Migration, Education and Urban Divergence: Evidence from United States Patent Counts

Oudom Hean*
Nattanicha Chairassamee†
Mark D. Partridge‡

November 25, 2019

Abstract

The skill-biased technical change empirical literature has been silent regarding two key channels of labor supply, migration and demand for college education, through which innovation increases divergence in incomes, skills, and productivity between urban areas. Using patents as a measure of skill-biased technical change, we shed new light on technology-induced spatial inequalities. Instrumental variables estimations show that between 2005 and 2015, which corresponds to the decade of relatively low U.S. migration rates, the local demand for college education played a greater role in explaining skill divergence between U.S. urban counties. We find positive relationships between urban income divergence and patent counts in computer and data processing, telecommunications, and automation. These technologies are potentially skill-biased.

JEL Classification: D31, J24, J61, R23
Keywords: Migration, Demand for Education, Innovation, Urban Inequalities

*Corresponding author. The Ohio State University, Department of Agricultural, Environmental, and Development Economics. Email: hean.1@osu.edu. Website: www.oudomhean.com.

†AED Economics, The Ohio State University.

‡AED Economics, The Ohio State University; School of Economics, Jinan University; and Social Sciences Department, Gran Sasso Science Institute.

1 Introduction

Unequal urban growth in the United States has long been a focus of economists. Among the hypotheses examined, skill-biased technical change (and routine-biased technical change) has received considerable attention (Autor, Katz, & Kearney, 2008; Acemoglu & Autor, 2011; Moretti, 2011; Autor & Dorn, 2013; Goos, Manning, & Salomons, 2014; Giannone, 2017). Theoretically, two key channels of labor supply through which technology can change disparities in regional skills and incomes are technology-induced migration (Diamond, 2016; Moretti, 2013a) and technology-induced local demand for education (Goldin & Katz, 2010; Galor & Moav, 2000). Skill-biased technical change can lead to wage differential across cities and consequently can induce in-migrations of college workers to innovative cities or to cities with an industry composition that intensely uses university graduates (Berry & Glaeser, 2005; Betz, Partridge, & Fallah, 2015). Because technology raises local demand for highly educated, highly skilled workers; therefore, it induces a geographic sorting of workers by education levels or skills (i.e., the migratory channel). Such change can also raise the economic returns to education and increase the demand for college education (Galor & Moav, 2000; Jones & Yang, 2016). It encourages local residents to pursue higher education, which, in turn, produces higher local supply of skilled workers (i.e., the non-migratory channel). Surprisingly, to the best of our knowledge, there has been no empirical work that simultaneously considers internal migration, local demand for college education, and spatial inequalities in the context of skill-biased technical change.

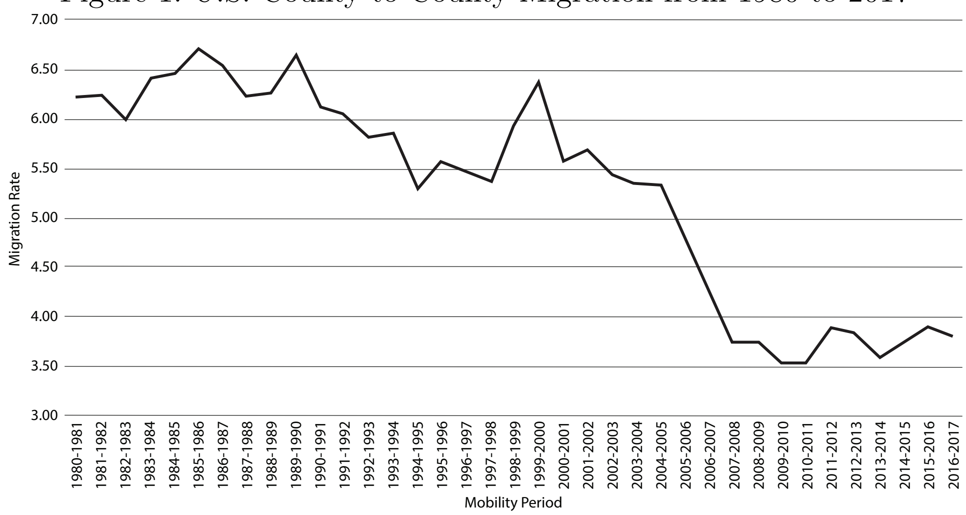
In this paper, we study the hypothesis of technology-induced urban divergence in skills and incomes/wages from an innovation-based perspective. Because of its impact on private returns and economic growth, patents are arguably an objective indicator of how technological change affects income inequality (Goldin & Katz, 2010; Aghion, Bergeaud, Blundell, & Griffith, 2017; Akcigit, Grigsby, & Nicholas, 2017; Kogan, Papanikolaou, Seru, & Stoffman, 2017). We use granted patents as a measure of skill-biased technical change and provide evidence that patent counts strongly capture the essence of technology-induced urban divergence. These findings are line with the recent finding that patents are a good indicator of technological progress and the knowledge advancement (Kogan, Papanikolaou, Seru, & Stoffman, 2017; Mann & Püttmann, 2018; Bottazzi & Peri, 2007; Jaffe, Trajtenberg, & Handerson, 1993; Grillches, 1990; Michaelsen, 2011).[1]

Consistent with previous works on technology-induced spatial inequalities (Diamond, 2016; Baum-Snow, Freedman, & Pavan, 2018; Moretti, 2013b), we measure skills by different

[1]We acknowledge that by focusing on patents, we could miss informal technological innovations that are not patented. We also recognize that patents do not differentiate between regular and radical technological innovations.

levels of education. We empirically identify the two technological channels that labor supply can satisfy labor demand for skills–i.e., (1) migration and (2) an increase in local demand for college education by currently remaining residents (which is simply referred to as the demand for college education or the non-migratory channel). Specifically, we assess how regional innovation affected inter-regional skill distributions from 2005 to 2015, during which the U.S. internal migration rate was low but divergence in urban skills and incomes did not diminish.

Figure 1: U.S. County-to-County Migration from 1980 to 2017



Note: The migration rate is calculated by dividing the total county-to-county migration by the total population. The data is obtained from the United States Census Bureau: https://www.census.gov/data/tables/time-series/demo/geographic-mobility/historic.html

Figure 1 illustrates a significant slowdown in internal U.S. migration rates since 2005 (Partridge, Rickman, Olfert, & Ali, 2012). Without taking the technology-induced local demand for college education into account, this pattern produces a mystery during the time period of this study because one would expect low migration rates to be most consistent with low wage/income differentials across U.S. regions (less incentive to migrate). This decline calls into question the degree to which the ongoing divergence in urban growth rates (Giannone, 2017) is affected by internal migration. On the other hand, from 2000 to 2017, total undergraduate enrollment increased by 27 percent (National Center For Education Statistics, 2019). Despite a slow pace, the country's college enrollment rates increased by 5 percent between 2000 and 2017 (National Center For Education Statistics, 2019).

Our paper offers three contributions to the literature on technology-induced spatial inequalities. First, to study urban divergence in skills and income/wages, we use patents as a measure of skill-biased technical change. Previous studies use various indicators to measure skill-biased technological change; however, to the best of our knowledge, none has yet employed patents to measure the effects of technological change on inequalities across U.S. urban counties. For instance, Beaudry, Doms, and Lewis (2010) use the adoption of personal computers to study the relationships between technological change, educational attainment, and the return to skill in U.S metropolitan areas. They conclude that the personal computer is a radical technological change. Analyzing the routine employment share at the commuting zone level, Autor and Dorn (2013) find that automation can both increase the demand for college-educated workers and shift less-educated workers toward service sectors. Diamond (2016) uses a structural spatial model and Batik labor demand shocks to study the sources and consequences of skill sorting across U.S. metropolitan regions. She finds that labor demand changes are the central cause of skill sorting. However, patents, as a measure of skill-biased technological change, have been underexploited in this literature.

Second, we identify two principal channels of labor supply, migration and the local demand for college education, through which innovation can affect the distributions of skills across U.S. urban counties. We use Bartik-style shift-share instruments to obtain the causal effects of innovation on the skill distributions. We find that from 2005 to 2015 technology-induced demand for college education or the non-migratory channel had a more significant impact on the skill distributions. Identifying these two channels as to how labor supply can respond to innovation is essential from both academic and policy perspectives. From an academic viewpoint, this study enhances our understanding of the links between innovation and divergence of urban skills and incomes/wages across cities. From a policy standpoint, we provide insights to bolster local economic growth by investing in new technologies, increasing college attainment, and attracting college workers.

Our third contribution responds to the perennial concern about interregional income/wage inequality. To this end, we conduct an analysis of how innovation affects incomes and wages across U.S. urban areas. Using patent data, we show strong positive relationships between income/wage divergence and specific technologies, including computer and data processing, telecommunications, and automation—relationships that have attracted the attention of scholars and policymakers. These findings cast some light on the black-box of skill-biased technical change—i.e., these technologies are potentially skill-biased technological changes.

The remainder of the paper is organized as follows. Section 2 discusses the advantages and disadvantages of using granted patents to measure skill-biased technical change. Section 3 outlines a conceptual framework that is a motivation of the empirical setup, while Section 4 discusses the data. Section 5 presents our empirical strategy of the links between innovation and the distributions of skills. Section 6 provides the estimated results, including those of the instrumental variables estimations. Section 7 gives direct evidence of how local demand for college education is influenced by innovation, and Section 8 performs a welfare analysis by examining per capita income and average wage. Section 9 probes several sensitivity analyses by examining some issues, such as immigration, commuting of high-skilled workers, knowledge spillovers, and serial correlation. Section 10 concludes.

2 Patents as a Measure of Skill-Biased Technological Change

Following the seminal works of Jaffe (1986); and Jaffe, Trajtenberg, and Handerson (1993), scholars have used patents as a proxy for knowledge, technology, and innovation. Acs, Anselin, and Varga (2002) also show that the patent count is a reasonable indicator of regional innovation. In fact, given the U.S. high patent intensity, patent counts can strongly capture innovations and technological progress. Additionally, Michaelsen (2011) uses patent applications as a measure of skill-biased technical change to explain the wage premium in the UK manufacturing industry from 1991 to 2006.

The use of granted patents in studies of how innovation affects spatial inequalities is appropriate for four key reasons. First, granted patents reflect the private economic return to firms and captures aggregate growth and productivity. Kogan, Papanikolaou, Seru, & Stoffman (2017) show that patent grants can account for firms' private economic value. Their measure of technological innovations constructed using patent grants can capture significant medium-run variations in aggregate economic growth and total-factor productivity. Second, there is a time lag between a discovery leading to a patent application and a granting of that patent. Therefore, the issue of endogeneity and simultaneity bias between high-educated workers and innovation can be lessened, as well as other measurement error issues. Third, to be consistent with the previous literature, we distinguish two different levels of skill/education: 1) high skilled or college workers (aged 25 to 64) who have a bachelor's or a higher degree; and 2) low skilled workers (aged 25 to 64), which we refer to as non-college workers, are defined as those who hold a high school diploma or have some college education. Although innovations are generally achieved by researchers who have more than a bachelor's degree, our use of the term "high-educated workers" in the U.S. refers mainly to those who have only a four-year college degree.[2] Therefore, it is less likely that there is

[2]Using survey data, Walsh and Nagaoka (2009) find that around 46 percent of U.S. inventors hold a doctorate degree. Unfortunately, they do not separate the education attainment data into Master's degree and Bachelor's degree, and so the percentage of inventors hold only a bachelor's degree is unknown.

a reverse causality between innovations and the percentage of "high-educated workers" (one of our variables of interest)—where we perform sensitivity analysis on whether this applies. Fourth, those with less than a college degree are less likely to patent.

In a separate analysis, we find strong evidence that patents capture the essence of skill-biased technical change. Specifically, we rank counties either by patent counts, percentage of high-educated workers, average income, average wage, or Gini index for household income. We find strongly positive relations between patent rank and other ranked variables. These results are consistent with the predictions of technology-induced spatial inequalities.[3]

Yet our estimates use crude patent counts as the lower bound estimates of technological effects because we do not control for patent quality. In other words, because all patents are given equal weight, we could be underestimating innovations that have high economic value. While citation counts are often used to indicate the quality of patents, they also have limitations.[4] Therefore, we use unweighted regional patent counts as the indicator of innovation and interpret the results as the lower bound estimates. However, our instrument variables (IVs) estimations should lessen this concern.

3 Conceptual Framework

The literature on skill-biased technical change is vast and includes the seminal studies of Katz and Murphy (1992); Bound and Johnson (1992); and Juhn, Murphy, and Pierce (1993). To illustrate how migration and local demand for college education affect the supply of skills and regional income/wage, we present a very stylized model inspired by Rosen (1979), Roback (1982), and the "canonical model" of skill-biased technical change.[5] To achieve theoretical tractability in empirical analysis, we assume the following. After dropping the time subscript to simplify the disposition of the model, the aggregate output produced by the representative firm in county i at time t is given by:

$$Y_i = [(X_{iH} \cdot H_i)^{\frac{\sigma-1}{\sigma}} + [(X_{iL} \cdot L_i)^{\frac{\sigma-1}{\sigma}})]^{\frac{\sigma}{1-\sigma}}, \quad \sigma > 1. \tag{1}$$

Here, Y is total output with the price normalized to be one; X_{ie} is skill-specific technology for workers of skill e. In this framework, X_{iH} is an innovation that has been patented in

[3]The results are available upon request.

[4]Citations may reflect the scientific value and may not capture the economic value of an innovation. Moreover, different industries tend to be characterized by different intensities of citation, and different regions have different types of industry composition. Using citation-weighted patents could bias our estimates in an unknown direction. Although Hall, Jaffe, & Trajtenberg (2001) suggest some methods to correct these issues, the methods cannot completely resolve the problems.

[5]For the canonical model of skill-biased technical change, see Acemoglu & Autor (2011).

county i, while X_{iL} is simply assumed to equal one. H_i and L_i are the numbers of high-skilled (college) and low-skilled workers (non-college) in county i, respectively.[6,7] In this specification, $Y_{ie} = X_{ie} \cdot e_i$ are goods produced by workers of skill e, for $e \in \{H, L\}$.

Wages are given by:

$$w_{iL} = \left[1 + X_{iH}^{\frac{\sigma-1}{\sigma}} \left(\frac{H_i}{L_i}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}}, \tag{2}$$

$$w_{iH} = X_{iH}^{\frac{\sigma-1}{\sigma}} \left[X_{iH}^{\frac{\sigma-1}{\sigma}} + \left(\frac{H_i}{L_i}\right)^{-\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}}. \tag{3}$$

Let ω_i denote the relative wages of high-educated and low-educated workers in region i, such that $\omega_i = \frac{w_{iH}}{w_{iL}}$. Additionally, Let Ω_{ik} be the relative wages of workers with skill e between regions i and k, such that $\Omega_{ike} = \frac{w_{ie}}{w_{ke}}$. It is readily shown that

$$\frac{\partial \omega_i}{\partial X_{iH}} > 0, \quad \text{and} \quad \frac{\partial \Omega_{ikH}}{\partial X_{iH}} > 0. \tag{4}$$

All else being equal, innovation raises the returns to education and increases income/wage inequality across regions. That is, innovations can increase the skill premium and induce spatial divergence in skill and income/wages.

In the U.S. labor market, migration and local demand for college education are two key determinants of labor supply. Workers migrate in response to the differential in wages (i.e., Ω_{ike}), amenity, and housing rents. An innovation in county i will increase differential wages and, therefore, induce high-skilled workers to in-migrate. Additionally, innovation can also raise the return to educational investment and, therefore, increase local demand for college education of low-skilled workers.

Let M_{it} be the total number of net migrants which is the total number out-migrants subtracted from the total number of in-migrants in region i. Migrants here may also include commuters from nearby regions. Also, denote $Enrollment_i(X_{iH})$ the total number of low-skilled workers who enroll in a college. $Enrollment_{i,t}(X_{iH})$ represents local demand for college

[6]That is, the production function simplifies the equilibrium solution, which involves migration and educational demand. Given data availability, the choice of this production function is relevant to our empirical analysis. Consistent with previous findings, Ciccone and Peri (2005), who use state-level data, find that the long-run elasticity of substitution between high- and low-educated workers is between 1 and 2. Their preferred value of the elasticity is 1.5.

[7]To simplify the exposition of the model, we assume away capital market in the theoretical model, in which perfect capital mobility implies a common interest rate–which is a reasonable assumption in the United States. In our sensitivity analysis, we use the estimated capital stocks at the state or county level to assess robustness of our baseline results. We find that our baseline results are very robust.

education in region i.[8] The migratory channel of innovation is simply $\frac{\partial e_i}{\partial M_i}\left(\frac{\partial M_i}{\partial X_{ie}}\right)$, and the non-migratory channel is given by $\frac{\partial e_i}{\partial Enrollment_i}$ or $\left(\frac{\partial e_i}{\partial X_{ie}}\right)|M_i$. That is, conditional on net migration, a change in the number of workers with skill e (due to innovation) should reflect the non-migratory channel which is the local supply satisfying the local demand for high-educated workers. Several hypotheses can be tested in the following empirical analyses. In theoretical Appendix A, we develop an extension of the current model to conceptualize these hypotheses and motivate a setup of our empirical analyses. In what follows, we summarize three key hypotheses.

Hypothesis 1 : Conditional on migration, innovation captures the local demand for college education of non-college workers. Therefore, $\left(\frac{\partial H_i}{\partial X_{iH}}\right)|M_i = -\left(\frac{\partial L_i}{\partial X_{iH}}\right)|M_i$. That is, conditional on migration, innovation increases the number of high-educated workers and decreases the number of low-educated workers. The magnitudes of these changes should be similar.

Hypothesis 2 : Net migration is an increasing function of innovation. Among high-skilled workers, innovation creates differential in wages between regions; therefore, innovation induces migration across urban counties. However, as in the canonical model of skill-biased technical change, technological change can also raise the wages of low-skilled workers. That is, low-skilled workers living in an innovative city can earn higher wages than those who live in a less innovative city (Moretti, 2013b).[9] By raising their wages, innovation could also attract some low-skilled workers. Yet given the sluggish mobility of low-skilled workers discussed in literature, migration of low-skilled workers should be less influenced by innovation in the short run.

Hypothesis 3 : The in-migration of high-skilled workers could lead to the out-migration of low-skilled workers, particularly over the long/medium term. An influx of highly mobile, highly skilled workers will lead to an increase in housing demand and, thus, housing prices, which could lead to an out-migration of low-skilled workers (Moretti, 2013a). However, given the low mobility of low-skilled workers, the out-migration rate of this group should be low, especially in the short term.

[8] The demand for education is a function of relative wages, which are functions of innovation, between high- and low-educated workers. However, because we assume $X_{iL} = 1$, educational demand is only the function of innovation, that constitutes a skill-biased technical change.

[9] In the canonical model of skill-biased technological change, high- and low-skilled workers are q-complements. That is, an increase in the productivity of high-skilled workers leads to an increase in the supply of high-skilled workers and a rise in the marginal productivity of low-skilled workers (Acemoglu & Autor, 2011). In fact, an increase in the productivity of high-skilled workers, which leads to an increase in the wage of high-skilled labor, can be accompanied by an increase demand for services and goods provided by low-skilled workers (Eeckhout, Pinheiro, & Schmidheiny, 2014).

4 Data

Ouranalysis focuses on urban counties that belong to metropolitan statistical areas (MSAs) delineated using 1999 metropolitan definitions of the Office of Management and Budget (OMB).[10] We are using the prior definitions of MSAs to reduce any endogeneity of MSA borders being expanded. The study is between 2005 and 2015, unless otherwise noted. The data are not available for all year or all MSA counties; therefore, the data set is an unbalanced panel. Domestic migration data is provided by the Internal Revenue Service (IRS).[11,12]

Because our main focus is on technology-induced spatial divergence between U.S. urban counties achieved through internal forces (i.e., internal migration and local demand for college), we ignore immigration.[13] This selection is consistent with many previous studies of skill-biased technical change.[14] In the section of sensitivity analysis, we find my baseline results robust to the inclusion of immigrants in the calculation of total net migration. To lessen concerns about commuting to work and outliers, we also exclude counties that host megacities.[15] However, the key findings of this paper are still maintained when we include these counties in the analysis.[16]

The number of utility patents is sourced from the U.S. Patent and Trademark Office (USPTO). In line with the innovation literature, the USPTO assigns a patent to each county on the basis of the residence of the first-named inventor when it is granted. We use this data set to identify the two technological channels that affect labor supply. The USPTO also provides patent counts within certain technology classes at the MSA level. We use this detailed

[10] The list of Metropolitan areas and components can be retrieved from https://www2.census.gov/programs-surveys/metro-micro/geographies/reference-files/1999/historical-delineation-files/99mfips.txt.

[11] The IRS computes migration data using year-to-year address changes reported on individual income tax returns. The annual deadline for tax returns is generally April 15. I therefore assign the initial year as the reference year for total migration. For example, the total migration data in a county for the period between year t and $(t+1)$ is the total migration in year t. My baseline results are similar when the end period serves as the reference year.

[12] IRS migration data does have its shortcomings. For example, it does not include those who do not file tax returns. It is also affected by features such as divorces, marriages, and when the tax-preparer uses their address on the return.

[13] The literature on the effects of immigration on the U.S. labor market is rich. See, for example, Borjas, Freeman, and Katz (1996) and Card (2001).

[14] Baum-Snow, Freedman, & Pavan (2018) use immigration shocks as the exogenous source of college and unskilled labor supplies. They find strong evidence supporting capital-skill complementarity and growth of skill bias of agglomeration economies in the presence of rapid skill-biased technical change.

[15] These megacities include New York (Kings County, New York County, Bronx County, Queens County, and Richmond County), Chicago (Cook County), Los Angeles (LA County), District of Columbia, Boston (Suffolk County), San Francisco (San Francisco County), Denver (Denver County), Houston (Harris county), Philadelphia (Philadelphia County), Dallas (Dallas County), and Seatle (King County). I also exclude Detroit (Wayne County) because it has suffered a major economic and demographic decline over several decades.

[16] The regression results of this exercise are available upon request. We still find that the non-migratory channel played a more significant role in explain skill divergence across urban areas between 2005 and 2015.

Table 1: Summary Statistics

	Observations	Mean	S.D.	Minimum	Maximum
Percentage of High-Educated Workers	5,189	0.14	0.049	0.032	0.36
Percentage of Low-Educated Workers	5,189	0.31	0.047	0.13	0.46
Net Migration Rate	6,782	0.003	0.011	-0.41	0.21
Patent Counts	6,916	56.36	154.35	0	3,293

Note: Patent count is the total number of granted patents. Net migration rate is defined as follows: $Net\ Migration = \frac{TotalInmigration - TotalOutmigration}{Population}$.

data set to study the relationship between urban income/wage divergence and specific types of technologies, including computer and data processing as well as telecommunications. To study the relationship between urban income/wage divergence and automation, we use the data at the commuting zone level provided by Mann & Püttmann (2018). Their definitions of commuting zones are based on Tolbert and Sizer (1996). Table 1 provides a summary statistics for the county level data on percentages of high- and low-educated workers, net migration rate, and patent counts.

The data on population by age group, educational levels, and enrollment in higher education are obtained from the American Community Survey (ACS). The U.S. Bureau of Economic Analysis (BEA) offers data on total population at the state and county levels. The BEA also provides county-level data on farm and non-farm employment, average income and wage, and median incomes of workers for each educational attainment level. MSA median housing rent is provided by the U.S. Department of Housing and Urban Development.[17]

5 Innovation and Skill Distributions Across Urban Counties: Empirical Model

The main variables of interest are the percentages of college and non-college workers in each county. To be consistent with related empirical works, we define as high-skilled or college workers who are between 25 and 64 and hold at least a bachelor's degree. Low-skilled or non-college workers are defined as those between 25 and 64 who have at least a high school diploma but not a bachelor's degree. The omitted categories are adult workers without a high school degree, people who are younger than 24 year olds, and the elderly.

Our analysis is at the county level. An issue with the county-level analysis is that we cannot account for commuting of high-skilled workers from nearby counties to work in urban

[17]The data is accessible at https://www.huduser.gov/portal/data sets/50per.html.

centers.[18] To circumvent this problem, some studies use either CZs or MSAs as units of analysis. However, analyses using CZs or MSAs cannot account for heterogeneity within each metropolitan region. Unlike MSA-level analysis, the county-level analysis is also consistent over time because the geographic boundary of each county is well defined over the period of this study. Therefore, our study is complementary to those studies using CZs or MSAs. We note that one of our main findings is that the migratory channel plays a lesser role in explaining skill distributions. If long-distance commuting reduces both in-migration of high-skilled workers from nearby regions and the total number of high-skilled workers in a highly innovative region, omitting commuting will cause our estimate of the migration effect to be larger. Consequentially, accounting for commuting should further lessen the effect of technological migration and strengthen this key finding. Nonetheless, one way to interpret the estimated coefficient of the migration variable is that it captures the effects of cross-county migration and commuting to work. In Section 9, we further explore these effects.

There might also be a concern about spatial autocorrelation in the patent variable, particularly knowledge spillovers. In Section 9, to capture these spillover effects from the surrounding regions, we control for an exponential distance-decay function of patent counts in other counties. The results of this sensitivity are virtually indistinguishable to those of our baseline.

Through the lens of the theoretical framework, we follow a two-step approach to identify the migratory and educational demand channels of labor supply. Together these two steps provide a recursive model, and the estimations are as follows:

1^{st} step: We obtain the effect of innovation on migration by estimating the following regression,

$$M_{ist} = \gamma Patent_{ist} + C_{ist} + FE_{st} + \epsilon_{ist}, \tag{5}$$

where M_{ist} is the net migration rate in county i, in state s, and at time t; $M_{ist} = \frac{Net\ Migration_{i,s,t}}{Total\ Population_{i,s,t}}$.[19] $Patent_{ist}$ is the inverse hyperbolic sine function (*IHS*) of patent counts in county i.[20] Concerning with zero values, we use *IHS* transformation (although less than

[18]Compared with less educated workers, high-educated workers are less sensitive to commuting time (Hårsman & Quigley, 1998). Therefore, the labor market for high-educated workers is relatively bigger than that of low-educated workers.

[19]We weigh the variables of interest by the total population at the county level. This weight has three nice properties. First, it allows us to control for the population in this estimation; therefore, it controls both for scaling effects and agglomeration effect. Second, it eliminates the problem of outliers in that some counties have a very high/low number of high- and low-educated workers. Third, it can avoid the issues of negative numbers and zeros which are the problems with the logarithmic transformations.

[20]The total patents counts is preferred to per capita patents because changes in the total patent counts better reflect the speed of innovation or technological progress. Yet we also conduct a sensitivity by replacing the *IHS* of total patent counts with *IHS* of per capita patents. These sensitivity results are robust to this modification.

2 percent of total observations have zero values). Following the migration literature, the control variables ($C's$) include personal per-capita income; median incomes of workers for education groups (less than a high school degree, high school degree, some college, bachelor's degree, and postgraduate); and MSA-level (not county-level) housing rents.[21] We use the median incomes as proxies for wages and to control for economic conditions for workers with different skill levels. The amenity variable is the USDA Economic Research Service natural amenity scale which is a function of January temperature, July temperature, July humidity, topography, and the percentage of the county area covered by water.

To account for a labor demand shock and the industrial composition effect, we use the natural log of industry mix employment growth (Bartik, 1991; Blanchard & Katz, 1992). We use the log transformation of this variable here because this specification fits the data better.[22] The industry mix variable comes from shift-share analysis, and its construction is as follows: $Industry\ Mix\ Employment\ Growth_{it} = \sum_j Share_{ji} \Delta Emp_{nationaljt}$. Here, $Share_{ji}$ is the employment share of industry j, in county i, and in 2005. $\Delta Emp_{national,j,t}$ is the national growth rate of industry j in year t relative to year 2005. FE_{st} contains the state and year effects. To capture any long-term trend within a state, we will also estimate this model using the state-specific time trends and state effects.

2nd step: We estimate the marginal effects of migration and college education on the skill distributions with the following model:

$$Education_{ist} = \theta_M M_{ist} + \theta_E Patent_{ist} + \tilde{C}_{ist} + FE_{st} + e_{ist} \tag{6}$$

where $Education_{ist}$ is the percentage of college or non-college population in county i. As previously defined, M_{ist} and $Patent_{ist}$ are the net migration rate and inverse hyperbolic sine function of the total number of granted patents, respectively. Using equations (5) and (6), the marginal effects of technological migration on either low- or high-skill distribution equals $\left(\frac{\partial Education}{\partial Patent}\right)_{migration} = \left(\frac{\partial Education}{\partial M}\right)\left(\frac{\partial M}{\partial Patent}\right) = \theta_M \cdot \gamma$. Conditional on the net migration rate, the patent variable captures the non-migratory channel of innovation—that is, the demand for college education. Therefore, the marginal effect of technology-induced demand for college education is $\left(\frac{\partial Collge}{\partial Patent}\right)_{education} = \theta_E^H \approx -\theta_E^L = -\left(\frac{\partial Noncollege}{\partial Patent}\right)_{education}$. In Section 7, we show additional evidence that, conditional on migration, innovation induces local college enrollment.

The model in the second step controls for regional labor market conditions and industrial

[21] We use a two bedroom apartment as a baseline for the estimations. Using different housing rents with a different number of rooms does not change the result of the baseline estimations.

[22] Davidson and MacKinnon (1981)'s test for functional misspecification strongly supports the use of the log industry mix employment growth in the regression (5), and it rejects the use of raw values (levels) of this variable.

composition because we include the industry mix employment growth and the percentages of the population working in the farm and non-farm sectors (the omitted group is the government employees). To further control for the agglomeration effect, a sensitivity analysis includes population density; the results resemble those of the baseline estimations. Finally, to control for labor-market and technological trends (and possibly agglomeration), we include the interaction term between population growth and patent growth.[23] Our baseline estimation includes both year and state effects. To check the sensitivity of our baseline results, we also replace the year effects with state-specific trends. In our study, employing the state and not county fixed effects in regressions (5) and (6) is appropriate because (1) this study focuses on spatial inequality between and not within urban counties and (2) most variations of migration are between county variations. Using the state fixed effects allow me to utilize between and within county (but within state) variations.

Specifically, because the net migration rate contains both college and non-college workers, in model (6) the estimated coefficients $\theta'_M s$ capture the effects of various educational groups on the skill distributions. As demonstrated in previous empirical work, among college workers, θ_M^H is likely to be positive because it is likely to capture the net migration of college migrants included in the percentage of college workers in a county. On the other hand, in the case of the distribution of non-college workers, θ_M^L is likely to be weakly negative because these workers have a sluggish mobility and could be crowded out of a county by an influx of high-skilled migrants as housing values rise and industrial composition changes (Moretti, 2013a). Therefore, due to this low mobility, internal migration is likely to be a weak predictor of the short-run regional stock of non-college workers.

The IRS migration data, unfortunately, does not allow me to separate the effects of high-educated migrants from those of low-educated migrants. The American Community Survey (ACS) and Current Population Survey offer more detailed data about educational attainment of migrants. However, given that ACS and CPS data are survey data, they are subject to sampling errors, particularly in less-populated MSA counties. Because our purpose is to compare magnitudes of different channels of technology, these migration measurement errors are problematic. In contrast, Foster, Ellis, & Fiorio (2018) have found the IRS administrative data to be relatively accurate for ongoing taxpayers.

---

[23]Population growth is calculated as $\Delta Population_{i,s,t} = \frac{\sum_{t-4}^{t} Population_{i,s,t} - Population_{i,s,t-1}}{5}$, and patent growth is defined by $\Delta Patent_{i,s,t} = \frac{\sum_{t-4}^{t} Patent_{i,s,t} - Patent_{i,s,t-1}}{5}$.

6 Innovation and Skill Distributions Across Urban Counties: Estimations

6.1 Baseline Estimations

The estimated results of specifications (5) and (6) are tabulated in Table 2. The regressions in Column (1) are the benchmark that control for year and state effects. In Column (2), the regressions control for the state effects and state-specific trends to account for possible long-term economic trends. While controlling for the year and state effects, the regressions in Column (3) use one-period lags of the net-migration rate and *IHS(total patent counts)* variables. In Column (4), we repeat the estimations in Column (1) but replace *IHS(total patent counts)* with the *IHS*($\frac{total\ patent\ counts}{population}$).

The results are highly robust across specifications. Regarding the effects of migration on skill distributions, the estimations yield contrasting results for college and non-college workers. Net migration has a positive effect on the percentage of college workers but an insignificant effect on the percentage of non-college workers. As noted, friction in the low-educated workers' labor market can be quite high.

The marginal effect of the patent variable on the distribution of college workers is around 0.5, while the marginal effect of patents on migration is 0.001. Consequently, the effect of technological migration on the distribution of college workers is around 0.0005 (i.e., $\widetilde{\theta}_M^H \cdot \widehat{\gamma} \approx 0.5 \times 0.001$). In other words, on average, doubling the patent counts is associated with an increase of a 0.05 percent of college workers through migration.[24] Yet for non-college workers, technology-induced migration does not appear to play a significant role in the short run.

Conditional on migration, Table 2 also shows that the effects of innovation on the college and non-college population percentages are similar in magnitude but opposite in sign. Therefore, the estimations are consistent with our theoretical framework. The results indicate that an increase in the local demand for education occurs in response to local progress in innovation. The estimated effect of a technology-induced demand for education is between 0.01 and 0.02. Therefore, doubling the patent counts is associated with, on average, an increase (decrease) of a 1 to 2 percent of college workers (non-college workers) through demand for college education.

Using the delta method and Wald test, we also test the null hypothesis that the migratory and non-migratory effects are equal in magnitude (i.e., $\theta_M \cdot \gamma = \theta_E$) for both education groups. We can reject these null hypotheses at the conventional levels.[25] That is, the non-migratory

[24] The sample annual average of patent counts is 56.

[25] We obtain the covariance matrix by jointly estimating (5) and (6) using Zellner's seemingly unrelated regressions.

Table 2: Technological migration, Education and Skill Distributions

Model	(1)	(2)	(3)	(4)
A. Dependent Variable: Net Migration Rate				
IHS(Patent)	0.001***	0.001***	0.001***	0.0007***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Observations	4,900	4,900	4,467	4,900
B. Dependent Variable: Percentage of College Workers				
Net Migration Rate	0.53***	0.55***	0.82***	0.50***
	(0.188)	(0.191)	(0.179)	(0.181)
IHS(Patent)	0.02***	0.02***	0.02***	0.02***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	5,146	5,146	4,704	5,146
C. Dependent Variable: Percentage of Non-college Workers				
Net Migration Rate	0.04	0.03	-0.08	0.04
	(0.047)	(0.046)	(0.049)	(0.049)
IHS(Patent)	-0.01***	-0.01***	-0.01***	-0.01***
	(0.0005)	(0.0005)	(0.0005)	(0.001)
Observations	5,146	5,146	4,704	5,146
Year Effects	Yes	-	Yes	Yes
State Effects	Yes	Yes	Yes	Yes
State-Specific Trends	-	Yes	-	-

Note: Robust standard errors are in parenthesis. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. When standard errors are clustered at the state level, all the results, except that of Panel A in Column (1), remain statistically significant at the conventional levels. In Column (3), the variables of interest—the net migration rate and *IHS* (Inverse Hyperbolic Sine function) of *total patent counts*–are lagged one year. When lagging these variables of interest, we lose some observations per county. In Column (4), the estimations are the same with those in Column (1), but the *total patent counts* is replaced by *per capita patent counts*.

channel significantly affects the distributions of skills across urban regions during the period under study.

6.2 Instrumental Variables Estimations

Although the results of the estimations in Section 5 are very robust, there are two concerns regarding the endogeneity of the patent and migration variables. First, there is a concern about the simultaneity bias of the patent variable; however, the time lags in the patenting process minimizes this concern for the number of granted patents. In other words, there is a time lag between new inventions and the resulting local demand for high-skilled workers. A second and greater concern is the migration variable, which is a regressor in equation

(6). We attempt to solve potential endogeneity of the migration and patent variables by the two-stage least squares (2SLS) estimations. As a robustness check to our benchmark IVs estimations (i.e., 2SLS with the state fixed effects), we replace the state effects with county effects. By replacing these state effects with the finer geographical controls, we are restricted to only within-county variations for estimations. The similarities between the findings of the *IVs with the state effects*, *IVs with the county effects*, and the baseline estimations strengthen our conclusions.[26]

We use variations in labor and industry demand factors to construct Bartik-style shift-share instruments for the labor supply factors (i.e., migration and demand for college education). We instrument the net migration rate variable with the following:

$$\hat{M}_{it} = \sum_{j} \widehat{\phi_j} \times (Emp_{i,j,t-10})(\Delta Emp_{national,j,t,t-10})\left(\Delta M_{national,t,t-10}\right), \tag{7}$$

where $\hat{M}_{it}$ is an instrument for the net migration rate variable. $Emp_{i,j,t-10}$ is an employment share of the one-digit industry j in county i during the previous decade, and $\Delta Emp_{national,s,t}$ is the national growth rate of employment in industry j during year t relative to the previous decade (*t-10*). $\Delta M_{national,t,t-10}$ is the national growth rate of net migration during year t relative to the previous decade. To ensure the exogeneity (and to alleviate the concern about spatial autocorrelation) of our instrument, when calculating the national growth rates, we exclude the state that county i belongs. Intuitively, our instrument is the predicated number of migrants who work in industry j, county i, and year t. This Bartik instrument is similar to the RegressM instrument developed by Détang-Dessendre, Partridge, and Piguet (2016). Because RegressM allows different multipliers (i.e., coefficients $\phi_j's$) across industries, the RegressM approach can produce a stronger instrument for migration, particularly when the national trend is weak. Our instrument is suitable when the national trend of migration has substantially declined and weakened during the last decades. The first-stage F-statistic of the migration instrument is 23 suggesting a strong instrument.[27]

Our Bartik instrument for the patent variable is constructed as follows:

$$\widehat{PATENT}_{it} = \widehat{\Phi} \sum_{J} \left(\frac{PATENT_{i,t-5}}{PATENT_{s,t-5}}\right) \left(\frac{PATENT_{j,s,t-5}}{PATENT_{s,t-5}}\right) \Delta PATENT_{national,j,t,t-5}. \tag{8}$$

[26]When the number of research institutions is (roughly) constant over the period of this study, using the state and county fixed effects can alleviate the concern from not controlling for the impact of these institutions on local labor markets and innovation.

[27]The first stage of the 2SLS estimations includes the migration instrument defined in (7), the patent instrument discussed in (8), industry mix employment growth, the interaction term between population growth and patent growth, the fractions of nonfarm and farm labor (excluding government workers), the state effects, and year effects.

We use five-year lags to create this instrument.[28] The first expression in parenthesis is the patent share of county i in state s. The second expression is the patent share of subsector j in the manufacturing industry in state s. Together, these two expressions create an estimate of the patent share of subsector j in county i based on state patterns. This construction lessens any influence from a county's socio-economic factors on that county's innovative activities. The last expression in (8) is the national growth rate of patent counts in the subsector j. To ensure the exogeneity of this instrument, when calculating national growth in patent counts, we also exclude the state to which county i belongs. Based on the national patent growth rate, our instrument is the predicted number of patent counts in subsector j of county i at time t. The ideas behind the construction of this instrument are geographic localization of knowledge spillovers (see, Jaffe, Trajtenberg, & Handerson (1993)) and local industry composition. The resulting first-stage F-statistic is 37, which suggests a strong instrument.[29]

Table 3 shows the 2SLS estimates for skill regressions. Column (1) reports the baseline results, while Column (2) shows the results of *IVs estimations with the state effects.* Finally, to check the validity of our study design, Column (3) presents the results of *IVs estimations with the county fixed effects.* It is important to reiterate that migration predominantly happens within a state; using county fixed effects significantly soak up between-country variations. With this caveat in mind, the results are reported in Table 3.[30]

All estimations control for industry mix employment growth, the interaction term between population growth and patent growth, and the fractions of nonfarm and farm labor. A recent concern with Bartik instruments is an issue of serial correlation (Jaeger, Ruist, & Sthler, 2018). However, controlling for industry mix employment growth between 2005 and 2015 along with other controls lessens the possibility of bias created when patent counts are

[28]We do not have data on the national patent counts in each industry from 2013 to 2015. We impute these data with the followings:$\Delta PATENT_{national,j,t,t-5} = \left(\frac{PATENT_{j,t-5}}{PATENT_{national,t-5}}\right) \Delta PATENT_{national,t,t-5}$

[29]In the first stage of the $2SLS$ estimations with the state effects of the skill regressions (i.e., equation (6)), the Cragg and Donald (1993) minimum eigenvalue statistic is 22. The Stock and Yogo (2005) 5% critical value of "2SLS relative bias" is 18.76. The Stock and Yogo (2005) 10% and 15% critical values of "2SLS Size of nominal 5% Wald test" are 29.32 and 16.16, respectively. Finally, The Stock and Yogo (2005) 10% critical value of "LIML Size of nominal 5% Wald test" is 3.64. Taking together, the instruments for the migration and patent variables in 2SLS estimations with the state effects of the skill regressions (i.e., equation (6)) are strong.

[30]When we replace the state fixed effects with county effects, the F values in the first stage of the 2SLS for the migration and patent variables fall to 9.17 and 1.19, respectively. Given that the results of $2SLS$ with the *county* fixed effects are comparable to those of 2SLS with the $state effects$, there should be a less concern about the weak instruments in the estimations *with* the county fixed effects. Additionally, we also estimate these IVs regressions with the limited information maximum likelihood (LIML) estimator. We find that the results of $LIML$ with $the county effects$ are virtually indistinguishable to those of 2SLS with the county effects. The comparability between the results of $LIML$ and $2SLS$ further lessen the concern about weak instruments (Goldsmith-Pinkham, Sorkin, & Swift, 2018; Angrist & Pischke, 2008).

Table 3: Instrumental Variables Estimations in the Skill Regressions

Model	Baseline	2SLS	2SLS
	(1)	(2)	(3)
A. Dependent Variable: Percentage of College Workers			
Net Migration Rate	0.53***	1.55***	0.23*
	(0.188)	(0.301)	(0.13)
IHS(Patent)	0.02***	0.02***	0.01
	(0.001)	(0.002)	(0.006)
Observations	5,146	5,124	5,124
B. Dependent Variable: Percentage of Non-college Workers			
Net Migration Rate	0.04	-0.12	0.02
	(0.047)	(0.200)	(0.183)
IHS(Patent)	-0.01***	-0.003	-0.02**
	(0.0005)	(0.002)	(0.010)
Observations	5,146	5,124	5,124
Year Effects	Yes	Yes	Yes
State Effects	Yes	Yes	-
County Effects	-	-	Yes

Note: Robust standard errors are in parenthesis. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. All estimations control for the industry mix employment growth. Column (1) is the baseline estimation; (2) is the IVs estimations with the state effects; and (3) reports the IVs estimations with the county effects. The instruments of net migration rate and *IHS* of patents are defined in equations (7) and (8), respectively. When constructing these instruments, there are some missing observations. These missing observations cause the number of observations in IVs estimations to be different from the number of observations in the benchmark estimations.

correlated with local economic growth. Yet in our sensitivity analysis, we further tackle the issue of serial correlation.

The results shown in Table 3 support the baseline. First, the mobility of non-college workers is relatively sluggish. Second, conditional on migration, the marginal effects of patents on the shares of college and non-college workers are similar in magnitude (although in the non-college workers specification, the estimated effect of patent in Column (2) is relatively small and insignificant); however, their signs are opposite. This second finding suggests that the demand for college education is induced by innovation.[31]

[31] As a robustness check to the results estimated by $2SLS$ with the state effects, we also use $LIML$ with the state effects because LIML is less biased, particularly when sample size is small (Goldsmith-Pinkham, Sorkin, & Swift, 2018; Angrist & Pischke, 2008). The results of LIMIL, which are similar to those of 2SLS and the baseline, suggest the validity of our instruments. For the migration regression (i.e., equation (5)), LIML estimates the coefficient of the patent variable to be 0.001. For the high-skilled regression (i.e., equation (5)),

To complete our analysis, we further examine the patent variable in the migration regression (i.e., equation (5)). We instrument this variable with the Bartik patent instrument shown in equation (8). In contrast to skill regressions, we only use state fixed effects and no county fixed effects because most variations of migration come from between counties and within individual states. To avoid any contamination from potential endogenous variables, we omit incomes and housing rent from the analysis and only control for the natural log of industry mix job growth.[32] The results, tabulated in Table 4, yields comparable results to those of the baseline.

Table 4: Instrumental Variable Estimations in the Migration Regression

Model	Baseline	2SLS
	(1)	(2)
Dependent Variable: Net Migration Rate		
IHS(Patent)	0.001***	0.004***
	(0.0002)	(0.001)
First-stage F-statistic	-	20
Observations	4,831	6,782
Year Effects	Yes	Yes
State Effects	Yes	Yes

Note: Robust standard errors are in parenthesis. *, ** and *** indicate significance at 10%, 5% and 1% levels, respectively. In Column (2), the instrumental estimations solely control for the natural log of industry mix employment growth. The instruments for the *IHS(Patent)* is defined in (8). There are missing observations in the incomes and housing rent variables. Therefore, by dropping those variables, the number of observations in the IVs estimation is higher than that in the baseline estimations.

Heretofore, our analysis has focused on the short run, an approach that is useful when the mobility of high-skilled workers is relatively high.[33] In the short run, conditional on migration, patents are likely to capture the intensive margin of the demand for education (the decision of whether to complete college education after enrollment). Yet the effects of innovation on low-educated workers might not be very evident in the short run.

In a separate analysis, we also analyze the impact of innovation on skills distributions over the medium term. The results echo our findings regarding high mobility of high-educated

LIMIL yields 0.80 and 0.02 for the estimated coefficients of the migration and patent variables, respectively. Finally, for the low-skilled regression (i.e., equation (5)), the coefficients of the migration and patent variables are -0.12 and -0.01, respectively.

[32] Whether one uses the industry mix employment growth or the log industry mix employment growth, IVs estimations yield similar results. The results are also comparable to the baseline. Therefore, using IVs decreases the concern about functional misspecification.

[33] Additionally, granted patents only capture the medium/long run effects of technical change on labor market whenever workers react to innovation on or prior to the date of patent applications.

workers and low mobility of low-educated workers. Because of high mobility, over the medium term, we find the past migration has a low predictive power when it comes to a change in the percentage of college workers. On the other hand, because of low mobility of non-college workers, the lagged net migration rate is a good predictor of a change in the share of non-college workers. Our estimations suggest that the out-migration of non-college workers is precipitated by high housing costs in the medium run. Finally, our results show that the non-migratory channel is still important.[34]

7 Evidence of Technology-Induced Demand for College Education

To shed light on how educational demand increases in response to innovations, we present suggestive evidence. Specifically, we study the relationship between enrollment in college, migration, and innovation. Conditional on net migration, the effects of the patent counts variable should reflect the effect of innovation on the local demand for college education. We estimate the aforementioned effects as follows:

$$E_{ist} = \alpha_1 M_{ist-l} + \alpha_2 Patent_{ist-l} + \tilde{C}_{ist-l} + FE_{st} + e_{ist}, \tag{9}$$

where E_{ist} is either the percentage of the population aged 18-24 that is enrolled in higher education or the percentage of the total population that is enrolled in higher education. As previously defined, M_{ist} and $Patent_{ist}$ are the net migration rate and *IHS(total patent counts)*, respectively. $\tilde{C}'s$ are the control variables, including the population growth-patent growth interaction term, industry mix employment growth, and percentages of population working in the farm and non-farm sectors.

It is important to note that individuals enrolling in college, especially college students aged 18-24, are unlikely to be innovators. Consequentially, in the regressions with the percentage of the population aged 18-24 enrolling in college as the dependent variable, the reverse causality in the patent variable should be of little concern. To lessen the concern about endogeneity, we also estimate the regression specification (9) with the independent variables lagged either four or six years.[35] These lagged specifications also enable me to examine the

[34] The analysis is available upon request.

[35] Because the structures of the error terms in (9) are different from those of the migration and skill regressions (specifications (5) and (6), respectively), it is not appropriate to apply my IVs in (9). When using IVs, the results are different between 2SLS and LIML; however, LIML is less biased than 2SLS when sample size is small (Angrist & Pischke, 2008; Goldsmith-Pinkham, Sorkin, & Swift, 2018). The LIML results resemble those of the baseline, which are shown in column (1) of Table 5.

dynamic effects of innovation on the local demand for education over the short-to-medium term.

Table 5: Evidence of Technology-Induced Demand for Education

Model	(1)	(2)	(3)
Number of Lags	(0)	(4)	(6)
A: Percentage of Population aged 18-24 Enrolling in Colleges			
Net Migration Rate	-0.34***	-0.51***	-0.43***
	(0.102)	(0.132)	(0.122)
IHS(Patent)	0.006***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)
Observations	4,915	3,122	2,214
B: Percentage of Total Population Enrolling in Colleges			
Net Migration Rate	-0.47***	-0.63***	-0.54***
	(0.134)	(0.154)	(0.133)
IHS(Patent)	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)
Observations	5,008	3,185	2,261

Note: Robust standard errors are in parenthesis. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. In Columns (2) and (3), all independent variables are lagged four and six years, respectively. When lagging the independent variables, we lose some observations per county.

The results shown in Column (1) of Table 5 are the base estimations (the independent variables are not lagged). The independent variables are lagged four and six years in columns (2) and (3), respectively. Panels A and B of Table 5 show the relationships between migration and patent counts, respectively, and (a) the percentage of the population aged 18 to 24 enrolled in higher education and (b) the percentage of the total population enrolled in colleges. From these estimations, we reach two conclusions. First, innovation is positively correlated with the enrollment in higher education, and the four-year-lagged specifications (corresponding to the average time required to obtain a college degree) yield the highest estimated coefficients of the patent variable. That is, conditional on migration, innovation is positively associated with local demand for college education. Second, migration is negatively correlated with the demand for higher education. One possibility is that because migrants tend to be highly educated, in-migration of high-educated workers increases the supply of skilled workers and crowds out some locals from obtaining college education.[36] Another pos-

[36]In Table 5, the migration coefficients are between -0.6 and -0.3—a magnitude close to the marginal effect of migration on the percentage of college workers estimated in Section 5 (i.e., $\frac{\partial College}{\partial M} \approx 0.5$). These results are consistent with the expectations outlined in our theoretical appendix. To see this, in a given county, let M be the net migration and *Enrollment* be the local demand for education. If θ_H is the marginal effect of migration on the stock of college workers, the total number of high-skilled labor is given by $N_H =$

sibility is that local college students could be crowded out by nonlocal college students. That is, there could be a competition between local and nonlocal students for enrollment in universities, especially at prestigious innovative institutions (Curs & Jaquette, 2017).

Perhaps a better way to show the relationships between the demand for college education and the distributions of low- and high-educated workers is to look at the association between college enrollment and the share of high- or low-educated workers in the medium run. We estimate the following regression specification:

$$\mathit{\Delta Education}_{i,s,t,t-4} = \breve{\theta}_M M_{i,s,t-4} + \breve{\theta}_E Enrollment_{i,s,t-4} + \tilde{C} + FE_t + \breve{e}_{i,s,t,t-4}. \tag{10}$$

Here, $\mathit{\Delta Education}_{i,s,t,t-4}$ is the change in either the population share of college or non-college workers in county i from *t and t-4*. This time span is the average years required to obtain a college degree. M_{ist-4} is the net migration rate in county i at time *t-4*. $Enrollment_{ist-4}$ is either the share of population aged 18-24 or the share of total population that enrolls in college lagged four year*s*. We also control for the percentages of the population that work in the farm and non-farm sectors lagged four periods. To control for regional labor market shocks related to industry composition, these estimations include industry mix employment growth and the interaction term between population growth and patent growth at time *t*. Finally, we control for the year effects.

From Table 6, the estimated effects of the net migration rate are very similar whether we estimate with patents or with college enrollment, and these give us confidence in the analysis of the technology-induced demand for college education. Conditional on the net migration rate, our estimations of the effects of college enrollment and patent variables (despite imprecise estimates) are also quite similar to one another. Given the high mobility of college workers, the non-migratory channel is better captured by the change in non-college workers than by the change in college workers. In particular, both patents and college enrollment produce similar estimated effects on changes in the percentage of non-college workers (i.e., from -0.002 to -0.004).

$\theta_H M + Enrollment$. Conditional on the demand for high-educated labor remaining fixed in the short run, $\frac{dEnrollment}{dM} = -\theta_H$. Thus, the marginal effect of the migration on the percentage of high-skilled workers ($\frac{\partial College}{\partial M}$) should be similar to the negative value of the marginal effect of migration on the demand for college education ($\frac{\partial Enrollment}{\partial M}$).

Table 6: Evidence of Medium Run Effects of Education on the Skill Distributions

Model	(1)	(2)	(3)
A. Dependent Variable: ΔPercentage of College Workers			
Lagged Net Migration Rate	0.12	0.09	0.09
	(0.071)	(0.065)	(0.064)
Lagged IHS(Patent)	0.003***	-	-
	(0.001)		
Lagged Share of Population Aged 18-24	-	0.006	-
Enrolling in College		(0.008)	
Lagged Share of Total Population	-	-	0.001
Enrolling in College			(0.007)
Observations	3,128	3,083	3,083
B. Dependent Variable: ΔPercentage of Non-college Workers			
Lagged Net Migration Rate	-0.32***	-0.27***	-0.27***
	(0.084)	(0.038)	(0.038)
Lagged IHS(Patent)	-0.004***	-	-
	(0.001)		
Lagged Share of Population Aged 18-24	-	-0.002	-
Enrolling in College		(0.011)	
Lagged Share of Total Population	-	-	-0.004
Enrolling in College			(0.009)
Observations	3,128	3,139	3,139
Year Effects	Yes	Yes	Yes

Note: Robust standard errors are in parenthesis. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. The dependent variable is the four-year change in the population share of college or non-college workers. Column (1) shows the results of the IVs estimations of migration and patent lagged four years. Because of missing observations in the college enrollment variables, the number of observations are not identical across columns.

8 Innovation and Urban Income/Wage Divergence

8.1 Innovation and Income/Wage Divergence over the Short-to-Medium Term

The issue of urban income/wage divergence has long prompted academic and political debate. Therefore, we provide evidence of the effects of innovation on interregional income and wage divergence over both the short and medium runs. Over the long run, as posited by endogenous growth theory, an increase in human capital and innovation can induce strong economic growth (Romer, 1990). Additionally, as pointed out by Moretti (2013b), highly innovative regions attract and employ highly educated workers. Therefore, innovation should lead to a high employment of high-educated workers and an increase in regional average income/wage.

Table 7: Innovation and Urban Income/Wage Divergence

Estimator	Fixed Effects			Two-Stage Least Squares		
Number of Lags	(0)	(4)	(8)	(0)	(4)	(8)
A. Dependent Variable: Natural Log of Income Per Capita						
IHS(Patent)	0.02***	0.02***	0.03***	0.05***	0.05***	0.05**
	(0.002)	(0.002)	(0.004)	(0.015)	(0.012)	(0.020)
First Stage F-statistic	-	-	-	38.11	27.87	14.07
Observations	5,146	3,140	1,306	5,146	3,140	1,306
B. Dependent Variable: Natural Log of Average Wage						
IHS(Patent)	0.04***	0.04***	0.04***	0.07***	0.06***	0.098
	(0.002)	(0.002)	(0.003)	(0.012)	(0.010)	(0.301)
First Stage F-statistic	-	-	-	30.62	25.40	14.21
Observations	4,581	2,809	1,186	4,581	2,809	1,186

Note: Robust standard errors are in parenthesis. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. The regressions include the natural log of college-to-non-college-workers ratio, state effects, and state-specific time trends. For instrumental variable estimations, the *IHS(Patent)* is instrumented by $\widehat{PATENT}_{it}$ defined in equation (8). When lagging the independent variables, we lose some observations per county.

As in Katz and Murphy (1992), we run the following specification to estimate the dynamic association between innovation and income/wage divergence between urban counties:

$$I_{ist} = \widetilde{\alpha}_1 Patent_{i,t-l} + \widetilde{\alpha}_2 ln\left(\frac{H}{L}\right)_{i,s,t-l} + FE_{st} + \tilde{e}_{ist}, \tag{11}$$

where I is either the natural log of per capita income ($\frac{income}{population}$) or average wage ($\frac{wage}{working\ population}$).[37] *Patent* is the inverse hyperbolic sine function of the total patent counts. We also conduct a sensitivity analysis by replacing *IHS(total patent counts)* with the *IHS(per capita patents).* These results are virtually indistinguishable.[38] This specification controls for the natural log of ratio of college to non-college workers (i.e., $ln\left(\frac{H}{L}\right)$), state effects, and state-specific trends. The estimated results are virtually identical when we either omit the state effects or replace the state-specific trends with the year effects.

To study the effects of innovation over the short-to-medium term, we estimate the regression (11) with no lag, a four-year lag, and eight-year lags of the independent variables.

[37] We also attempt to control for the cost of living by using the median housing rent at the MSA level to adjust per capita income and average wage. The findings after this adjustment resemble those presented in this paper.

[38] The results are available upon request.

Table 7 shows the results of both fixed effects and IV estimators.[39][40] Following suggestions of Angrist & Pischke (2008); and Goldsmith-Pinkham, Sorkin, & Swift (2018), we also probe the validity of our instrument used in (14) by employing three different IV estimators, including two-stage least square (2SLS), generalized method of moments (GMM), and limited information likelihood (LIML). Particularly, LIML has a very good small sample property, such that it can lessen the 2SLS bias (Hahn, Hausman, & Kuersteiner, 2004). The similar results between 2SLS, GMM, and LIML suggest the validity of our instrument.

The results shown in Table 7 support the hypothesis that innovation increases both income/wage divergence between urban counties over the short-to-medium term. This relationship appears to be very stable over time.

8.2 Inside the Black Box of Skilled–Biased Technological Change

We have examined the impact of innovation on urban income/wage divergence. Our results raise an interesting question: What types of technology cause spatial income/wage inequality? To address this question, we examine three specific types of technology that have received considerable attention: computer and data processing (Beaudry, Doms, & Lewis, 2010; Jorgenson, 2001); telecommunications (Roller & Waverman, 2001); and automation (Autor, Levy, & Murnane, 2003).

To study the relationship between these technologies and urban income/wage divergence, we use the patent data provided by the U.S. Patent and Trademark Office (USPTO); and Mann and Püttman (2018). The USPTO provides patent counts by primary classifications in metropolitan areas. We use technology classes 700-720, 725, and 726 to get patent counts for computer and data processing. Class 455 is used for telecommunications.[41,42,43] Mann and Püttman (2018) provide data on automation patents at the commuting zones (CZs),

---

[39] For IV estimations, the $IHS(Total\ Patent\ Counts\ )$ is instrumented by $\widehat{PATENT}_{it}$ defined in (8).

[40] To reiterate, our instruments are specifically designed for the skill and migration regressions (i.e., specifications (5) and (6). Since the structures of the error terms in income/wage regression (i.e., specification (11)) might be different from those in the skill and migration regressions, there could be confounding factors which invalidate the use of our patent IV in the model (11).

[41] The data on patenting in technology classes breakout by metropolitan areas was retrieved on December 31, 2018, from https://www.uspto.gov/web/offices/ac/ido/oeip/taf/cls_cbsa/explan_cls_cbsa.htm.

[42] USPTO defines metropolitan areas on the basis of U.S. Office of Management and Budget (OMB) Bulletin No. 10-02. To conform to this definition, we use the December 2009 metropolitan definitions published by the U.S. Census. The delineation files are accessible at https://www2.census.gov/programs-surveys/metro-micro/geographies/reference-files/2009/historical-delineation-files/list1.txt.

[43] Using December 2009 MSA definitions of the Census Bureau, we identify counties that belong to each MSA. Then, using all available data, we aggregate all county-level data to metropolitan-level data.

whose definitions are based on Tolbert and Sizer (1996).[44,45]

As we have done in our previous analyses in (11), we run the following regression specification,

$$I_{ist} = \widetilde{\alpha}_1 Patent_{ist} + \widetilde{\alpha}_2 ln\left(\frac{H}{L}\right)_{ist} + FE + \tilde{e}_{ist}, \tag{12}$$

where the inverse hyperbolic sine function (*IHS*) of patent counts now is the *IHS* of patent counts in either computer and data processing, telecommunications, or automation. To reiterate, due to the availability of our data, we study computer and data processing, and telecommunications at the MSA level; we study automation at the CZ level. The estimated results are consistent with our findings in section 8.1 using the county data as well as with previous studies using data at the MSA or CZ level.

Because individual commuting zones can cross state boundaries, we cannot control for state-specific trends in the regression using automation patents. Therefore, we control for trends in each CZ and the CZ fixed effects. Note, our results are robust even we omit either state or commuting zone fixed effects. They are also robust when we replace the CZ- or state-trend effects with the year effects. Table 8 shows the relations between innovations in computer and data processing, telecommunications, automation, and cross-urban income/wage inequality. Consistent with previous findings, we find that all of these technologies have significantly positive relations with cross-urban income measures. Yet during the period of this study, computer and data processing had a stronger association with urban income/wage than telecommunications. It is worth mentioning that patents reflect both the quantitative and qualitative effects of technology. For instance, unlike the amount of workplace computer use, which in the U.S. has dramatically increased, innovation in computer and data processing could better capture the impact of skill-biased technological change on urban divergence. Both high- and low-educated workers widely use computers, but perhaps high-educated workers are first adopters of new innovations in computer and data processing technology. High-educated workers, in other words, might easily adapt to new technology (Galor & Moav, 2000).

[44] Automation patents were retrieved on December 31, 2018, from https://github.com/lpuettmann/automation-patents.

[45] To be consistent with our MSA analysis, we identify counties that belong to each metropolitan area using the December 2009 metropolitan definition of the Census Bureau. Then we construct data at the CZ level by aggregating available county data.

Table 8: Computer and Data Processing, Telecommunications, Automation, and Urban Income/Wage Divergence

Model	(1)	(2)
A. Dependent Variable: Natural Log of Per Capita Income		
Computer and Data Processing	0.030***	-
	(0.002)	
Telecommunications	0.010***	-
	(0.002)	
Automation	-	0.007***
		(0.001)
Observations	2,592	2,803
B. Dependent Variable: Natural Log of Average Wage		
Computer and Data Processing	0.034***	-
	(0.002)	
Telecommunications	0.020***	-
	(0.002)	
Automation	-	0.003***
		(0.0004)
Observations	2,592	2,803
Regional Level	MSA	CZ

Note: Robust standard errors are in parenthesis. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

9 Sensitivity Analysis

To probe the robustness of our estimates of the two channels of labor supply, we further examine four potential issues: immigration, commuting of high-skilled workers, knowledge spillovers, and serial correlation. We find the base results, which are presented in Column (1) of Table 2, soundly robust to the consideration of these issues.

First, one might worry about how immigration affects our estimates. Consistent with the literature, the baseline only focuses on internal migration. Yet we re-estimate the regressions (5) and (6) by replacing the internal net migration rate with the total net migration rate—i.e., the internal net migration rate plus the net immigration rate. Our base results are resemble those reported in Table B.1 of Appendix B.

Second, there is the concern about the analysis at the county level that could fail to consider commuting to work of high-educated workers. To reiterate, if commuting reduces in-migration of high-educated workers and percentage of high-skilled workers living in a highly innovative region, omitting commuting will bias our results upward—though as we noted, the result could still be interpreted as the total human capital that is imported into a county. Since we find that the migratory channel played a lesser role in explaining skill distributions from 2005 to 2015, accounting for commuting will only strengthen this finding. Following Partridge, Rickman, Ali, and Olfert (2009), we account for commuting and other economic spillovers by controlling for the distance from the population-weighted center of a county to the population-weighed center of its own MSA.[46] In Appendix B, the results, which are shown in Column (2) of Table B.2, confirm our hypothesis. That is, controlling for commuting to work in the estimations reduces the estimated effect of migration on distribution of high-educated workers.

Third, there could be another concern about spatial autocorrelation, particularly knowledge spillovers, in our patent variable. To capture the potential spillovers, we include an exponential distance-decay function of patent counts in the baseline estimations. Specifically, for any county i, this function is defined as $\sum_{j \neq i} \exp(-distance_{ij}) patent_{jt}$, where $distance_{ij}$ is the distance between counties i and j, and $patent_{jt}$ is patent counts in other contiguous U.S. counties. Andersson and Gråsjö (2009) show that this exponential distance-decay function can effectively control for spatial autocorrelation in the analysis of knowledge production function.[47] The results, shown in Column (3) of Table B.2, are similar to those of the baseline.

The fourth concern is the issue of serial correlation. To address this concern, we employ the Blundell-Bond estimator, which is an instrumental variables estimation. The results shown in Appendix C support the baseline.

Finally, we also analyze other issues, including differing definitions of high- and low-educated workers, capital-skill complementarity (Krusell, Ohanian, Ríos-Rull, & Violante, 2000), and the agglomeration effect (Baum-Snow, Freedman, & Pavan, 2018). We find our results robust.[48]

10 Conclusion

We seek to understand the role of technology in increasing skill and income inequalities across U.S. urban areas. The previous literature on skill-biased technical change has been empirically silent about the channels of labor supply through which technology causes skill inequality across cities. Consequently, our analysis of these channels has considerable value. Furthermore, our analysis offers significant information to policymakers who attempt to promote regional growth by investing in new technologies, attracting skilled workers, and

[46]Follow Partridge, Rickman, Ali, and Olfert (2009), we use 1990 population–which could lessen the endogeneity, such as agglomeration, patent growth, and population growth.

[47]The county distance database is compiled by Roth (2014) using Census 2000 Summary File 1. The data is accessible at https://www.nber.org/data/county-distance-database.html.

[48]The analyses are available upon request.

increasing educational attainment.

Using the U.S. patents as a novel measure of the inducement of urban skill divergence through skill-biased technical change, we identify two key technological channels of labor supply: migration and local college demand prompting a local labor supply response. Our IVs estimations show that the demand for college education plays a key role in explaining skill divergence between U.S. urban counties. This finding conforms to two empirical facts: a post-2005 shape decline in the U.S. internal migration rate; and a simultaneous increase in educational attainment. The patent data also allows me to shed light on the inside of the black box of skill-biased technical change and urban income/wage divergence. Specifically, we show strong positive relations between urban income/wage inequality and three specific high technologies: computer and data processing, telecommunications, and automation.

In conclusion, this paper advances the study of innovation and spatial inequality. The use of patents as a measure of technological change provides new insights into the relationship between technology and urban inequalities in skills and income/wage. Further, by examining the impacts of the two technological channels on skill inequality across urban areas, this paper offers valuable guidance to policymakers interested in place-based policies and economic development.

References

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, 4, 1043-1171.

Acs, Z. J., Anselin, L., & Varga, A. (2002). Patents and innovation counts as measures of regional production of new knowledge. *Research policy*, 31(7), 1069-1085.

Aghion, P., Bergeaud, A., Blundell, R., & Griffith, R. (2017). Innovation, firms and wage inequality. Retrieved December 12, 2018, from https://scholar.harvard.edu/files/aghion/files/innovations_firms_and_wage.pdf

Akcigit, U., Grigsby, J., & Nicholas, T. (2017). Immigration and the Rise of American Ingenuity. *American Economic Review*, 107(5), 327-331.

Andersson, M., & Gråsjö, U. (2009). Spatial dependence and the representation of space in empirical models. *Annals of Regional Science*, 43(1), 159-180.

Angrist, J. D., & Pischke, J.-S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton, NJ: Princeton University Press.

Autor, D. H., & Dorn, D. (2013). The Growth of low-skill service jobs and the polarization of the U.S. labor market. *American Economic Review*, 103(5), 1553-97.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The Skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4), 1279-1333.

Autor, D., Katz, L. F., & Kearney, M. S. (2008). Trends in U.S. wage inequality: Revising the revisonists. *Review of Economics and Statistics*, 90(2), 300-323.

Bartik, T. J. (1991). Who benefits from state and local economics development policies? Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Baum-Snow, N., Freedman, M., & Pavan, R. (2018). Why has urban inequality increased? *American Economic Journal: Applied Economics*, 10(4), 1-42.

Beaudry, P., Doms, M., & Lewis, E. (2010). Should the personal computer be considered a technological revolution? Evidence from US metropolitan areas. *Journal of Political Economy*, 118(5), 988-1036.

Berry, C. R., & Glaeser, E. L. (2005). The divergence of human capital levels across cities. *Papers in Regional Science*, 84(3), 404-444.

Betz, M. R., Partridge, M. D., & Fallah, B. (2015). Smart cities and attracting knowledge workers: Which cities attract highly-educated workers in the 21st century. *Papers in Regional Science*, 95(4).

Blanchard, O. J., & Katz, L. F. (1992). Regional evolutions. Brookling Papers on Economic Activity, 23(1), 1-76.

Borjas, G. J., Freeman, R. B., & Katz, L. F. (1996). Searching for the effect of immigration on the labor market. *American Economic Review*, 86(2), 246-251.

Bottazzi, L., & Peri, G. (2007). The international dynamics of R&D and innovation in the long run and in the short run. *The Economic Journal*, 117(518), 486-511.

Bound, J., & Johnson, G. (1992). Changes in the structure of wages in the 1980s: An evaluation of alternatives explanations. *American Economic Review*, 133(1), 207-241.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. J*Journal of Labor Economics*, 19(1), 22-64.

Ciccone, A., & Peri, G. (2005). Long-run substitutability between more and less educated workers: Evidence from US states, 1950–1990. *Review of Economics and Statistics*, 87(4), 652-663.

Cragg, G. J., & Donald, G. S. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, 222-240.

Curs, B. R., & Jaquette, O. (2017). Crowded out? The effect of nonresident enrollment on resident access to public research universities. *Educational Evaluation and Policy Analysis*, 39(4), 535-565.

Davidson, R., & MacKinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, 49(3), 781-793. doi:10.2307/1911522.

Détang-Dessendre, C., Partridge, M. D., & Piguet, V. (2016). Local labor market flexibility in a perceived low migration country: The case of French labor markets. *Regional Science and Urban Economics*, 58, 89/103.

Diamond, R. (2016). The Determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000. *American Economic Review*, 106(3), 479-524.

Eeckhout, J., Pinheiro, R., & Schmidheiny, K. (2014). Spatial sorting. *Journal of Political Economy*, 122(3), 554-620.

Ellen, G. I., & O'Regan, K. (2011). Gentrification: Perspectives of economists and planners. In The *Oxford Handbook of Urban Economics and Planning*. Oxford University Press.

Foster, T. R., Ellis, M., & Fiorio, L. (2018). The Opportunities and Challenges of Linked IRS Adminstrative and Census Survey Records in the Study of Migration. Center of Economic Studies, US Census Bureau, No. 2018-06.

Galor, O., & Moav, O. (2000). Ability-biased technological transition, wage inequality, and economic growth. *Quarterly Journal of Economics*, 115(2), 469-497.

Giannone, E. (2017). Skill-biased technical change and regional convergence. Retrieved 12 6, 2018, from https://home.uchicago.edu/∼elisagiannone/files/ JMP_ElisaG.pdf

Goldin, C., & Katz, L. F. (2010). The Race between education and technology. Belknap Press.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2018). Bartik instruments: What, when, why, and how (No. W24408). National Bureau of Economic Research.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review*, 104(8), 2509-26.

Grillches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, 28(4), 1661-1707.

Hahn, J., Hausman, J., & Kuersteiner, G. (2004). Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. *Econometrics Journal*, 7(1), 272-306. doi:https://doi.org/10.1111/j.1368-423X.2004.00131.x

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001).The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper No. 8498. DOI: 10.3386/w8498

Hårsman, B., & Quigley, J. M. (1998). Education, job requirements, and commuting: An analysis of network flows. In M. J. Beckmann, B. Johannsson, F. Snickars, & R. Thord, Knowledge and Networks in a Dynamic Economy (pp. 261-272). Berlin, Heidelberg: Springer.

Jaeger, D. A., Ruist, J., & Sthler, J. (2018). Shift-share instruments and the impact of immigration (No. w24285). National Bureau of Economic Research.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits and market value. *American Economics Review*, 76, 431-437.

Jaffe, A. B., Trajtenberg, M., & Handerson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly journal of Economics*, 108(3), 577-598.

Jones, J. B., & Yang, F. (2016). Skill-biased technical change and the cost of higher education. Journal of Labor Economics, 34(3), 621-662.

Jorgenson, D. W. (2001). Information Technology and U.S. Economy. *American Economic Review*, 91(1), 1-32.

Juhn, C., Murphy, K., & Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Econonomy*, 101(3), 35-78.

Katz, L., & Murphy, K. (1992). Change in relative wages, 1963-1987: supply and demand factors. *Quaterly Journal of Economics*, 107(1), 35-78.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *Quarterly Journal of Economics*, 132(2), 665-712.

Krusell, P., Ohanian, L. E., Ríos-Rull, J.-V., & Violante, G. L. (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica*, 68(5), 1029-1053.

Mann, K., & Püttmann, L. (2018). Benign effects of automation: New evidence from patent texts. Retrieved October 3, 2018, from SSRN: https://ssrn.com/abstract=2959584 or http://dx.doi.org/10.2139/ssrn.2959584

Michaelsen, M. M. (2011). The Hidden increase in wage inequality: Skill-biased and ability-biased technological change . Ruhr Economic Papers No. 262, RWI Essen.

Moretti, E. (2011). Local labor markets. *Handbook of Labor Economics*, 4, 1237-1313.

Moretti, E. (2013a). Real wage inequality. *American Economics Journal: Applied Economics*, 5(1), 65-103.

Moretti, E. (2013b). The new geography of jobs. Boston, MA: Mariner Books.

National Center For Education Statistics (2019). The condition for education. Retrieved September 21, 2019, from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2019144

Partridge, M. D., Rickman, D. S., Ali, K., & Olfert, M. R. (2009). Agglomeration spillovers and wage and housing cost gradients across urban hierarchy. *Journal of International Economics*, 78(1), 126-140.

Partridge, M. D., Rickman, D. S., Olfert, M. R., & Ali, K. (2012). Dwindling US internal migration: Evidence of spatial equilibrium or structural shifts in local labor markets. *Regional Science and Urban Economics*, 42(1-2), 375-388.

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

Roller, L.-H., & Waverman, L. (2001). Telecommunications infrastructure and economic development: A simultaneous approach. *American Economic Review*, 91(4), 909-923.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5, Part 2), S71-S102.

Rosen, S. (1979). Wage-based indexes of urban quality of life. In P. Mieszkoski M. Straszheim (Eds), *Current Issues in Urban Economics*. Baltimore: John Hopkins University Press.

Roth, J. (2014, June 11). County Distance Database. The National Bureau of Economic Research. Retrieved March 21, 2019, from https://www.nber.org/data/county-distance-database.html

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

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. Andrews, & J. H. Stock (Eds.), In identification and inference for econometric models: Essays in honor of Thomas Rothenberg (pp. 80-108). New York: Cambridge University Press.

Tolbert, C. M., & Sizer, M. (1996). U.S. commuting zones and labor market areas: A 1999 update. Economic Research Service Staff Paper No. 9614.

Walsh, J. P., & Sadao, N. (2009). Who invents? Evidence from the Japan-U.S. inventor survey. RIETI Discussion Papers.

Appendix A: Theoretical Framework

To illustrate the effects of innovation on the distributions of skills and regional income through migration and college educational demand, we present a very stylized model which builds on the Rosen-Roback model and the "canonical model" of skill-biased technical change. To obtain a tractable model, we adopt the following assumptions. First, we assume that there are two regions, and the economy exists for two periods. We further simplify the regional production function by assuming perfect substitution between high- and low-educated workers. After dropping the time subscript, the aggregate output for the representative firm in county i at time t is given by:

$$Y_i = (X_{i,H} \cdot N_{i,H}^{1-\beta} + X_{i,L} \cdot N_{i,L}^{1-\beta}) \cdot \overline{K}_i^{\beta}, \tag{A.1}$$

where Y is the total output with the price normalized to be one; $X_{i,e}$ is exogenous skill-specific technology for workers of skill e; $N_{i,e}$ is the total number of workers of skill e in county i, and $\overline{K}_i$ is the stock of capital which is assumed to be constant.

At the beginning of time, we assume that the total number of high- and low-educated workers are equal and normalized to be one, $N_{H,0} = N_{L,0} = 1$. At the time, $t = 0$, let us also assume that the two regions are endowed with the same skill-specific technologies. That is, $X_{i,e,0} = X_{k,e,0} = 1$. In period 1, there is a positive shock to the productivity of high-educated workers in county i. In other words, there is new innovation introduced in county i in period 1.

The migration and education decisions of workers is assumed as follows. First, workers migrate in response to the differential in wages, amenity and housing rents between the two regions. Then, a fraction of low-educated workers who are living in region i at the beginning decide whether to invest in education in response to innovation in region i. Perhaps, this might not perfectly capture the reality of the decision-making processes of workers. However, it helps us simplify the specifications of labor supply of college and non-college workers. This assumption can circumvent the mobility of low-educated workers to innovative regions in order to accumulate more human capital and earn higher wage. This assumption is likely to hold given that mobility cost of low-educated workers is high. Additionally, it might be costly to obtain education abroad than at the home region. That is, in-state tuition is likely to be lower than out-of-state tuition. While this assumption is used to simplify the conceptual analysis, it is not required in the empirical study because we exam the effect of net migration rate on skill distributions.

The labor supply functions of high- and low-educated workers at time t are specified as follows:

$$N_{i,H} = \left\{ \frac{1}{2} + \theta_H ln \left[\left(\frac{W_{i,H}}{W_{k,H}} \right)^{\delta_1} \left(\frac{A_i}{A_k} \right)^{\delta_2} \left(\frac{R_{i,H}}{R_{k,H}} \right)^{-\delta_3} \right] \right\} + \theta_E ln \left(\frac{X_{i,H}}{X_{i,L}} \right), \tag{A.2}$$

and

$$N_{i,L} = \left\{ \frac{1}{2} + \theta_L ln \left[\left(\frac{W_{i,L}}{W_{k,L}} \right)^{\delta_1} \left(\frac{A_i}{A_k} \right)^{\delta_2} \left(\frac{R_{i,L}}{R_{k,L}} \right)^{-\delta_3} \right] \right\} - \theta_E ln \left(\frac{X_{i,H}}{X_{i,L}} \right), \tag{A.3}$$

where A_i is the amenity value in county i, and $R_{i,e}$ is the housing rent paid by a worker of skill e in county i. The first term in the curly brackets capture the relative values of wages, amenities and housing rents between counties i and k. Therefore, they reflect the effects of innovation through migration on skill distributions, with the marginal effects of $\theta'_M s$ for $\in \{H, L\}$. We assume all parameters, including $\theta' s$ and $\delta' s$, to be weakly positive.[1] The second term captures another channel through which innovation can affect the distributions of skills. Given a weakly positive value of θ_E, an increase in innovation will increase the number of college workers and reduce the same number of low-educated workers. We call this channel the educational demand, and it is induced by innovation. We assume that $\theta_E ln \left(\frac{X_{i,H}}{X_{i,L}} \right) < N_{i,L}$. That is, there are some low-educated workers who do not demand more schooling at time $t = 1$.

The housing demand is just the total population of each type of workers in each county. The supply of housing for high- and low-educated workers of at time t are given as,

$$R_{i,H} = B_{i,H} N_{i,H}, \tag{A.4}$$

and

$$R_{i,L} = B_{i,L} \left(N_{i,L} \right) \cdot h(N_{i,H}). \tag{A.5}$$

Here, $B_{i,e} \geq 0$ is a constant elasticity of housing supply in county i, and $h'(N_{i,H}) > 0$. Therefore, a large number of $B_{i,e}$ indicates that houses workers with skill e demand are limited in county i. We model different housing supply functions for different types of workers. For high-educated workers, the housing supply is only a function of the number of college workers living in that county. On the contrary, the housing supply function for low-educated workers depends on both the number of low- and high-educated workers. Therefore, an

[1] For the first part, which represents the migratory channel, of the aggregate labor supply equations, one can derive this part by set up a utility maximization of a representative consumer/worker who enjoys benefits of wage, housing, and amenity. See, for example, Moretti (2013a) Moretti (2013a) who specifies similar aggregate labor supply equations of college and non-college workers.

increase in the demand for housing of high-educated workers can lead to higher housing rent for low-educated workers. The rationale is that an increase in the high- population of educated workers could lead to gentrification and finally out-migration of low-educated workers.[2]

While we model different supply functions for different types of workers, by no mean this assumption always holds in reality. Yet we find suggestive evidence in our empirical analysis that the influx of high-educated migrants could lead to an outflow of local low-educated workers. Yet the cross-price effect is not an interest of this paper; we leave this interesting issue to future research.

Proposition 1 ***Innovation in county i increases the total supply of high-educated workers through migration and educational attainment. Additionally, an influx of high-educated migrant due to innovation could lead to an out-migrant of low-educated workers.***

Proof. First, a representative firm's maximization in region i yields,

$$W_{i,e} = (1-\beta)\, X_{i,e}\ \overline{K_i}^{\beta} N_{i,e}^{-\beta}. \tag{A.6}$$

Substitute (A.4) and (A.6) into (A.2), we can rewrite the labor supply function for skill H as follows,

$$f\left(n_{i,H}\right) = Q + \theta_H \beta\delta_1 \left(x_{i,H} - x_{k,H}\right) + \theta_E (x_{i,H} - x_{k,L}), \tag{A.7}$$

where the lower case letters denote the natural logarithmic functions. $Q = \frac{1}{2} + \theta_H[\beta\delta_1(k_i - \beta k_k) + \delta_2\left(a_i - a_k\right) - \delta_3\left(b_{i,H} - b_{k,H}\right)]$ and $f\left(n_{i,H}\right) = n_{i,H} + \theta_H \left(ln\frac{N_{i,H}}{1-N_{i,H}}\right)\left(\beta\delta_1 + \delta_3\right)$. Apply the implicit function theorem to (A.7), we obtain:

$$\frac{\partial N_{i,H}}{\partial x_{i,H}} = \left(\frac{\partial f}{\partial N_{i,H}}\right)^{-1} \left(\beta\delta_1\theta_H + \theta_E\right)\ \geq 0. \tag{A.8}$$

θ_H captures the response of migrants toward innovation in county i, and θ_E captures the response of local people by increasing demand in education. Whenever low-educated workers do not respond to a differential in wages across regions (that is, $\theta_L = 0$), innovation only affects the supply of non-college jobs through increasing educational attainment.

Second, since an innovation increases the number of high-educated workers in county i and therefore reduces the number of high-educated workers in county k, an innovation will lead to higher housing rent for low-educated workers in county i than in county k. That is,

[2] For an overview of gentrification, see Ellen and O'Regan (2011) Ellen and O'Regan (2011)

$\frac{\partial R_L}{\partial N_H} = B_L(N_L) \cdot h'(N_H) > 0$. Therefore, low-educated workers in county i migrate out in response to relatively high housing cost caused by an influx of high-educated migrants. ■

Proposition 2 ***In the short/medium run, for any given population of high skill workers and with a fixed number of new innovation, the increase in the marginal effect of migration reduces the marginal demand for education.***

Proof. Fixed $N_{i,H} = \overline{N}_{i,H}$ and total differentiate (A.2) with respect to θ_H and θ_E,

$$\frac{d}{d\theta_H}(\theta_E) = -\frac{x_{i,H} - x_{i,L}}{D} \leq 0 \ if \ x_{i,H} > x_{i,L}, \tag{A.9}$$

where $D = N_{H,0}\{\frac{1}{2} + ln\left[\left(\frac{W_{i,H}}{W_{k,H}}\right)^{\delta_1}\left(\frac{A_i}{A_k}\right)^{\delta_1}\left(\frac{R_i}{R_k}\right)^{-\delta_3}\right]\}$. That is, conditional on the number of jobs for high-educated workers available in the county, an increase in high-educated migrants in a region affected by innovation will reduce the educational demand for low-educated workers. ■

Proposition 3 ***Innovation increases regional average wage.***

Proof. The average wage in region i is given by:

$$\mu_i^w = \frac{N_{i,H}}{N_i}W_{i,H} + (1 - \frac{N_{i,H}}{N_i})W_{i,L}, \tag{A.10}$$

where μ_i^w is the average wage in region i. Take derivative of μ_i^w with respect to $X_{i,H}$, we get:

$$\frac{\partial \mu_i^w}{\partial X_{i,H}} = \frac{\partial}{\partial X_{i,H}}\left(\frac{N_{i,H}}{N_i}\right)(W_{i,H} - W_{i,L}) + \frac{N_{i,H}}{N_i}(\frac{\partial W_{i,H}}{\partial X_{i,H}}). \tag{A.11}$$

First, innovation in region i attracts high-educated workers and increases the demand for college education. Consequentially, $\frac{\partial}{\partial X_{i,H}}\left(\frac{N_{i,H}}{N_i}\right) > 0$. Second, innovation also raises wages of high-educated workers; therefore, $\frac{\partial W_{i,H}}{\partial X_{i,H}} > 0$. If we assume that high-educated workers earn more than low-educated workers (this situation is generally true in reality), then $W_{i,H} - W_{i,L} > 0$ and $\frac{\partial \mu_i^w}{\partial X_{i,H}} > 0$. That is, innovation raises regional average wage (or per capita income). ■

Our model is static; therefore, it cannot effectively explain the impact of innovation in the long run. If some urban areas experience a temporarily positive productivity shock, spatial income divergence between urban areas will only be temporary. Over the long run, regional income convergence could be achieved through factor mobility (i.e., migration, knowledge spillovers, and capital flows) or local supply of high-skilled workers. However, if the technological stocks of those urban areas continue to grow (endogenously or exogenously) over time and if factor mobility or local supply of skilled workers is relatively low, then income divergence between urban regions will persist.

Appendix B: Sensitivity Analysis

Table B.1: Re-Assessment of Net Migration Rate

Model	Internal Migration Rate	Total Migration Rate
A. Dependent Variable: Net Migration Rate		
IHS(Patent)	0.001***	0.001***
	(0.0002)	(0.0002)
Observations	4,900	4,900
B. Dependent Variable: Percentage of College Workers		
Net Migration Rate	0.53***	0.55***
	(0.188)	(0.193)
IHS(Patent)	0.02***	0.02***
	(0.001)	(0.001)
Observations	5,146	5,146
C. Dependent Variable: Percentage of Non-college Workers		
Net Migration Rate	0.04	0.02
	(0.047)	(0.048)
IHS(Patent)	-0.01***	-0.01***
	(0.0005)	(0.0005)
Observations	5,146	5,146
Population Density	No	Yes

Note: Robust standard errors are in parenthesis. *, ** and *** indicate significance at the 10%, 5% and1% levels, respectively.

Table B.2: Re-Assessment of Commuting and Knowledge Spillovers

Model	(1)	(2)	(3)
	Dependent Variable: Net Migration Rate		
IHS(Patent)	0.001***	0.001***	0.001***
	(0.0002)	(0.0002)	(0.0002)
Observations	4,900	4,900	4,900
	Dependent Variable: Percentage of College Workers		
Net Migration Rate	0.53***	0.44**	0.44***
	(0.188)	(0.170)	(0.171)
IHS(Patent)	0.02***	0.02***	0.02***
	(0.001)	(0.0005)	(0.001)
Observations	5,146	5,146	5,146
	Dependent Variable: Percentage of Non-college Workers		
Net Migration Rate	0.04	0.02	0.02
	(0.047)	(0.048)	(0.048)
IHS(Patent)	-0.01***	-0.01***	-0.01***
	(0.0005)	(0.0005)	(0.0005)
Observations	5,146	5,146	5,146
Distance to the center of MSA	No	Yes	Yes
Distance-weighted Patents	No	No	Yes

Note: Robust standard errors are in parenthesis. *, ** and *** indicate significance at the 10%, 5% and1% levels, respectively. Column (1) is the baseline estimation from Table 2 in the text. Column (2) include the distance to the center of one's own MSA. In Column (2), zero distance is assigned to Bedford and other five independent cities in Virginia. To capture knowledge spillovers, Column (3) include the distance-weighted patent counts of the surrounding counties.

Appendix C: Serial Correlation

Table C.1: Re-Assessment of Serial Correlation in Skill Distribution

	Percentage of College Workers	Percentage of Non-College Workers
L1 of Dependent Variable	0.28***	0.20***
	(0.036)	(0.055)
L2 of Dependent Variable	0.21***	0.11***
	(0.033)	(0.037)
L3 of Dependent Variable	0.16***	0.10***
	(0.023)	(0.025)
L1 of Net migration Rate	0.06	0.16***
	(0.055)	(0.057)
L2 of Net migration Rate	0.07	0.08
	(0.054)	(0.056)
L3 of Net migration Rate	0.05	-0.02
	(0.037)	(0.048)
L1 of IHS(Patent)	0.0017**	-0.00004
	(0.001)	(0.001)
L2 of IHS(Patent)	0.0017**	-0.001
	(0.001)	(0.001)
L3 of IHS(Patent)	0.0015**	0.00073
	(0.001)	(0.001)
Arellano-Bond Test (AR 2)	0.39	0.60
Number of Instruments	76	76
Observations	3,592	3,592

Note: L.n stands for nth-period lag. Arellano-Bond test of no serial correlation (p-value reported). Robust standard errors are in parenthesis. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

There is a concern that serial correlation in the baseline estimations of equation (6), the supply of skill regression, could also affect the results. To deal with this concern (and simultaneity bias), we re-estimate (6) using Blundell-Bond estimator, which is an instrumental variables estimation. We proceed with the following regression,

$$Skill_{i,s,t} = \sum_{1}^{3} \theta_{Skill,t-l} Skill_{i,s,t-l} + \sum_{1}^{3} (\theta_{M,t-l} M_{i,s,t-l} + \theta_{E,t-l} Patent_{i,s,t-l} + \tilde{C}_{i,s,t-l}) + FE_t + e_{i,s,t}, \tag{C.1}$$

where all variables are defined as in regression (6). The results are shown in Table C.1. First, the Arellano-Bond of no serial correlation cannot be rejected at any conventional level. Second, the some coefficients are imprecisely estimated which is expected from using Blundell-Bond estimator. Using estimated coefficients from Table C.1, we can calculate the long-run equilibrium effects of the migratory and non-migratory channels of innovations. The equilibrium effect of net migration rate on the percentage of college workers is around 0.51 ($\approx \frac{0.06+0.07+0.05}{1-0.28-0.21-016}$), and its effect on non-college workers is approximately 0.37 ($\approx \frac{0.16+0.08-0.02}{1-0.2-0.11-01}$),. On the other hand, the effect of patents on the percentage of college labor is around 0.01 ($\approx \frac{0.0017+0.0017+0.0015}{1-0.28-0.21-016}$), and it's effect on the non-college workers is arround -0.0005 ($\approx \frac{-0.00004-0.001+0.0073}{1-0.28-0.21-016}$). Qualitatively, the results of this sensitivity analysis are comparable to the baseline. The main findings are therefore maintained.

Table C.2: Re-Assessment of Serial Correlation in Technological Migration

	Dependent Variable: Net Migration Rate
L1 of Dependent Variable	0.36***
	(0.04)
L2 of Dependent Variable	-0.001
	(0.02)
L1 of IHS(Patent)	0.0007***
	(0.0002)
L2 of IHS(Patent)	0.001***
	(0.0002)
Arellano-Bond Test (AR 2)	0.90
Number of Instruments	52
Observations	5,423

Note: L.n stands for nth-period lag. Arellano-Bond test of no serial correlation (p-value reported). Robust standard errors are in parenthesis. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Finally, to complete the analysis of serial correlation, we investigate this issue in technological migration in equation (5). As in the previous analysis of serial correlation, we proceed as follows,

$$M_{i,s,t} = \sum_{1}^{2} \ddot{\theta}_{M,t-l} M_{i,s,t-l} + \sum_{1}^{2} (\gamma_{t-l} Patent_{i,s,t-l} + \ddot{C}_{ist}) + FE_t + \epsilon_{ist}. \tag{C.2}$$

However, because of collinearity issue, we only include 2-periods lags of the dependent and independent variables and 2006 to 2010 year effects (which are correspondent to the great recession). To further lessen the concern about serial correlation, we only control for the average patent counts in the U.S. and natural log of industry-mix employment growth; we omitted the income and housing rent variables. From Table C.2, the long-run equilibrium effect of technological migration is 0.003 (i.e., $\frac{0.0007+0.001}{1-0.36+0.001}$) which is comparable to the baseline (i.e., 0.001).