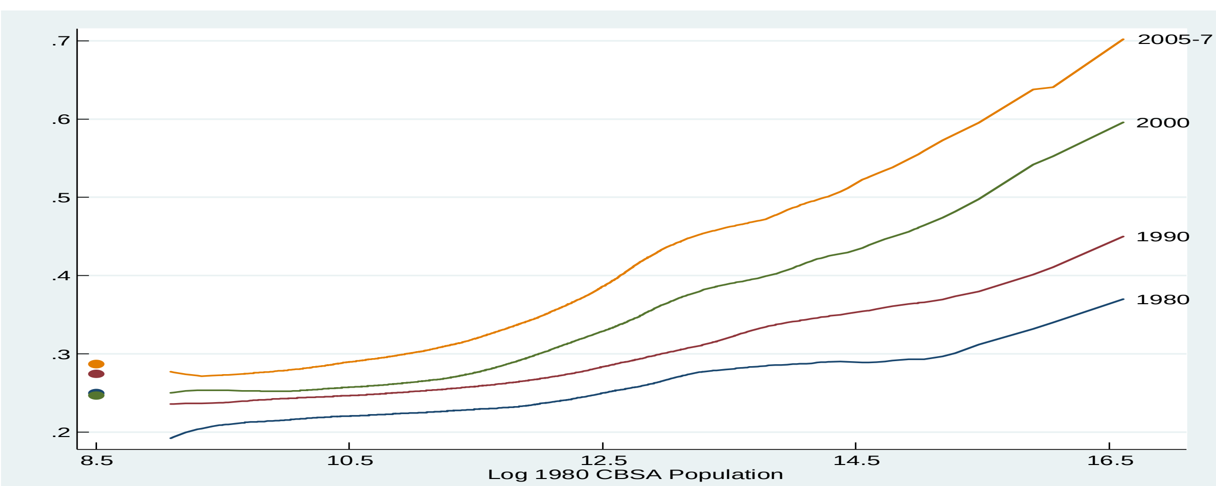
Panel A: Log Wage, Some College or More



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.