

Spatial Changes in the College Wage Premium

Joanne Lindley* and Stephen Machin**

August 2012

* *Department of Economics, University of Surrey*

** *Department of Economics, University College London and Centre for Economic Performance, London School of Economics*

Abstract

We study spatial changes in the US college wage premium for states and cities using US Census and American Community Survey data between 1980 and 2010. We report evidence of significant and persistent spatial disparities in the college wage premium for US states and MSAs. We use estimates of spatial relative demand and supply models to calculate implied relative demand shifts for college graduates vis-à-vis high school graduates and show that relative demand has shifted faster in places that have experienced faster increases in R&D intensity and shifted slower in places where union decline has been faster. Overall, our spatial analysis complements research findings from labour economics on wage inequality trends and from urban economics on agglomeration effects connected to education and technology.

JEL Keywords: College wage premium; Relative demand; Spatial changes.

JEL Classifications: J31; R11.

Corresponding author: Joanne Lindley: j.lindley@surrey.ac.uk.

Acknowledgements

We would like to thank Gilles Duranton, Lowell Taylor and participants in the II Workshop on Urban Economics in Barcelona for a number of helpful comments and suggestions.

1. Introduction

Over the last twenty years, study of the evolution of the wage distribution over time has become a major preoccupation of empirical economists. A widening of the wage distribution showing rising wage inequality in a number of countries has been very clearly documented in this work.¹ Empirical studies in this area have highlighted the temporal evolution of particular wage differentials linked to, for example, education or experience emphasising increase in the college wage premium or the return to experience that have gone hand-in-hand with rising wage inequality.

Despite there being a sizable urban economics literature studying the urban wage premium², work studying the spatial dimensions of rising wage inequality is more sparse.³

In part, this is because within/between type decompositions show that a significant part of the increase in overall wage inequality, or in particular wage differentials, has been within, rather than between, spatial units of observation like regions, states, cities or local labour markets.

Nonetheless, at a given point in time there are very sizable spatial differences in wages and in wage differentials between different groups of workers. In the past, specifically in the first few decades in the post-war period, these spatial differences tended to show persistence through time, with if anything there being regional and spatial convergence in wage or income differences. It is interesting to note that, in the period since wage inequality started to rise in the US (since the mid-to-late 1970s), this pattern

¹ See Katz and Autor (1999) or Acemoglu and Autor (2010) for reviews of the large literature in labour economics and Hornstein et al. (2005) for a review of the work in macroeconomics.

² See Puga (2010) or Rosenthal and Strange (2004) for discussions of the literature on urban wage premia and how they relate to agglomeration effects that raise productivity in cities.

³ Although less concerned with inequality rises over time, see the recent work on spatial wage differences and sorting (e.g. Combes, Duranton and Gobillon, 2008 or Baum-Snow and Pavan, 2012) and on local wage and skill distributions (Combes et al., 2012).

seems to have stopped. Since then mean reversion in spatial wage differences is absent as wages have tended to rise faster in places with higher initial wages. Moretti (2010), for example, shows plots of the wages of college graduates and high school graduates in 288 US Metropolitan Statistical Areas (MSAs) in 1980 and 2000 where wages grow faster in MSAs with higher wage levels in 1980 for both groups of workers. We find the same using data between 1980 and 2010 for 216 MSAs, as demonstrated in Figure 1.⁴ Figure 1 very much shows there to be faster increasing higher wage levels for college and high school workers between 1980 and 2010 in MSAs where wages were already higher in 1980.

In this paper, our interest is in relative wage differentials between college and high school workers. We study changing patterns of spatial college wage premia in the context of changing relative supply and demand of college educated versus high school educated workers. We begin by documenting the nature of changes in education employment shares and the college wage premium across different spatial units, looking at their evolution over time at state and MSA level. To do so, we use US Census and American Community Survey (ACS) data from 1980 through 2010. We uncover an interesting spatial dimension where, despite very rapid increases in the supply of college workers, the college wage premia has risen almost everywhere, but to varying degrees as the spatial variation in the wage gap between college educated and high school educated workers has become more persistent over time.

In the wage inequality literature, rising wage gaps between college and high school workers have been connected to shifts in the relative demand and supply of these groups of workers. Indeed, aggregate evidence shows that a key driver of rising wage college

⁴ The Figure is based on the 5 percent 1980, 1990 and 2000 Censuses and the 1 percent 2010 ACS which we collapse to 216 consistently defined MSAs. The Figure replicates Moretti's (2010) Figure based on 1980 and 2000 Census data. Moretti (2010) reported slope coefficients (and associated standard errors) of 1.82 (0.89) for high school graduates and 3.54 (0.11) for college graduates.

premium has been an increased relative demand for college educated workers (see Katz and Murphy, 1992; Katz and Autor, 1999; Acemoglu and Angrist, 2010). The presence of rising spatial college wage premia at different rates in the face of rapidly rising supply also suggests there may be differential relative demand shifts occurring at the spatial level. We thus modify the commonly used relative demand and supply model to calculate the extent of spatial relative demand shifts and examine the variations in their evolution through time. We also consider what factors have been correlated with the observed spatial shifts in relative demand, exploring whether technology measures (like R&D intensity or computer usage) and the reduced importance of labour market institutions (through union decline) display spatial correlations with changes in relative demand.

Previewing our key results, we report evidence of significant spatial variations in the college wage premium for US states and MSAs, and show that the pattern of shifts through time has resulted in increased spatial persistence. Because relative supply of college versus high school educated has also risen faster at the spatial level in places with higher initial supply levels, we also report a strong persistence in spatial relative demand. These relative demand increases are bigger in more technologically advanced states that have experienced faster increases in R&D intensity and computer usage, and in states where union decline has been fastest. Finally, we report evidence that the relationship between cross-state migration and college education has stayed relatively constant between 1980 and 2010.

The rest of the paper is structured as follows. Section 2 offers a descriptive analysis of changes in college shares in employment and changing college/high school wage differentials in the US at different levels of spatial aggregation. Section 3 then considers these differentials in the context of a relative supply-demand framework, showing spatial variations in the nature of relative demand shifts over time. Section 4

considers the relationship between state level relative demand shifts and some potential drivers of the shifts. Section 5 studies cross-state migration differences by education, whilst Section 6 concludes.

2. Spatial Employment Shares and Wage Differentials - Descriptive Analysis

To investigate spatial changes in the college wage premium we use data from the US Census in 1980, 1990 and 2000, and the American Community Survey (ACS) of 2010. We use the 5 percent Census samples and the 1 percent ACS to study the evolution of wage differentials for the 48 contiguous US states (dropping Alaska, Hawaii and the District of Columbia) and for 216 consistently defined MSAs. We focus on US born individuals aged 26-55 throughout our analysis.⁵

It has been widely documented that the employment shares of more educated workers have increased over time in the US (see, *inter alia*, Acemoglu and Autor, 2010). As with much of the work studying these changes, we consider changes in the relative employment of two composite education groups, 'college equivalent' and 'high school equivalent' workers. To form these composite groups, we first define five education groups, namely high school drop outs (with less than twelve years of schooling), high school graduates (with exactly twelve years of schooling), those with some college (thirteen to fifteen years of schooling), college graduates (sixteen years of schooling) and postgraduates (with over sixteen years of schooling). The college equivalent group then comprises college graduates plus postgraduates and 30 percent of the some college group (both weighted by their wage relative to college graduates).⁶ High school equivalent

⁵ See the Data Appendix for more detail on the data used throughout the paper.

⁶ The wage weights used are the average wage of the respective education group over all time periods. 30 percent of the some college group are assigned to the college equivalent group and 70 percent to the high school equivalent group because the some college wage is closer to the high school wage than to the college wage.

workers are defined analogously as high school graduates or high school dropouts plus 70 percent of some college workers (with the high school dropouts and some college workers having efficiency weights defined as their wage relative to high school graduates).

Table 1 shows the average college equivalent hours share between 1980 and 2010 for the 48 states and 216 MSAs. It reports that, on average across the 48 states we look at, the hours share of college equivalents rose by 11.6 percentage points between 1980 and 2010, going from a share of 29 percent in 1980 to just under 41 percent by 2010. The comparable average increase is much the same in the MSAs, at 9.7 percentage points (going from 30 to 40 percent).

However, what this Table of means does not show is the sizable spatial disparities in this relative education supply variable. For example, for states, the lowest hours share of college equivalents is 22 percent in Wyoming in 1980 and the highest is 55 percent in Massachusetts in 2010. Figures 2a and 2b therefore plot the state and MSA values of the college equivalent hours shares for the three periods 1980-90, 1990-2000 and 2000-2010. The same scale is utilised (from 0.2 to 0.55 for states, 0.15 to 0.65 for MSAs) so as to clearly show how the shares have moved through time over the three ten year intervals.

The pattern in the Figures is quite striking. First of all, in all time periods there is a strong persistence in rankings of high and low college equivalent states/MSAs.^{7 8} Second, there is evidence of relative supply increases in all states through time, as the scatterplot moves in a North Eastern direction when moving from the 1980-90 plot at the top, to the 2000-2010 plot at the bottom. But this movement is to varying degrees in different states

⁷ Spearman rank correlation coefficients are strongly significant for all three time periods. For states, they are: 0.94 (p-value = 0.00) for 1980-90; 0.97 (p-value = 0.00) for 1990-2000; 0.94 (p-value = 0.00) for 2000-10. For MSAs, they are: 0.94 (p-value = 0.00) for 1980-90; 0.93 (p-value = 0.00) for 1990-200; 0.94 (p-value = 0.00) for 2000-10.

⁸ Our focus is on relative education differences in labour supply and demand. Other papers show that broader sets of skills are concentrated in particular cities. Some recent examples are Bacold, Blum and Strange (2009) who place a focus on the distribution of a range of cognitive and non-cognitive skills in US cities and Hendricks (2011) who considers city skill compositions looking at spatial complementarities between business services and the skill structure of employment.

and MSAs. For example, for states over the whole 1980-2010 time period, the smallest increase is in New Mexico with a 4 percentage points increase and the largest in Massachusetts, which rises by 18 percentage points. Thus, the spread widens: in 1980 the range was 16 percentage points, by 2010 it was 25. For MSAs, the spread rises from 36 to 45 percentage points between 1980 and 2010. Thirdly, the slopes on the Figures show that, if anything, the college hours shares are diverging over time (as the coefficient of above unity on the slopes in the Figures shows).

Thus, the relative supply of college workers has risen sharply, and with differential spatial evolutions. What about the spatial college wage premium? Our analysis of the Census/ACS data makes it evident that, at the same time as the hours shares of the college educated have risen, so have their relative wage differentials. Table 2 presents mean composition adjusted log weekly wage differentials for college graduates relative to high school graduates across states and MSAs.⁹ One can see the well known average pattern of increasing wage payoffs to college graduates as the college/high school log wage premium rises by 0.178 percentage points between 1980 and 2010 across states of residence and by 0.185 percentage points across MSAs.

Previous research by Black, Kolesnikova and Taylor (2009) has noted that, at a point in time (in their case the cross-sections from the 1980, 1990 and 2000 US Census), there are sizable spatial disparities in education related wage differentials. This is also the case for our analysis, both in terms of the yearly variations across spatial units (thus

⁹ These wages are composition adjusted on the basis of estimating log weekly wage equations for full time full year workers separately for each year, for three birth cohorts/ages and by gender in each year for the 48 states and 216 MSAs respectively. The equations include dummies for age and race. To derive the educational wage differentials, education dummies are included for postgraduates (more than 16 years of education), college only (16 years of education), some college (13 to 15 years of education) and high school dropouts (less than 12 years of education) relative to the omitted group of high school graduates (12 years of schooling). The college/high school graduate log wage differential is the estimated coefficient on the college only variable, which we weight across the sub-groups for which we estimate using the average share of hours worked for each of the groups over 1980 to 2010.

confirming the Black, Kolesnikova and Taylor findings) and in terms of the increase in the college premium.

This can be shown in the same way as for our earlier analysis of relative education supply, as Figures 3a and 3b show the spatial distributions of the college wage premium for the three sub-periods, 1980-90, 1990-2000 and 2000-2010. There is very clear evidence of wide variations in the college premium. The lowest college/high school wage differential is 0.21 log points in Utah in 1980 and the highest is 0.61 log points in New York in 2010. Within years there is a wide spread which widens over time, from 0.20 log points in 1980, reaching 0.26 log points by 2010. A comparison of the three Panels in the Figures also reveals, as was the case for the relative supplies, a significant North Eastern movement over time. Thus, the college wage premium rises in all states, mirroring the national pattern, but it goes up by more in some places. In terms of states, the smallest rise is in Delaware (at 0.07 log points) and the largest in Illinois (at 0.27 log points).

This upward movement in spatial wage differentials is also characterised by persistence over time, and a persistence that seems to get stronger. For the state level analysis in Figure 3a, the estimated slope over the ten year interval rises from 0.75 in 1980-90 to 0.99 in 1990-2010 and 0.90 in 2000-2010. Thus, in the first decade there is some evidence of catch up, or mean reversion in the wage premia, from the states that has lower wage premia to begin with. However, this alters in 1990-2000 and 2000-2010 where the slope steepens as the premia move in the North Eastern direction in the Figures and become insignificantly different from unity. The same qualitative pattern of a steepening slope also occurs in the MSA data in Figure 3b, although there is also more evidence of mean reversion at this more disaggregated level (possibly due to more noise because of smaller samples of individuals from the Census/ACS data at this level of aggregation).

This descriptive section of the paper has highlighted spatial disparities in education supply and in the college wage premium in the US over the last thirty years. In the remainder of the paper, we focus on reasons why these disparities are present and why they are persistent. In the next section we consider spatial demand and supply models that enable us to use these patterns of change in education supply and the college wage premium to calculate spatial shifts in relative demand.

3. Relative Supply and Demand Models

The spatial dimensions of a rising supply of more educated workers and simultaneously rising college wage premia suggests a need to consider how these empirical phenomena map into a supply-demand model of spatial labour markets. Consequently, we now draw upon the Katz and Murphy (1992) canonical model of relative supply and demand to see if there are differential relative demand shifts by state and MSA.

The starting point is a Constant Elasticity of Substitution production function where output in for state or MSA i in period t (Y_{it}) is produced by two education groups (E_{1it} and E_{2it}) with associated technical efficiency parameters (θ_{1it} and θ_{2it}) as follows:

$$Y_{it} = [\delta_{it} (\theta_{1it} E_{1it})^\rho + (1 - \delta_{it}) (\theta_{2it} E_{2it})^\rho]^{1/\rho} \quad (1)$$

where δ indexes the share of work activities allocated to education group 1 and $\rho = 1 - 1/\sigma_E$, where σ_E is the elasticity of substitution between the two education groups.

Equating wages to marginal products for each education group, taking logs and expressing as a ratio leads to a relative wage equation (or inverse relative demand function) of the form

$$\log\left(\frac{W_{1it}}{W_{2it}}\right) = \frac{1}{\sigma_E} \left[D_{it} - \log\left(\frac{E_{1it}}{E_{2it}}\right) \right] \quad (2)$$

where $D_{it} = \sigma_E \log\left(\frac{\delta_{it}}{1-\delta_{it}}\right) + (\sigma_E - 1) \log\left(\frac{\theta_{1it}}{\theta_{2it}}\right)$ is an index of relative demand shifts that depends upon the (skill-biased) technological change parameters and the reflects shifts in relative demand.

This equation therefore relates the relative wage to relative demand and supply factors, and this is why the approach is sometimes framed as a race between supply and demand.¹⁰ A critical parameter determining the extent to which increases in relative supply affect relative wages is the elasticity of substitution between the education groups of interest, σ_E . A by now quite large literature has, in various ways, attempted to estimate σ_E .¹¹ For our purposes, we would like an estimate of σ_E at the spatial level, so that we can construct a measure of implied relative demand at the spatial level by rearranging equation (2) as:

$$D_{it} = \log\left(\frac{E_{1it}}{E_{2it}}\right) + \sigma_E \log\left(\frac{W_{1it}}{W_{2it}}\right) \quad (3)$$

where spatial relative demand is the relative supply plus the product of the elasticity of substitution and the relative wage.

There are two main routes to obtaining an estimate of σ_E which will enable us to put together the patterns of spatial college wage premia and spatial relative education supplies we described in Section 2 of the paper to form this index of spatial relative demand. First, we could use estimates that exist. However, there are only a few at state level as most estimates are at the aggregate level. Moreover, the ones that do exist do not match our samples and time period of study. Thus, we decided to follow the second route and estimate σ_E ourselves. To check robustness, we do also benchmark our estimates to

¹⁰ This dates back to Tinbergen (1974).

¹¹ For the traditional labour demand work, see Hamermesh (1993). For the wage inequality research, see Acemoglu and Autor (2010).

other state level estimates of the substitution elasticity (albeit from different samples as in Ciccone and Peri, 2005, or Fortin, 2006).

Probably the key issue estimating σ_E , and particularly at the sub-national level as we wish to, are possible biases emerging from geographical migration or because of potential endogeneity. In addition, when estimating at the spatial level there may be issues of measurement error that can cause attenuation bias. Consequently, we adopt a Two Stage Least Squares (2SLS) approach. To instrument relative labour supply, we draw upon the paper by Ciccone and Peri (2005) who use data from Acemoglu and Angrist (2000) to show that changes in state level compulsory attendance laws are correlated with the labour supply of high school graduates relative to high school drop outs. For our purposes it was necessary to update the data used in Acemoglu and Angrist (2000) to include state level compulsory attendance laws up to 2002.¹² We then use these laws as instruments for changes in relative supply at the state and MSA level.¹³

The instruments are set up as five dichotomous variables associated with each individual in the Census/ACS sample and then aggregated to the appropriate spatial unit in each year. The dummies CA8, CA9, CA10, CA11 and CA12 are equal to 1, and all other compulsory attendance law dummies are equal to 0, if the state where the individual is likely to have lived when aged 14 had compulsory attendance laws imposing a minimum of 8, 9, 10, 11 and 12 plus years of schooling. The five dummies are used to calculate the share of individuals for whom each of the five dummies is equal to 1 in each state and MSA of residence. We omit the share for CA8 (as the variables add up to 1) and these are used as instruments for the relative supply of college graduates. The data do not include

¹² The Acemoglu and Angrist (2000) data run from 1915 to 1978 and are available from the authors on request. We updated these data by obtaining compulsory schooling data up to 2002 from the digest of education statistics, supplemented by looking up the laws themselves in the state statutes. We cross checked our new data with that derived in Oreopoulos (2009) and found we had very similar measures.

¹³ Ciccone and Peri (2005) adopt a similar instrumentation approach when they estimate state level demand and supply models for high school graduates relative to high school drop outs using 1950-1990 Census data.

precise information on where individuals lived when aged 14, we therefore assume (as do others who use state compulsory school leaving laws as instruments in various contexts) that at age 14 individuals still lived in the state where they were born.

On a practical level, to be able to estimate equation (2) we need to model the demand shift term in some way. We specify that D_{it} is a function of state fixed effects and time so that $D_{it} = a_i + f(t) + e_{it}$ where $f(t)$ is a function of time (e.g. proxied by a time trend in the economy wide approaches of Katz and Murphy, 1992, Autor, Katz and Kearney, 2008, and Card and Lemieux, 2001), a_i are state level fixed effects which are captured using state/MSA level dichotomous variables and e_{it} is an error term. Thus the estimating equation becomes:

$$\log \left(\frac{W_{1it}}{W_{2it}} \right) = a_i + f(t) + \gamma \log \left(\frac{E_{1it}}{E_{2it}} \right) + e_{it} \quad (4)$$

where $\gamma = -1/\sigma_E$.

We specify the $f(t)$ function in its most general way, using a full set of time dummy variables, so that the estimating equation expresses the relative wage as a function of a time, state/MSA fixed effects and relative supply. As with our earlier descriptive analysis our focus is upon the college only/high school wage differential and we consider relative supply in terms of the definitions of college equivalent and high school equivalent workers introduced earlier.

Estimates of Relative Demand and Supply Models

Table 3 reports the estimates from the first stage regressions of the relative supply on the state of birth instruments. The estimates also include time and state/MSA fixed effects. These estimated coefficients are mostly negative and significant (relative to the omitted group CA8) showing relative boosts to high school equivalent supply (as would

be expected) from the raising of state compulsory school leaving ages). The F tests show the instruments to be significant, though they are not that strong leaving some possible weak instrument concerns. We deal with this issue below, by using our estimate of the elasticity of substitution from our 2SLS models, but also showing what happens if we use other estimates from the literature by making assumptions on the magnitude of σ_E in plausible bounds.

Table 4 provides the 2SLS estimates of equation (4). Again, these include time and geographical fixed effects. The estimate of the labour supply parameter γ is negative, as expected, in all cases. At state level, the 2SLS estimate is -0.396, which provides an estimated elasticity of substitution of 2.53. For the MSA level estimates, the estimate of γ is smaller (in absolute magnitude) with an estimated elasticity of substitution of 3.91. These estimates are in line with estimates in the aggregate literature, especially the state level estimate: for example, Lindley and Machin (2011) derive an estimate of around 2.6 using aggregate CPS data from 1963 to 2010 which is in the same ballpark as Autor, Katz and Kearney's (2008) estimate of 2.4, who also use the CPS data from 1963 to 2005.¹⁴

The estimated parameters on the time dummies in Table 4 also tell us something about the relative demand shifts that have occurred on a decade by decade basis. Relative to the 1980s the relative demand for college graduates has increased across all time periods although these incremental changes get smaller over time. This supports the idea of a quadratic relationship between relative labour demand and relative wages over the thirty years we study.¹⁵

Implied Relative Demand Shifts

¹⁴ Like Ciccone and Peri (2005) we also considered other estimation methods that may be robust to issues of potentially weak instruments. We used limited information maximum likelihood (LIML) estimation methods to produce similar estimates that were not statistically different to the 2SLS ones. For example, for our state level 2SLS estimate (and standard error) of $\gamma = -0.396$ (0.160) a comparable LIML estimate was $\gamma = -0.465$ (0.190), implying an elasticity of substitution of 2.15.

¹⁵ If a trend and trend squared were entered into the equation in place of the year dummies they confirm this.

We are now in a position to combine the spatial changes in wage differentials and supply into an implied relative demand index using our estimates of σ_E . As noted above, the demand index for spatial unit i in year t can be calculated as $D_{it} = \log(E_{1it}/E_{2it}) + \sigma_E \log(W_{1it}/W_{2it})$ for any two particular education groups. We construct the demand index based upon our relative supply measure of college equivalent (CE) versus high school equivalent (HE) workers and our composition adjusted college/high school wage differentials (W^C/W^H) as $D_{it} = \log(E_{it}^{CE}/E_{it}^{HE}) + \sigma_E \log(W_{it}^C/W_{it}^H)$.

Given that our estimates of relative demand depend on our elasticities of substitution (2.53 and 3.91 for state and MSA level analysis respectively), which in turn depend on the validity of our instruments, for robustness purposes we also bound our estimates by imposing two polar assumptions on the size of σ_E . Firstly, we assume that the elasticity of substitution σ_E is equal to unity (as for a Cobb-Douglas production function), which is just below the range of estimates in Ciccone and Peri's (2005) state level study.¹⁶ Second, we assume a larger (upper bound) with an elasticity of substitution equal to 5 (which is close to Fortin's, 2006, more recent study which focuses only on younger age cohorts).¹⁷

Table 5 compares the slopes for the different values of σ_E from regressions of the spatial relative demand shifts on their ten year lag, for the time periods 1980-90, 1990-2000 and 2000-2010 at both state and MSA level. The first row shows these for our estimated σ_E values and reveals that putting together the relative supply and relative wage measures to compute this demand index in this way produces a pattern of highly persistent relative demand shifts at the spatial level. The persistence also becomes more marked in

¹⁶ Ciccone and Peri (2005) present a range of estimates derived from different estimation approaches. Their Panels B and C of their Table 2 report estimates between 1.20 and 1.50 for data from 1950 to 1980.

¹⁷ Fortin (2006) presents state level estimates for age 26-35 workers between 1979 and 2002. Her 2SLS estimates from her Table 3 are in the range of 4.39 to 5.68.

the 1990-2000 and 2000-2010 period where, in statistical terms, the estimated persistence parameter is unity or above. This represents a shift from 1980-90, where there was also strong persistence, but also some convergence as the estimated coefficient on the 1980 level was below unity. The second row imposes the assumption $\sigma_E = 1$ and the third row that $\sigma_E = 5$. In both cases, for both states and MSAs, whilst the estimated parameters do shift a little, the same qualitative pattern of persistence remains.

To see more clearly what is going on, Figures 4a and 4b show the spatial distributions of the demand shift measure, using our estimated elasticities of substitution for states and MSAs (i.e. the first row of Table 5).¹⁸ These show that demand has shifted strongly in favour of the college educated. But these also allow us to eyeball which states and MSAs have increased their relative demand for college graduates the most (and the least) over the three decades we analyse. In Figure 4a we can see that the Eastern states like Massachusetts, Connecticut and New York have increased their relative demand the most and that this is consistent over time. More Southerly states like West Virginia, Wyoming and Nevada have experienced much smaller relative demand shifts.

Similarly, Figure 4b identifies two MSAs that have demonstrated relatively large and consistent increases in college graduate demand over time. These are 188 (Stamford, Connecticut) and 170 (San Jose, California). It is well known that Stamford has a large cluster of corporate headquarters for international companies¹⁹ (including banks like UBS and RBS), whilst San Jose is the largest city in Silicon Valley. MSAs that stand out as having relatively low demand shifts for college graduates (especially in the 1990s) are 107 (Lima, Ohio) and 66 (Flint, Michigan). Throughout the 1980s and 1990s Lima suffered

¹⁸ Figures A1a to A2b in the Appendix report the same Figures for demand shifts calculated under the assumption of $\sigma_E = 1$ and $\sigma_E = 5$ to very clearly show that the picture of increasing relative demand for college graduates is very robust to those derived using our estimated spatial elasticities of substitution shown in Figures 4a and 4b.

¹⁹ See David and Henderson (2005) for a discussion of the notion that places, like Stamford, generate significant agglomeration effects (including education and technology agglomeration).

economic decline as a consequence of many large company closures and Lima's plight and its subsequent efforts to re-define itself were captured in the PBS documentary *Lost in Middle America*. In a similar way to Lima, Flint is a large city that experienced severe economic decline but specifically this decline was in the automobile industry and in particular the closure of the General Motors headquarters. Flint's economic and social downfall has also been the subject a television documentary in *Roger & Me* by Michael Moore, as well as featuring in the movies *Bowling for Columbine* and *Fahrenheit 9/11*.

4. Potential Drivers of Implied Relative Demand Shifts

We can relate our estimated spatial relative demand shifts to potential demand side drivers of rising wage inequality that can be directly measured at the state level. We look at three different potential drivers of spatial relative demand shifts at state level that have been considered in some of the existing literature, but which are usually analysed at the aggregate or industry level. These are expenditures on R&D (measured relative to state GDP), computer use and union coverage. The latter two are measured in state level proportions.

Table 6 reports results from undertaking this exercise. The first three columns report the individual correlations between these potential drivers and relative demand shifts between 1980, 1990, 2000 and 2010, whilst the fourth column includes all three potential drivers together for a horse race between the three. The final two columns provide a robustness check for column 4 by assuming our two polar extremes for an elasticity of substitution equal to 1 and 5. All equations include state and year fixed effects.

Table 6 clearly shows that state level R&D intensity and computer use are positively correlated, whilst union coverage is negatively correlated with our implied

demand shifts. Entering all three together in one equation shows that only R&D intensity and union coverage are still statistically significant, although it is likely that R&D and computer use are to an extent multi-collinear. Hence, states with higher union coverage have also experienced lower shifts in relative demand for college graduates.

The final two columns show that although the negative union effect remains robust to the assumptions of the value for the elasticity of substitution between college and high school graduates, the horse race between R&D intensity and computer use is not. If one assumes that college graduates and high school graduates have a unit elasticity of substitution, then computer use is significantly correlated with the spatial demand shifts. If one assumes an elasticity of substitution of five, then it is R&D intensity (and not computer use) that is significantly correlated with implied demand shifts. Therefore, whilst it seems that the story that technology and union decline matter are robust, the precise form of the technology impact may be sensitive to the size of σ_E .

Figure 5 plots the long run 1980-2010 demand shifts against all three of our potential demand side drivers of inequality. Again, presenting these correlations in graphical form shows us which states are the most and least correlated with the proximate determinants. For example, Massachusetts demonstrates the largest long run increase in R&D, followed by Washington, Connecticut and New Jersey and these all show significant increases in relative demand. The interpretation for identifying the main states driving the changes in computer use is less obvious, mainly because of the mass implementation of general purpose computer technology (especially in more recent decades) which probably makes computers a less good proxy for technical change.

Notice in the plot of the demand shifts against change in the proportion of union covered workers there is union decline in all states. This reflects the overall long run decline in union coverage. However, some of the largest declines occurred in Michigan,

Indiana, Pennsylvania, Ohio, Illinois and West Virginia. These are states that have been much more affected by de-industrialisation, with sectoral shifts away from unionised large scale manufacturing firms and towards non-unionised service sector firms who are also likely to employ more graduates.

5. Cross-State Migration and College Education

The US is a highly mobile society and so it is possible that some of the reported results to date could be attributed to increased sorting of more educated individuals to potentially more prosperous states over time. In this Section we therefore consider cross-state migration differences by education and we study possible differences in relative supply and demand changes for individuals who remain in their state of birth as compared to those who move to another state.²⁰

We are able to consider cross-state migration since the Census/ACS data not only asks individuals to report their state of residence at the time of the survey, it also asks in which state that US born individuals were born.²¹ There is a lot of cross-state migration. In the 1980, 1990 and 2000 Census and 2010 ACS the proportion of 26-55 year old US born individuals respectively reported that 38 percent (in 1980, 1990 and 2000) and 37 percent (in 2010) resided in a different state from the one in which they were born. This proportion is significantly higher for college educated individuals as compared to high school educated individuals as is shown in Table 7.

For our purposes, it is noteworthy just how stable the college/high school differences in inter-state migration are between 1980 and 2010, suggesting little change in education gaps from spatial sorting. Controlling for age, race and gender, and for state of

²⁰ For early work on mobility and education see Ladinsky (1975) and the review of Greenwood (1975). More recent studies include Wozniak (2010) and Malamud and Wozniak (2011).

²¹ It also asks individuals where they lived five years before. We look at the longer run migration measure in this paper.

birth fixed effects, makes little difference to this, as is shown in the statistical models of cross-state migration reported in Table 8.

The other aspect of migration that is of relevance to our earlier analysis is whether migrants to a state operate in the same labour markets as those born in the state. Put differently, we would like to know whether migrants and indigenous individuals compete for the same jobs and can be thought of as perfect substitutes in production. If this were the case, then our earlier analysis of supply and demand which made no distinction between movers and stayers would be robust to this.²²

In the context of the CES production function in Section 3 (equation (1) above), we can add a second nest to the production function to allow for potential imperfect substitutability of movers and stayers within education groups. To do so we define CES

sub-aggregates for the two education groups as $E_{1it} = \left[\sum_j \beta_{1j} E_{1jit}^\eta \right]^{1/\eta}$ and $E_{2it} = \left[\sum_j \beta_{2j} E_{2jit}^\eta \right]^{1/\eta}$,

where j denotes movers/stayers and $\eta = 1 - 1/\sigma_M$, with σ_M being the elasticity of substitution between movers and stayers. If $\eta = 1$ (i.e. when σ_M is infinity due to perfect substitution) this collapses back to the standard model as there is no need to nest.

In an analogous manner to earlier, by deriving wages and setting them equal to marginal products, we obtain the following estimation equation, which is a generalised version of the earlier model we estimated at the level of mover/stayer j in state i in year t

$$\log\left(\frac{W_{1jit}}{W_{2jit}}\right) = a_i + a_j + f(t) + \gamma_1 \log\left(\frac{E_{1it}}{E_{2it}}\right) + \gamma_2 \left[\log\left(\frac{E_{1jit}}{E_{2jit}}\right) - \log\left(\frac{E_{1it}}{E_{2it}}\right) \right] + v_{jit} \quad (5)$$

²² The analysis of 1990 Census data in Dahl (2002) is suggestive that correcting college returns for self-selected migration does not make that much difference to the state-specific college wage premia, at least in the cross-section he considers.

where $\gamma_1 = -1/\sigma_E$, $\gamma_2 = -1/\sigma_M$ and v is an error term.²³ Thus a statistical test of whether $\gamma_2 = 0$ is a test of whether or not the movers and stayers are perfect substitutes within the college and high school groups of workers.

Estimates of equation (5) are reported in Table 8. The models are again estimated by 2SLS and show that we cannot reject the null hypothesis that $\gamma_2 = 0$. Thus an assumption that, at state level, the hypothesis that movers and stayers act as perfect substitutes is consistent with the data. This is reassuring as it offers confirmation of the robustness of our earlier findings.

6. Concluding Comments

We study spatial changes in the US college wage premium for states and cities using US Census and American Community Survey data between 1980 and 2010. We report evidence of significant spatial variations in the college wage premium for US states and MSAs. We use estimates of spatial relative demand and supply models to calculate implied relative demand shifts for college graduates vis-à-vis high school graduates. These also show significant spatial disparities. Considering potential drivers of the differential spatial trends, we show that relative demand has increased faster in those states and cities that have experienced faster increases in R&D intensity and computer use, and increased slower in those states and cities where union decline has been more marked.

These findings complement findings from the more aggregated work in labour economics on trends in wage inequality and on shifts in the relative demand and supply of more and less educated workers. They are also in line with the work in urban economics

²³ In practice, the equation from the two-level nested CES model is estimated as a two step procedure. First, the coefficient γ_1 can be estimated from regressions of the relative wages of movers and stayers to their relative supplies to derive a first estimate of σ_M and a set of efficiency parameters (the β_{1j} 's and β_{2j} 's in the CES sub-aggregates) can be obtained for each education group from a regression of wages on supply including mover/stayer dummy with spatial and year fixed effects. Given these, one can then compute E_{1t} and E_{2t} to obtain a model based estimate of aggregate supply. See Card and Lemieux (2001) for more detail.

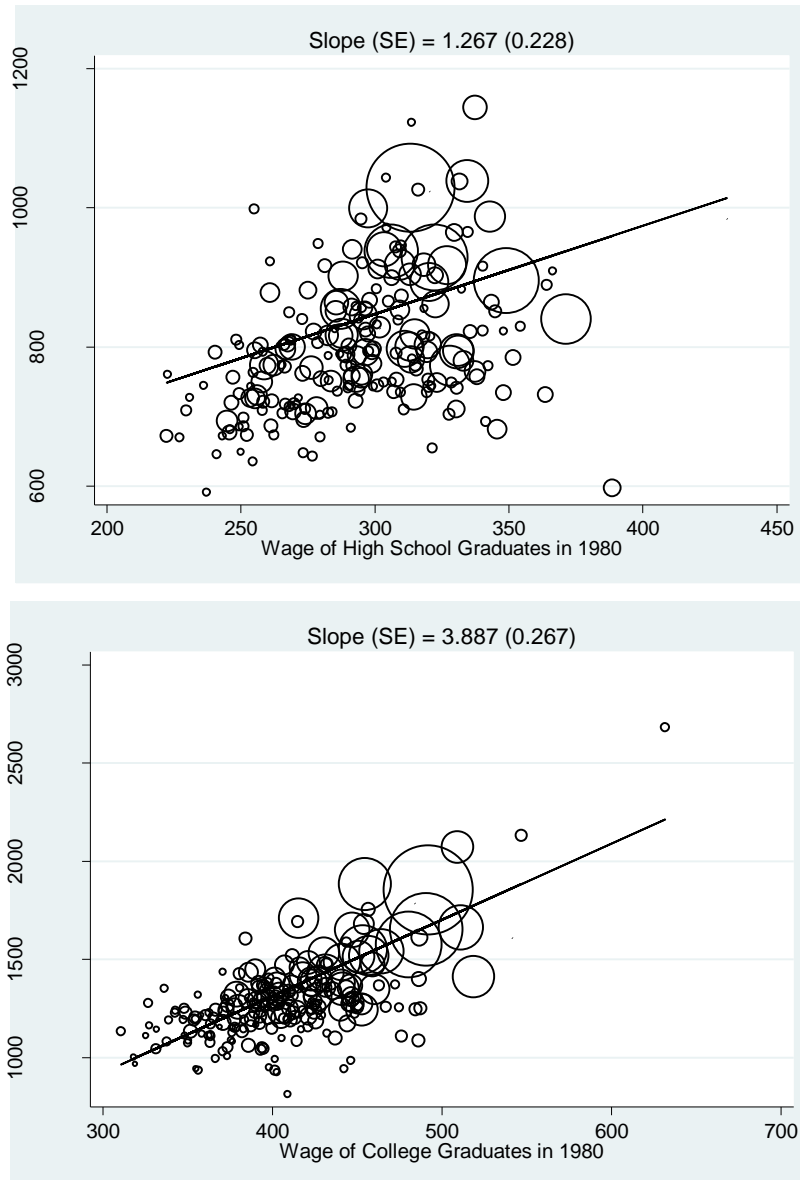
that emphasises agglomeration effects in locations that are strongly connected to education and technology. Our analysis brings these two areas of work together to an extent, by emphasising that the US has seen significant rises in the college wage premium despite rapid increases in education supply, and that there have been important spatial aspects to this and to the relative demand shifts by education that have occurred in the last thirty years.

References

- Acemoglu, D. and J. Angrist (2001) How Large are Human-Capital Externalities? Evidence from Compulsory Schooling Laws, NBER Macroeconomics Annual, 15, 9-59.
- Acemoglu, D. and D. Autor (2010) Skills, Tasks and Technologies: Implications for Employment and Earnings, in Ashenfelter, O. and D. Card (eds.) Handbook of Labor Economics Volume 4, Amsterdam: Elsevier.
- Autor, D., L. Katz, and M. Kearney (2008) Trends in U.S. Wage Inequality: Re-Assessing the Revisionists, Review of Economics and Statistics, 90, 300-323.
- Bacolod, M., B. Blum and W. Strange (2009) Skills and the City, Journal of Urban Economics, 67, 127-35.
- Baum-Snow, N. and R. Pavan (2012) Understanding the City Size Wage Gap, Review of Economic Studies, 79, 88-127.
- Black, D., N. Kolesnikova and L. Taylor (2009) Earnings Functions When Wages and Prices Vary by Location, Journal of Labor Economics, 27, 21-47.
- Card, D. and T. Lemieux (2001) Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis, Quarterly Journal of Economics, 116, 705-46.
- Ciccone, A. and G. Peri (2005) Long Run Substitutability Between More and Less Educated Workers: Evidence From US States 1950-1990, Review of Economics and Statistics, 87, 652-63.
- Combes, P., G. Duranton and L. Gobillon (2008) Spatial Wage Disparities: Sorting Matters!, Journal of Urban Economics, 63, 253-66.
- Combes, P., G. Duranton, L. Gobillon and S. Roux (2012) Sorting and Local Wage and Skill Distributions in France, IZA Discussion Paper No. 6501.
- Dahl, G. (2002) Mobility and the Return to Education: Testing a Roy Model with Multiple Markets, Econometrica, 70, 2367-2420.
- David, J. and V. Henderson (2005) The Agglomeration of Headquarters, Regional Science and Urban Economics, 38, 445-60.
- Fortin, N. (2006) Higher-Education Policies and the College Wage Premium: Cross-State Evidence from the 1990s, American Economic Review, 96, 959-87.
- Greenwood, M. (1975) Research on Internal Migration in the United States: A Survey, Journal of Economic Literature, 13, 397-443.
- Hamermesh, D. (1993) Labor Demand, Princeton, NJ: Princeton University Press..

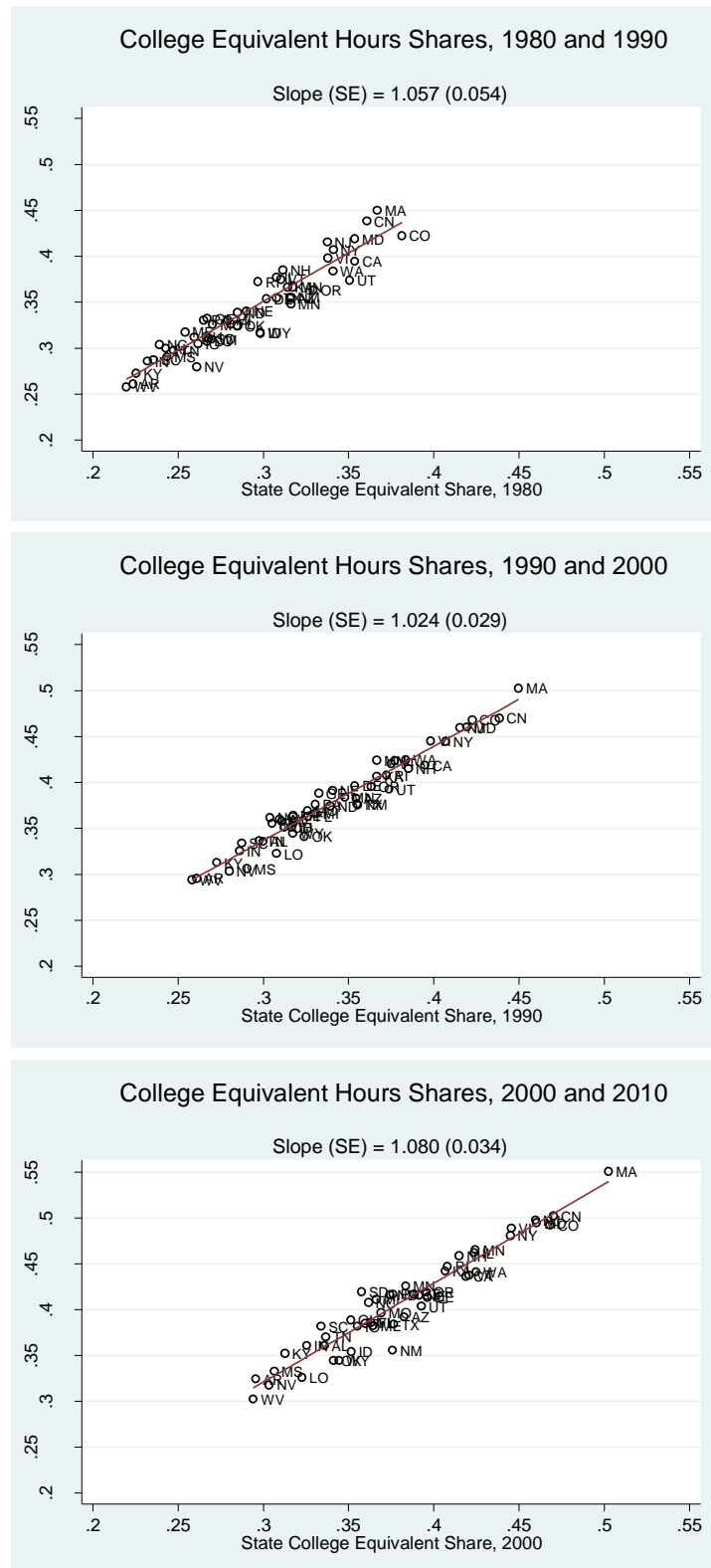
- Hendricks, L. (2011) The Skill Composition of US Cities, International Economic Review, 52, 1-32.
- Hirsch, B. and D. Macpherson (2003) Union Membership and Coverage Database from the Current Population Survey, Industrial and Labor Relations Review, 56, 349-54.
- Hornstein, A., P. Krusell and G. Violante (2005) The Effects of Technical Change on Labour Market Inequalities, in P. Aghion and P. Howitt (eds.) Handbook of Economic Growth, North Holland.
- Jaeger, D. (1997) Reconciling the Old and New Census Bureau Education Questions: Recommendations for Researchers, Journal of Business and Economic Statistics, 15, 300-309.
- Katz, L. and D. Autor (1999) Changes in the Wage Structure and Earnings Inequality, in O. Ashenfelter and D. Card (eds.) Handbook of Labor Economics, Volume 3, North Holland.
- Katz, L. and K. Murphy (1992) Changes in Relative Wages, 1963-87: Supply and Demand Factors, Quarterly Journal of Economics, 107, 35-78.
- Ladinsky, J. (1967) The Geographical Mobility of Professional and Technical Manpower, Journal of Human Resources, 2, 475-94.
- Lindley, J. and S. Machin (2011) Postgraduate Education and Rising Wage Inequality, IZA Discussion Paper No. 5981.
- Malamud, O. and A. Wozniak (2011) The Impact of College on Migration: Evidence from the Vietnam Generation, Journal of Human Resources, forthcoming.
- Moretti, E. (2010) Local Labor Markets, in Ashenfelter, O. and D. Card (eds.) Handbook of Labor Economics Volume 4, Amsterdam: Elsevier.
- Oreopoulos, P. (2009) Would More Compulsory Schooling Help Disadvantaged Youth? Evidence from Recent Changes to School- Leaving Laws, in J. Gruber (ed) The Problems of Disadvantaged Youth: An Economic Perspective, University of Chicago Press.
- Puga, D. (2010) The Magnitude and Causes of Agglomeration Economies, Journal of Regional Science, 50, 203-19.
- Rosenthal, S. and W. Strange (2004) Evidence on the Nature and Sources of Agglomeration Economies, in V. Henderson and J-F. Thisse (eds.) Handbook of Urban and Regional Economics, North Holland.
- Tinbergen, J. (1974) Substitution of Graduate by Other Labour, Kyklos, 27, 217-26.
- Wozniak, A. (2010) Are College Graduates More Responsive to Distant Labor Market Opportunities, Journal of Human Resources, 45, 944-70.

Figure 1:
Change Over Time in the Average Log Weekly Wage of High School and College Graduates by Metropolitan Area



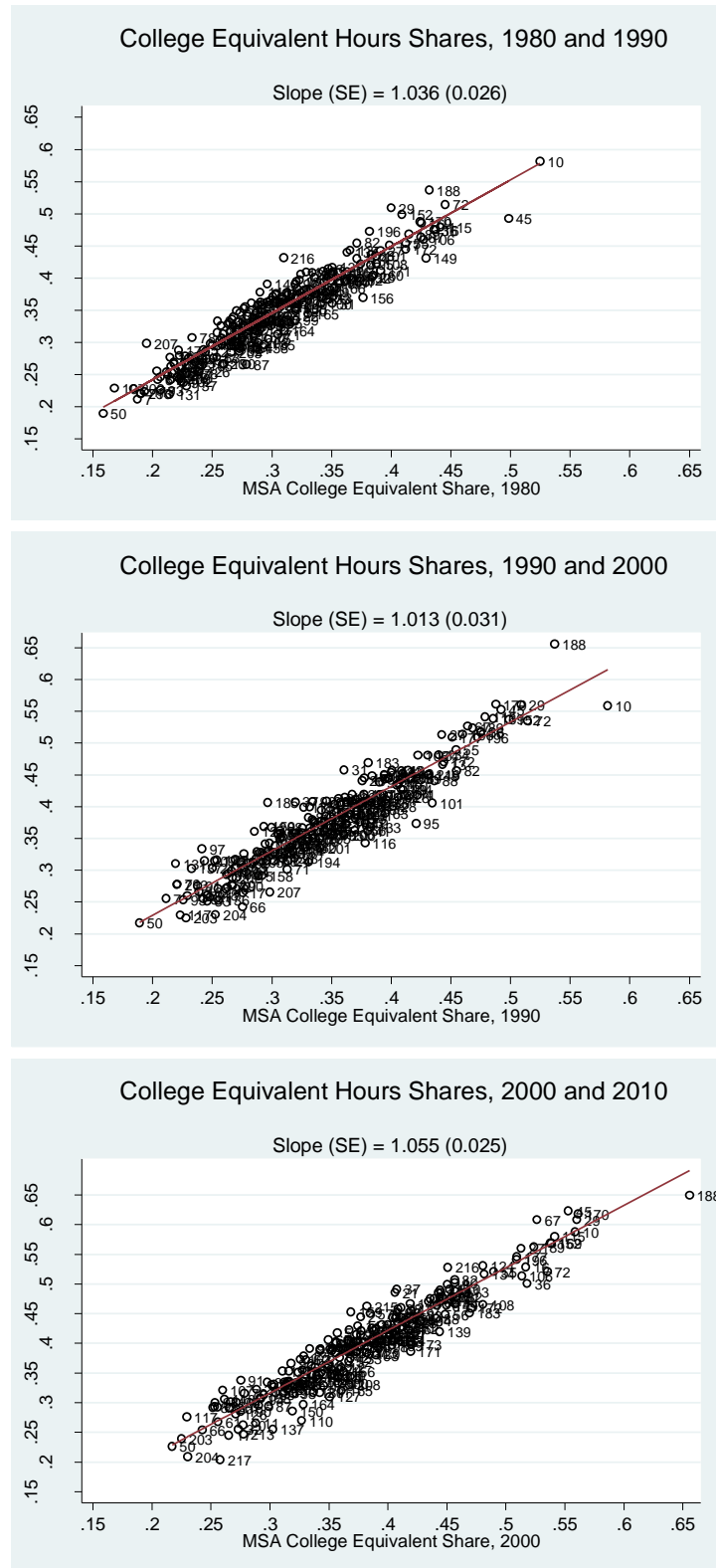
Notes: Each panel plots the nominal wage in 1980 against the nominal wage in 2010 by metropolitan area. The top panel is for high school graduates and the bottom panel is for college graduates. These are weighted using the number of workers in the relevant metropolitan area and skill group in 1980. There are 216 metropolitan areas. The regression line is the predicted log wage in 2010 from a weighted OLS regression. The slope is 1.267 (0.228) for high school graduates and 3.887 (0.267) for college graduates. Data are from the Census of Population. The sample includes all full time US born workers age between 26 and 55 who worked at least 40 weeks in the previous year.

Figure 2a:
State Level College Equivalent Hours Shares, 1980 to 2010



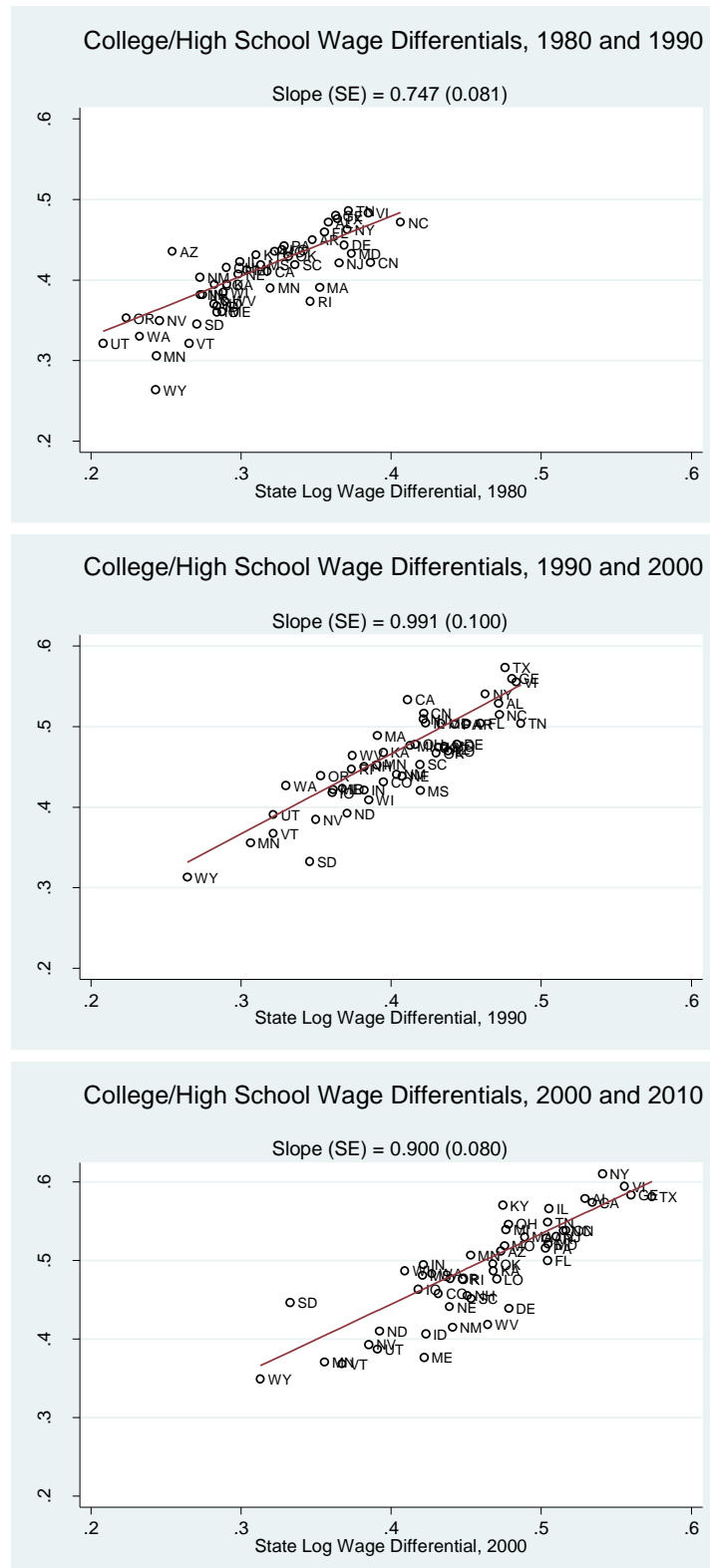
Notes: These are college equivalent hours shares for workers aged 26-55 in 48 states in the 1980, 1990 and 2000 Census (where wages refer to the previous calendar years, 1979, 1989 and 1999 respectively) and the 2010 American Community Survey. For definitions of college and high school equivalent see the main text and the Data Appendix. Standard errors in parentheses for the reported slope coefficients.

Figure 2b:
MSA Level College Equivalent Hours Shares, 1980 to 2010



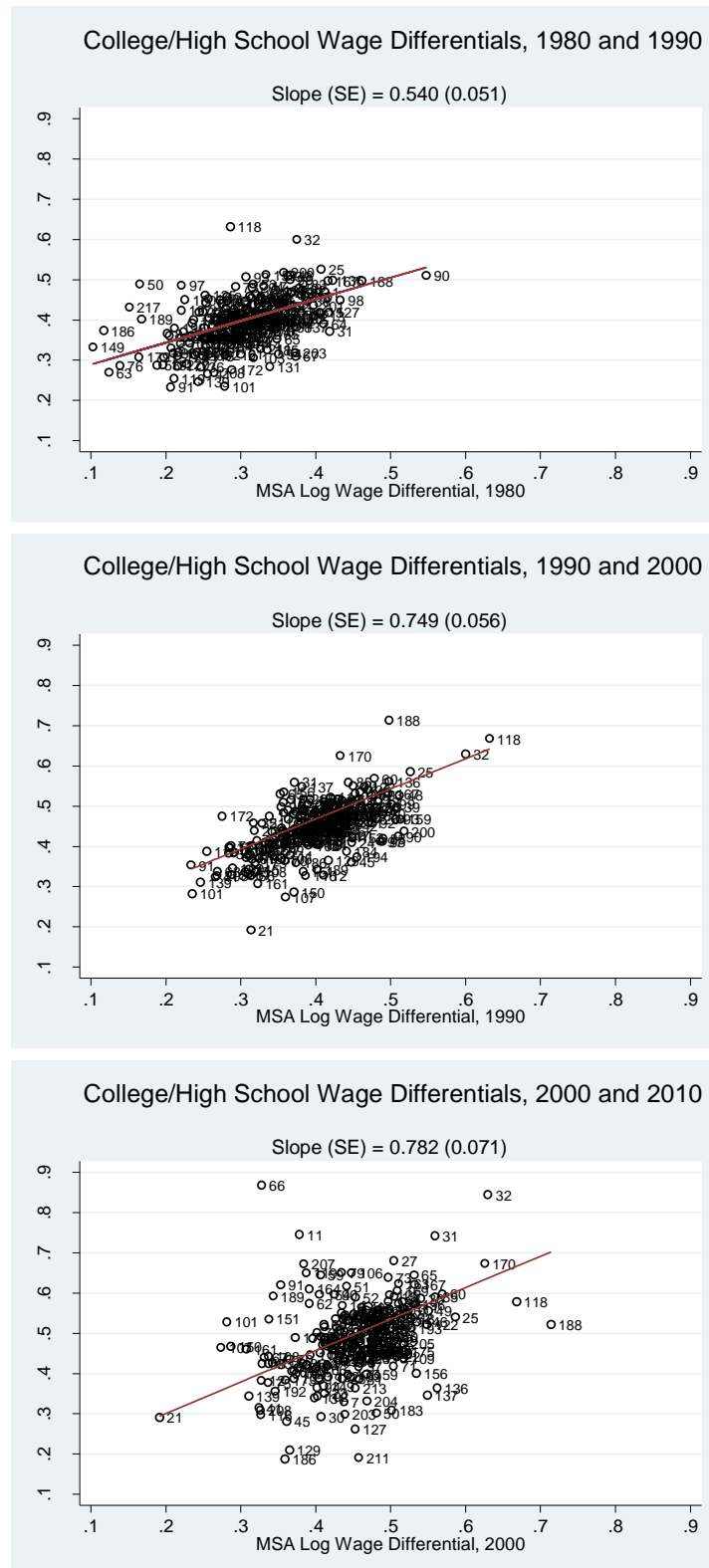
Notes: These are college equivalent hours shares for workers aged 26-55 in 219 MSAs in the 1980, 1990 and 2000 Census (where wages refer to the previous calendar years, 1979, 1989 and 1999 respectively) and the 2010 American Community Survey. For definitions of college and high school equivalent see the main text and the Data Appendix. Standard errors in parentheses for the reported slope coefficients.

Figure 3a:
State Level College/High School Log Wage Differentials, 1980 to 2010



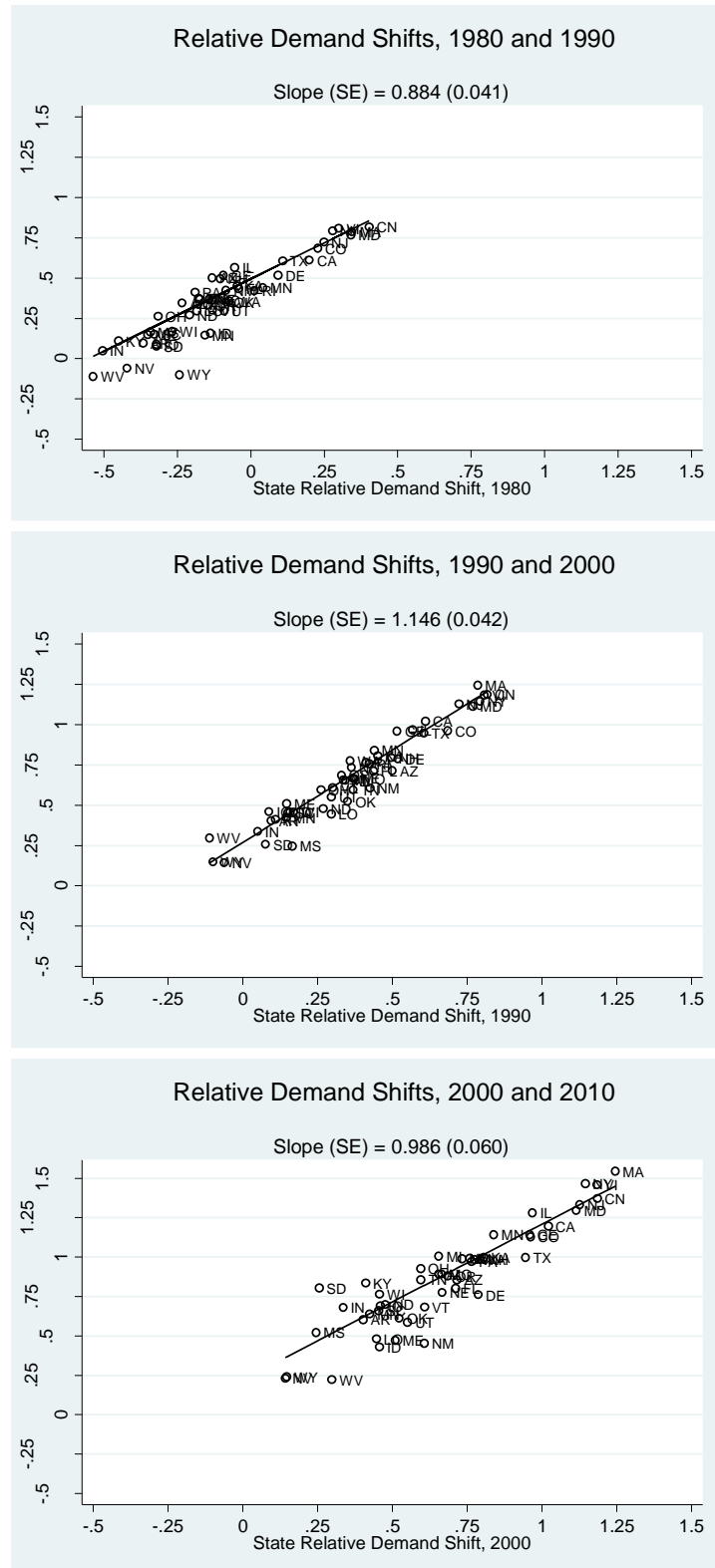
Notes: These are fixed hours weighted composition adjusted college only/high school log wage differentials for full time full year workers aged 26-55 in 48 states in the 1980, 1990 and 2000 Census (where wages refer to the previous calendar years, 1979, 1989 and 1999 respectively) and the 2010 American Community Survey. The composition adjustment is described in the main text and in the Data Appendix. The estimated slope coefficients (and associated standard errors) are weighted by the inverse sampling variance of the state level wage differentials.

Figure 3b:
MSA Level College/High School Log Wage Differentials, 1980 to 2010



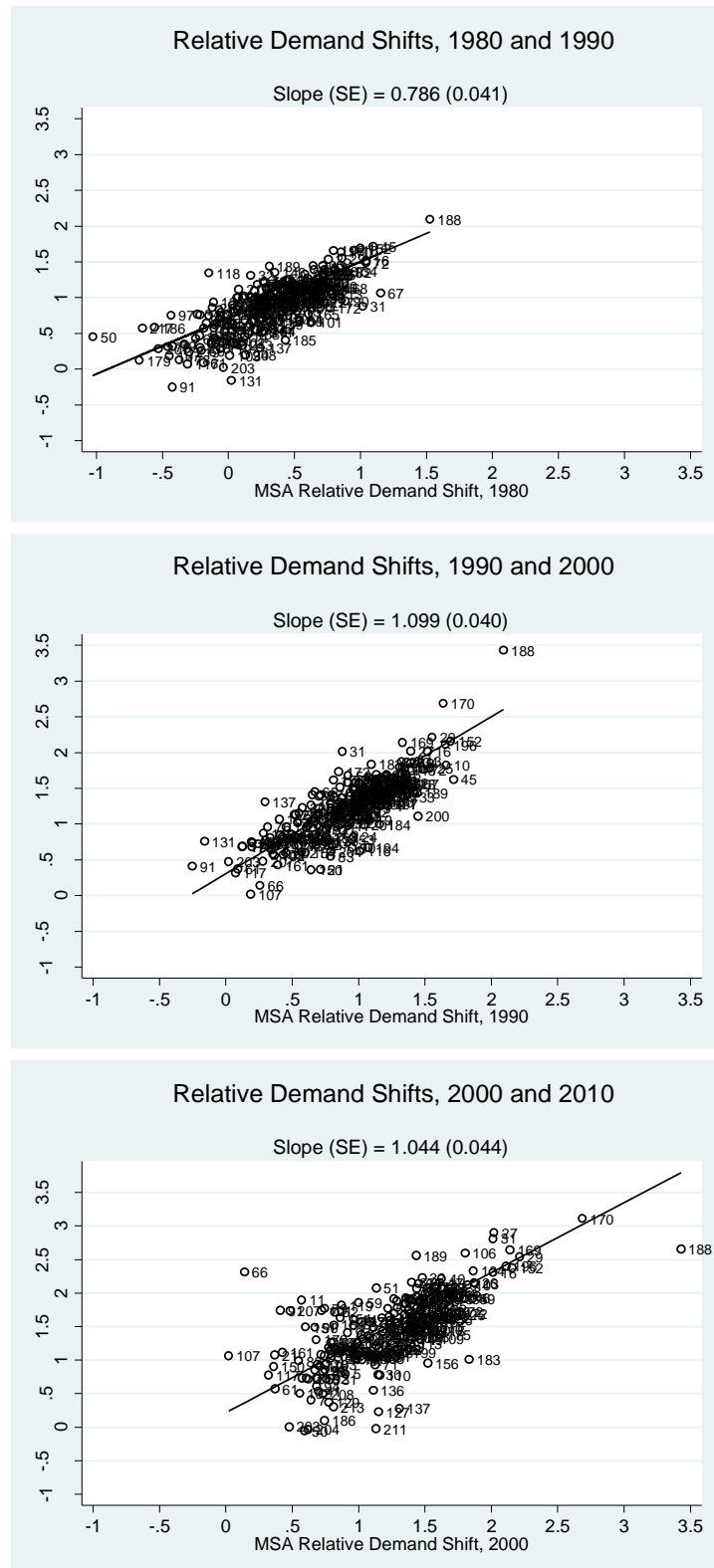
Notes: These are fixed hours weighted composition adjusted college only/high school log wage differentials for full time full year workers aged 26-55 in 219 MSAs in the 1980, 1990 and 2000 Census (where wages refer to the previous calendar years, 1979, 1989 and 1999 respectively) and the 2010 American Community Survey. The composition adjustment is described in the main text and in the Data Appendix. The estimated slope coefficients (and associated standard errors reported in parentheses) are weighted by the inverse sampling variance of the MSA level wage differentials.

Figure 4a:
Implied Relative Demand Shifts, State Level, 1980 to 2010



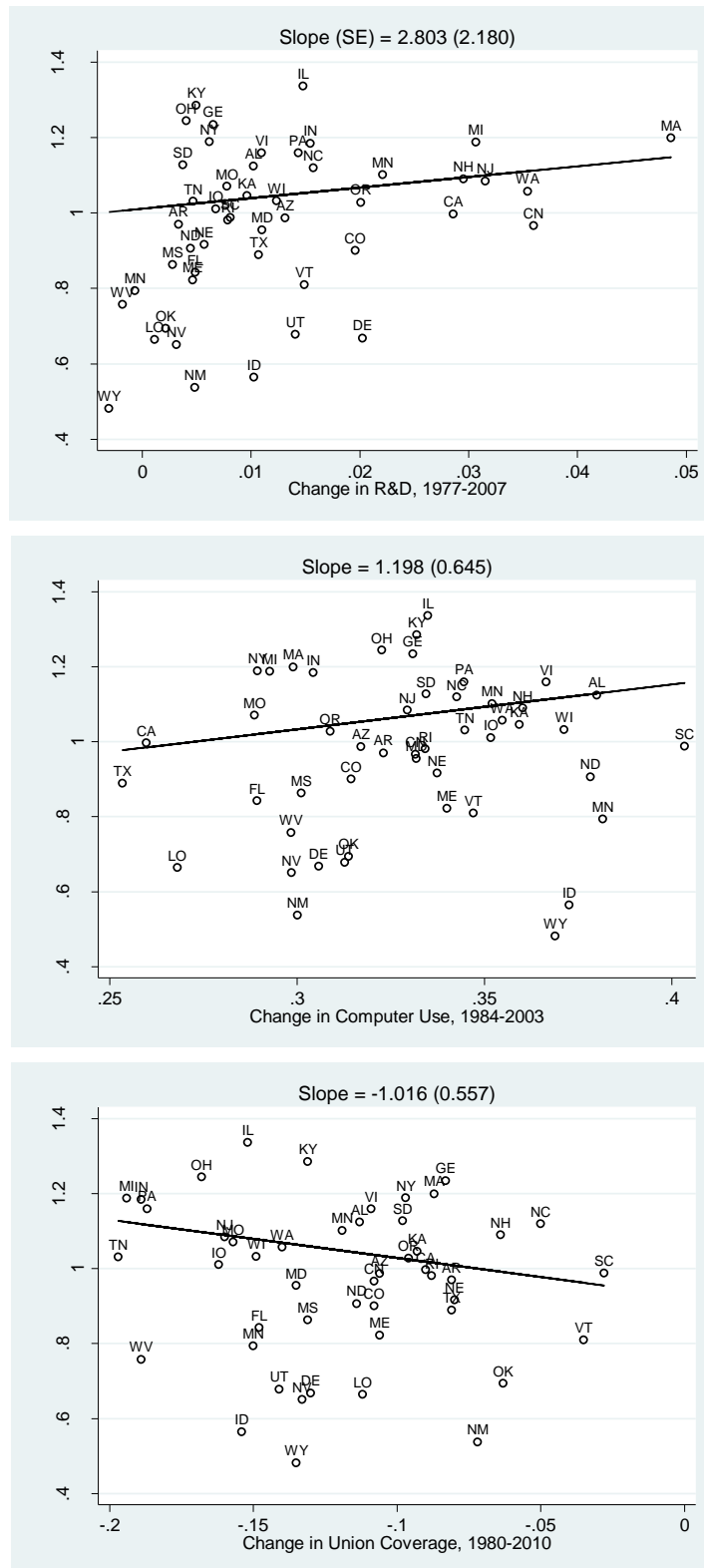
Notes: The relative demand shifts are calculated as $\log(L^{CE}/L^{HE}) + \sigma_E \log(W^C/W^H)$, where $\log(L^{CE}/L^{HE})$ is the log relative supply of college equivalent versus high school equivalent hours, $\sigma_E (= 2.53)$ is the elasticity of substitution between college and high school workers and $\log(W^C/W^H)$ is the fixed weighted composition adjusted college/high school log wage differential. The estimated slope coefficient (and associated standard error reported in parentheses) are weighted by the inverse sampling variance of the state level wage differentials.

Figure 4b:
Implied Relative Demand Shifts, MSA Level, 1980 to 2010



Notes: The relative demand shifts are calculated as $\log(L^{CE}/L^{HE}) + \sigma_E \log(W^C/W^H)$, where $\log(L^{CE}/L^{HE})$ is the log relative supply of college equivalent versus high school equivalent hours, $\sigma_E (= 3.91)$ is the elasticity of substitution between college and high school workers and $\log(W^C/W^H)$ is the fixed weighted composition adjusted college/high school log wage differential. The estimated slope coefficient (and associated standard error) are weighted by the inverse sampling variance of the MSA level wage differentials.

Figure 5:
State Level Relative Demand Shifts and Changes in R&D Intensity



Notes: The estimated slope coefficient (and associated standard error reported in parentheses) are weighted by the inverse sampling variance of the state level wage differentials.

**Table 1:
Average State/MSA Hours Shares of College Equivalent Workers - 1980, 1990 and 2000 Census and 2010 ACS**

	Mean College Equivalent Hours Share	
	State of Residence	MSA of Residence
1980	0.292	0.300
1990	0.343	0.346
2000	0.381	0.377
2010	0.408	0.397
Change 2010-1980	0.116 (0.010)	0.097 (0.008)

Notes: Hours shares are for all workers aged 26-55. To construct the shares, college equivalent workers are defined as college or college plus workers plus 30 percent of some college workers (where the college plus and some college workers have efficiency weights defined as their wage relative to college graduates). High school equivalent workers are defined analogously as high school graduates or high school dropouts plus 70 percent of some college workers (with the high school dropouts and some college workers having efficiency weights defined as their wage relative to high school graduates). The college equivalent hours share is then hours of college equivalent workers divided by the sum of hours of college equivalent and high school equivalent workers. See the Data Appendix for more detail.

Table 2:
Composition Adjusted College Wage Premia - 1980, 1990 and 2000 Census and 2010 ACS

Mean College/High School Log Wage Differential		
	State of Residence	MSA of Residence
1980	0.310	0.299
1990	0.404	0.395
2000	0.460	0.447
2010	0.488	0.484
Change 2010-1980	0.178 (0.012)	0.185 (0.008)

Notes: The composition adjusted state of birth college plus/high school log wage differential are derived from estimated log wage equations estimated separately for each year, age group (3) and gender for 48 states and 216 MSAs respectively (i.e. six equations per year for each state/MSA). The equations include dummies for age and race. Three education dummies are included for college plus (16 or more years of education), some college (13 to 15 years of education) and high school dropouts (less than 12 years of education) relative to the omitted group of high school graduates (12 years of schooling). The college graduate/high school graduate log wage differential is the estimated coefficient on the college graduate variable. The wage sample consists of US born full time full year workers age 26-55. For the change 2010-1980 standard errors are in parentheses.

Table 3:
First Stage 2SLS Regressions - 1980, 1990 and 2000 Census and 2010 ACS
(Relative Supply = College Equivalents/Non-College Equivalents).

	State of Residence			MSA of Residence		
	Relative Supply	College Equivalents	Non-College Equivalents	Relative Supply	College Equivalents	Non-College Equivalents
CA9	-0.052 (0.091)	0.179 (0.114)	0.231 (0.138)	-0.065 (0.063)	-0.068 (0.100)	-0.003 (0.110)
CA10	-0.167 (0.097)	0.120 (0.122)	0.288 (0.147)	-0.143 (0.066)	-0.141 (0.104)	0.003 (0.115)
CA11	0.047 (0.173)	0.620 (0.218)	0.573 (0.263)	-0.100 (0.126)	0.361 (0.198)	0.461 (0.219)
CA12	-0.221 (0.128)	0.112 (0.161)	0.333 (0.194)	-0.177 (0.084)	-0.274 (0.133)	-0.097 (0.147)
F-Test	2.07	2.20	1.71	2.36	3.51	1.81
P-Value	0.08	0.07	0.15	0.05	0.01	0.12
Spatial Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	192	192	192	864	864	864

Notes: Where CA9 is the proportion of individuals residing in a geographical area with 9 years Compulsory Schooling Attendance (CSA) in the state in which they were born, when they were age 14. Similarly CA10 is the same but for CSA of ten years, CA11 for CSA of 11 years and CA12 for CSA of 12 years and over. College equivalents contain the hours of college graduates and half of the hours for some college. Non-college equivalents include the hours for high school drop outs, high school graduates and half of the hours for some college.

Table 4:
2SLS Estimates of Supply-Demand Models of College Plus/High School Wage Differentials - 1980, 1990 and 2000 Census and 2010 ACS
(Relative Supply = College Equivalents/Non-College Equivalents)

	State of Residence	MSA of Residence
Log(Relative Supply)	-0.396* (0.160)	-0.256* (0.155)
Year = 1990	0.198* (0.041)	0.159* (0.040)
Year = 2000	0.333* (0.067)	0.268* (0.063)
Year= 2010	0.416* (0.087)	0.340* (0.081)
Spatial Fixed Effects	Yes	Yes
Sample Size	192	864

Notes: The dependent variable is the log of the fixed weighted composition adjusted college/high school wage differential. State Compulsory School Leaving Laws used as instruments for Log(Relative Supply) - the first stages are in Table 3). Estimates are weighted by the inverse sampling variance of the state/MSA level wage differentials. Standard errors in parentheses.

Table 5:
Spatial Persistence in Implied Relative Demand Shifts For Different σ_E Estimates

	Estimates of ψ_1 from:					
	$D_{it} = \psi_0 + \psi_1 D_{i,t-10} + u_{it}$					
	States			MSAs		
	1980-1990	1990-2000	2000-2010	1980-1990	1990-2000	2000-2010
Estimated $\sigma_E = 2.53$ (State), $\sigma_E = 3.91$ (MSA)	0.884 (0.041)	1.146 (0.042)	0.986 (0.060)	0.786 (0.040)	1.099 (0.040)	1.044 (0.044)
$\sigma_E = 1$	0.932 (0.037)	1.039 (0.033)	1.103 (0.049)	0.977 (0.027)	1.056 (0.025)	1.077 (0.024)
$\sigma_E = 5$	0.819 (0.053)	1.207 (0.062)	0.953 (0.070)	0.725 (0.042)	1.078 (0.045)	1.020 (0.049)

Notes: The dependent variable is the implied relative demand shift $\log(L^{CE}/L^{HE}) + \sigma_E \log(W^C/W^H)$, where $\log(L^{CE}/L^{HE})$ is the log relative supply of college equivalent versus high school equivalent hours, σ_E is the elasticity of substitution between college and high school workers and $\log(W^C/W^H)$ is the fixed weighted composition adjusted college/high school log wage differential. Estimates are weighted by the inverse sampling variance of the state level wage differentials $\log(W^C/W^H)$. Standard errors in parentheses.

Table 6:
State Level Demand Shifts, Technological Change and Union Coverage

Implied Relative Demand Shifts, $\log(L^{CE}/L^{HE}) + \sigma_E \log(W^C/W^H)$, 1980, 1990 and 2000 Census and 2010 ACS						
	(1)	(2)	(3)	(4)	(5)	(6)
	Estimated $\sigma_E = 2.53$				$\sigma_E = 1$	$\sigma_E = 5$
R&D/GDP	2.251 (1.089)			2.172 (1.054)	1.053 (0.792)	3.983 (1.708)
Computer Usage		0.684 (0.322)		0.571 (0.314)	0.874 (0.236)	0.080 (0.509)
Union Coverage			-0.960 (0.342)	-0.874 (0.338)	-0.646 (0.253)	-1.244 (0.547)
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	192	192	192	192	192	192

Notes: The dependent variable is the state level implied relative demand shift $\log(L^{CE}/L^{HE}) + \sigma_E \log(W^C/W^H)$, where $\log(L^{CE}/L^{HE})$ is the log relative supply of college equivalent versus high school equivalent hours, σ_E is the elasticity of substitution between college and high school workers and $\log(W^C/W^H)$ is the fixed weighted composition adjusted college/high school log wage differential. Estimates are weighted by the inverse sampling variance of the state level wage differentials $\log(W^C/W^H)$. Standard errors in parentheses.

Table 7:
Cross-State Migration and Education in the Census and ACS Data, 1980 to 2010

	Proportion Moving State			
	All	College	High School	Gap (Standard Error)
1980	0.378	0.472	0.326	0.146 (0.001)
1990	0.379	0.477	0.306	0.171 (0.001)
2000	0.378	0.465	0.298	0.168 (0.001)
2010	0.366	0.303	0.139	0.139 (0.001)

Notes: Individuals aged 26-55. Sample sizes are: 1980 - 3797299; 1990 - 4556006; 2000 - 5036408; 948062.

Table 8:
Cross-State Migration and College Education, 1980 to 2010

	Pr(State of Residence \neq State of Birth)			
	1980	1990	2000	2010
A. Without State of Birth Fixed Effects				
College	0.159 (0.001)	0.177 (0.001)	0.168 (0.001)	0.141 (0.001)
Control Variables	Yes	Yes	Yes	Yes
State of Birth Fixed Effects	No	No	No	No
Sample Size	3797299	4556006	5036408	948062
B. With State of Birth Fixed Effects				
College	0.165 (0.001)	0.180 (0.001)	0.165 (0.001)	0.136 (0.001)
Control Variables	Yes	Yes	Yes	Yes
State of Birth Fixed Effects	Yes	Yes	Yes	Yes
Sample Size	3797299	4556006	5036408	948062

Notes: Individuals aged 26-55. These are linear probability estimates where the dependent variable is a 0-1 dummy coded to 1 for movers and 0 for stayers. The control variables are a full set of age dummies and dummies for gender and race. Standard errors in parentheses.

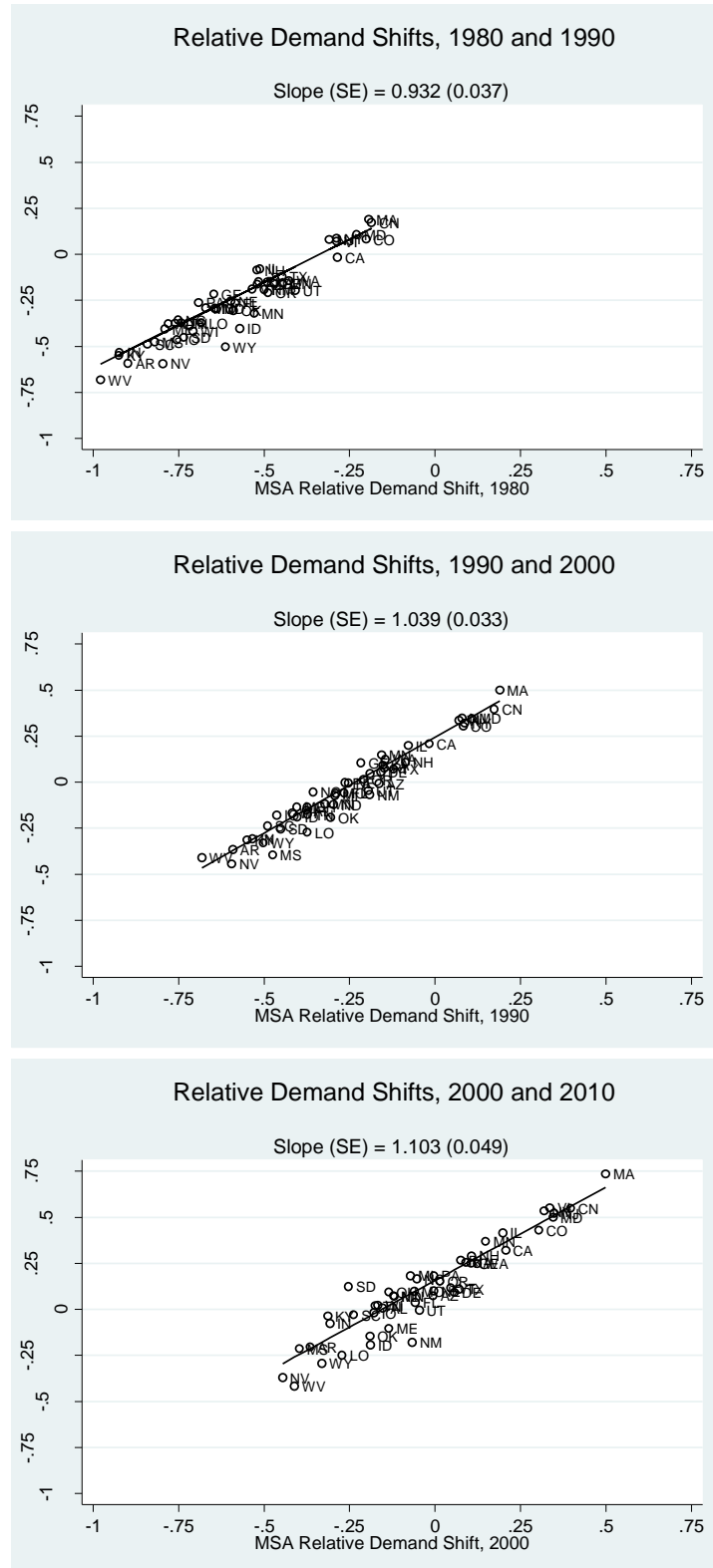
**Table 9:
Test of Substitutability of Movers and Stayers**

	State of Residence
Log(State Relative Supply)	-0.307 (0.138)
Log(Mover/Stayer State Relative Supply) - Log(State Relative Supply)	0.050 (0.034)
Mover	0.029 (0.018)
Year = 1990	0.176 (0.035)
Year = 2000	0.295 (0.508)
Year= 2010	0.367 (0.075)
State Fixed Effects	Yes
Sample Size	384

Notes: The dependent variable is the log of the fixed weighted composition adjusted college/high school wage differential. State Compulsory School Leaving Laws used as instruments for Log(Relative Supply) - the first stages are in Table 3). Estimates are weighted by the inverse sampling variance of the state/MSA level wage differentials. Standard errors in parentheses.

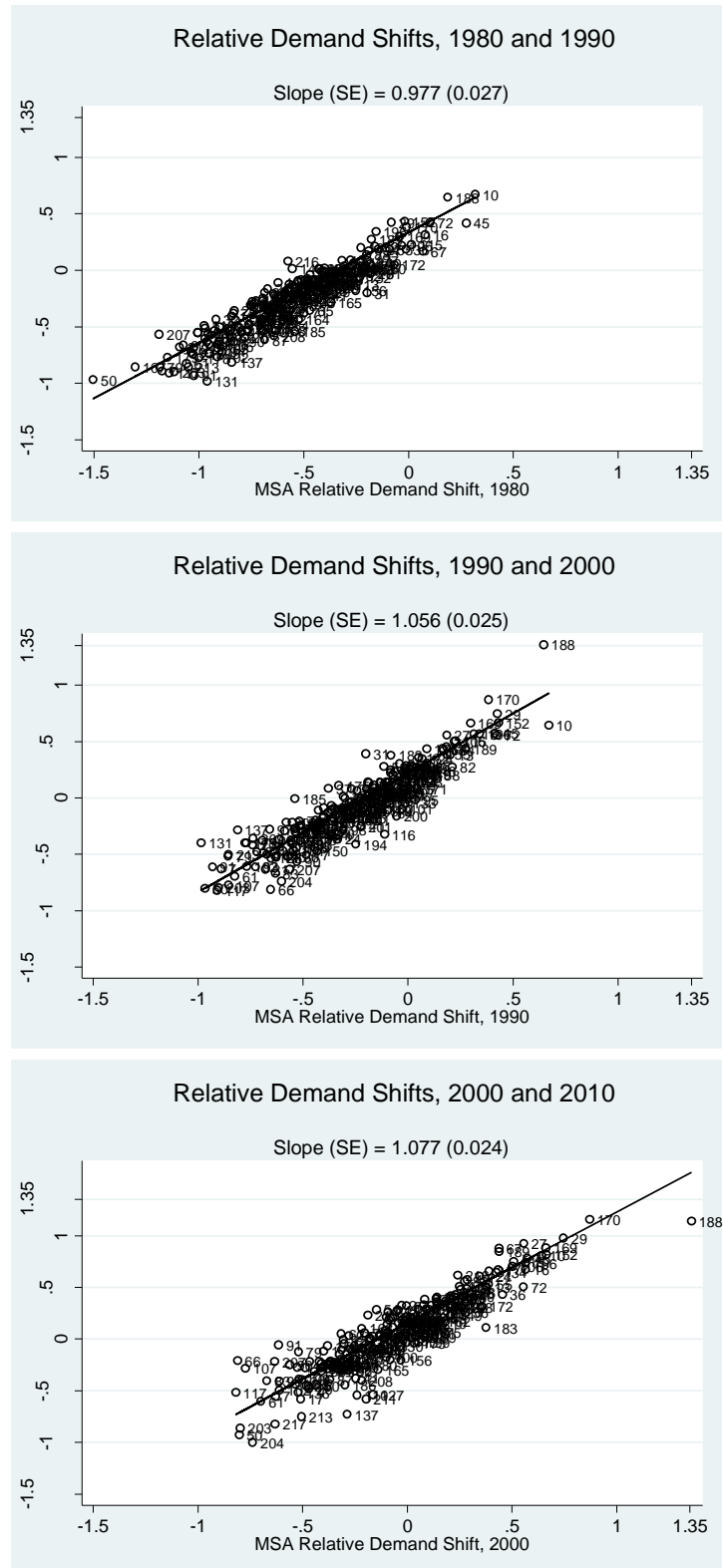
Appendix

Figure A1a: Implied Relative Demand Shifts by State, 1980 to 2010 ($\sigma_E = 1$)



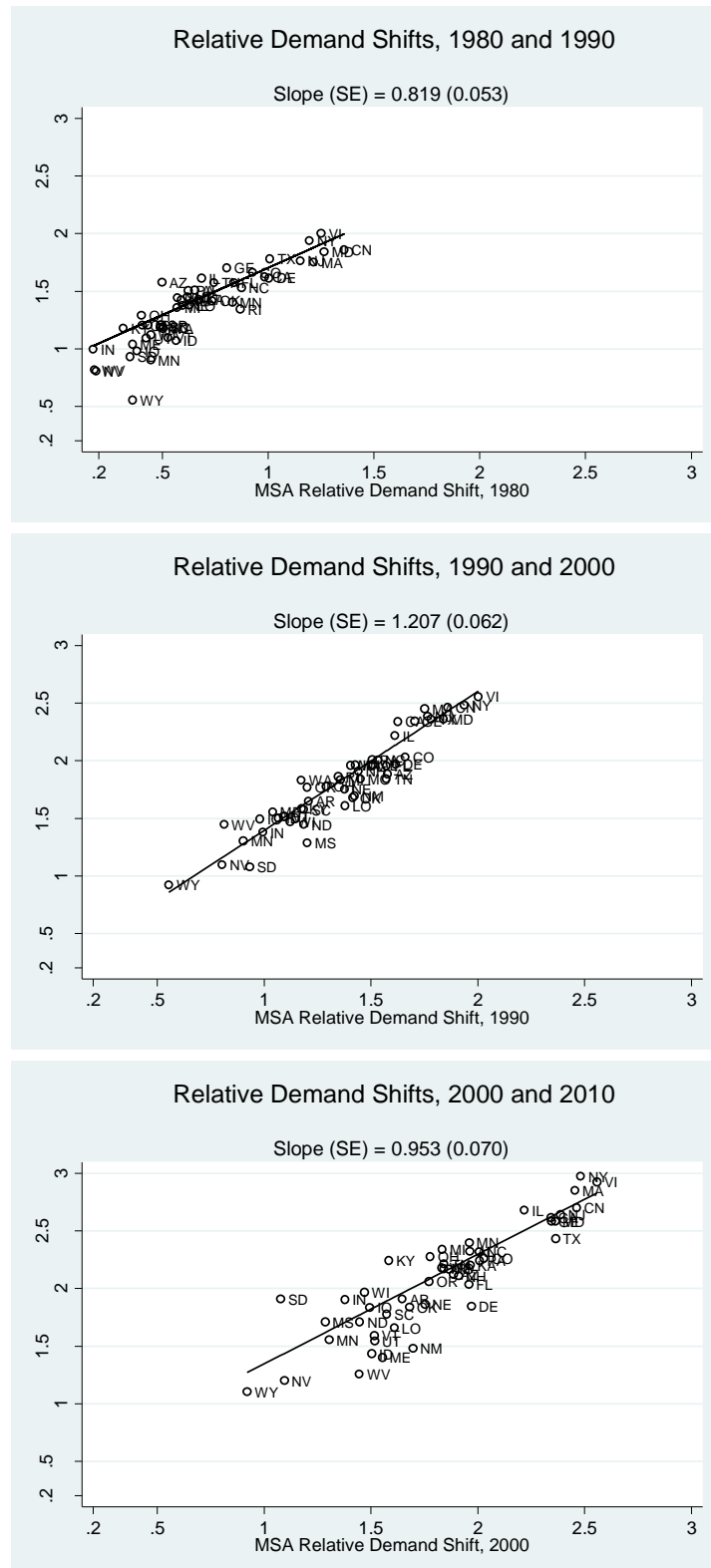
Notes: The relative demand shifts are calculated as $\log(L^{CE}/L^{HE}) + \sigma_E \log(W^C/W^H)$, where $\log(L^{CE}/L^{HE})$ is the log relative supply of college equivalent versus high school equivalent hours, $\sigma_E (= 1)$ is the elasticity of substitution between college and high school workers and $\log(W^C/W^H)$ is the fixed weighted composition adjusted college/high school log wage differential. The estimated slope coefficient (and associated standard error reported in parentheses) are weighted by the inverse sampling variance of the state level wage differentials.

Figure A1b: Implied Relative Demand Shifts by MSA, 1980 to 2010 ($\sigma_E = 1$)



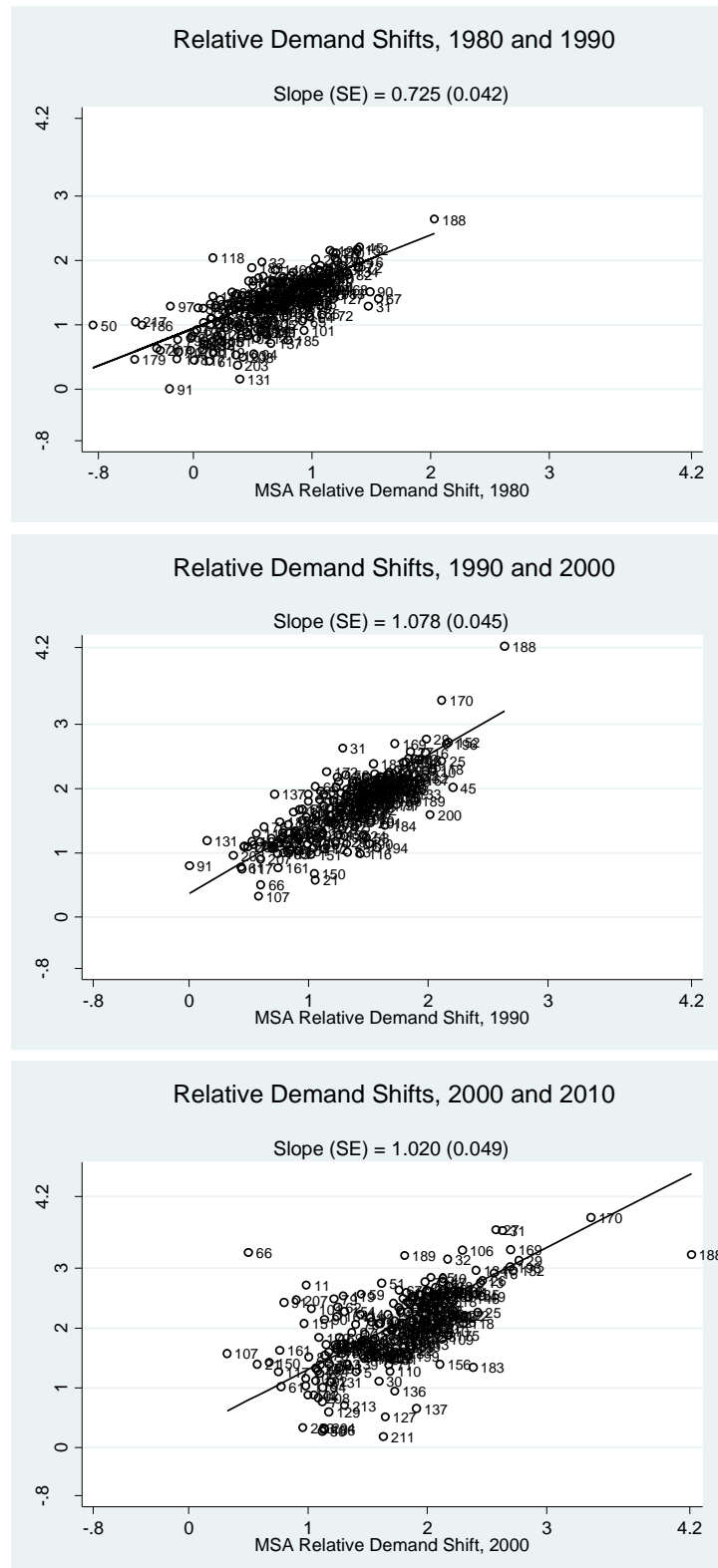
Notes: The relative demand shifts are calculated as $\log(L^{CE}/L^{HE}) + \sigma_E \log(W^C/W^H)$, where $\log(L^{CE}/L^{HE})$ is the log relative supply of college equivalent versus high school equivalent hours, $\sigma_E (= 1)$ is the elasticity of substitution between college and high school workers and $\log(W^C/W^H)$ is the fixed weighted composition adjusted college/high school log wage differential. The estimated slope coefficient (and associated standard error reported in parentheses) are weighted by the inverse sampling variance of the MSA level wage differentials.

Figure A2a: Implied Relative Demand Shifts by State, 1980 to 2010 ($\sigma_E = 5$)



Notes: The relative demand shifts are calculated as $\log(L^{CE}/L^{HE}) + \sigma_E \log(W^C/W^H)$, where $\log(L^{CE}/L^{HE})$ is the log relative supply of college equivalent versus high school equivalent hours, $\sigma_E (= 5)$ is the elasticity of substitution between college and high school workers and $\log(W^C/W^H)$ is the fixed weighted composition adjusted college/high school log wage differential. The estimated slope coefficient (and associated standard error reported in parentheses) are weighted by the inverse sampling variance of the state level wage differentials.

Figure A2b: Implied Relative Demand Shifts by MSA, 1980 to 2010 ($\sigma_E = 5$)



Notes: The relative demand shifts are calculated as $\log(L^{CE}/L^{HE}) + \sigma_E \log(W^C/W^H)$, where $\log(L^{CE}/L^{HE})$ is the log relative supply of college equivalent versus high school equivalent hours, $\sigma_E (= 5)$ is the elasticity of substitution between college and high school workers and $\log(W^C/W^H)$ is the fixed weighted composition adjusted college/high school log wage differential. The estimated slope coefficient (and associated standard error reported in parentheses) are weighted by the inverse sampling variance of the MSA level wage differentials.

Data Appendix

1. Basic Processing of the Census and ACS Data

We use the 5 % PUMS 1980, 1990 and 2000 Decennial Census data, as well as the 1 % 2010 ACS. We drop Alaska, Hawaii and the District of Columbia from all of our analyses. We consistently defined 216 MSAs between 1980 and 2010. Our basic sample consists of all working individuals aged 26-55. Hours are measured using usual hours worked in the previous year. Full-time weekly earnings are calculated as the logarithm of annual earnings over weeks worked for full-time, full-year UK born workers. Weights are used in all calculations. Full-time earnings are weighted by the product of the CPS sampling weight and weeks worked. All wage and salary income was reported in a single variable, which was top-coded at values between \$75,000 in 1980 and \$200,000 in 2010. Following Katz and Murphy (1992), we multiply the top-coded earnings value by 1.5. Earnings numbers are inflated into 2010 prices using the PCE deflator.

2. Coding of the Education Categories

We construct consistent educational categories using the method proposed by Jaeger (1997). For the pre 1990 education question, we defined high school dropouts as those with fewer than twelve years of completed schooling; high school graduates as those having twelve years of completed schooling; some college attendees as those with any schooling beyond twelve years (completed or not) and less than sixteen completed years; college-only graduates as those with sixteen or seventeen years of completed schooling and postgraduates with eighteen or more years of completed schooling. In samples coded with the post Census 1990 revised education question, we define high school dropouts as those with fewer than twelve years of completed schooling; high school graduates as those with either twelve completed years of schooling and/or a high school diploma or G.E.D.; some college as those attending some college or holding an associate's degree; college

only as those with a bachelor degree; and postgraduate as a masters, professional or doctorate degree.

3. Construction of the Relative Wage Series

We calculate composition-adjusted relative wages overall and by age cohorts using the wage sample described above, excluding the self-employed. The data are sorted into gender-education-age groups based on a breakdown of the data by gender, the five education categories described above, and three age categories (26-35, 36-45 and 46-55). We predict wages separately by sex and age groups. Hence, we estimate six separate regressions for each state/MSA and year including education and age dummies (as well as two dummies for race). The (composition-adjusted) mean log wage for each of the thirty groups in a given state/MSA and year is the predicted log wage from these regressions for each relevant education group. These wages are then weighted by the hours shares of each group for the whole time period.

4. Construction of the Relative Supply Measures

We calculate relative supply measures using the sample above. We form a labour quantity sample equal to total hours worked by all employed workers (including those in self-employment) age 26 to 55 to form education cells in each state/MSA and year. Education groups are high school dropout, high school graduate, some college, college graduate, and postgraduate. This provides our efficiency units by education group.

The quantity data are merged to a corresponding price sample containing real mean full-time weekly wages by state/MSA, year and education group (Wage data used for the price sample correspond to the earnings samples described above). For each state and year we calculate aggregate college equivalent labour supply as the total efficiency units of labour supplied by postgraduates weighted by the postgraduate-college graduate

relative wage from the price sample, plus college graduate efficiency units, plus 30 percent of the efficiency units of labour supplied by workers with some college weighted by the some college-college graduate relative wage. Similarly, aggregate high school equivalent labour supply is the sum of efficiency units supplied by high school or lower workers, plus 70 percent of the efficiency units supplied by workers with some college, all weighted by respective relative high school graduate average wages. Hence, the college-only/high school log relative supply index is the natural logarithm of the ratio of college-only equivalent to non-college equivalent labour supply (in efficiency units) in each state/MSA and year.

5. Compulsory School Attendance Laws

We update the data used in Acemoglu and Angrist (2001) which run from 1915 to 1978 to include state level compulsory attendance laws up to 2002. We obtained compulsory schooling law data up to 2002 from the Digest of Education Statistics, supplemented by information from the state statutes. We cross checked our new data with that derived in Oreopoulos (2009) and found we had very similar measures. This provides five variables (CA8, CA9, CA10, CA11 and CA12) for 48 states and years between 1915 and 2002 that capture compulsory years of schooling. For example CA8 equals one in the states that had a minimum of 8 years of compulsory schooling for the relevant years and zero otherwise, whereas CA12 equals one in the states that had over 12 years of compulsory schooling for the relevant years and zero otherwise. We match these into the individual Census and ACS data by state of birth and year aged 14.

6. R&D, Computer Use and Union Coverage Data

Our Research and Development (R&D) intensity measures are generated using R&D expenditure divided by nominal GDP for 1977, 1987, 1997 and 2007. These are taken

from the National Industrial Productivity Accounts (NIPA) made available by the Bureau of Economic Analysis. State level changes in R&D performance are measured using Total (company, Federal and Other) funds for industrial R&D performance in millions of dollars.

The computer use data are measured in proportions per state in each year. These are taken from the October 1984, 1987, 1997 and 2003 CPS supplements and derived from the question 'Do you use a computer at work?'. Computer use is the proportion of employed workers in the CPS that use computers at work.

The union coverage data are also in state level proportions per year and are taken from the Union Membership and Coverage Database provided by Hirsch and Macpherson (2003).²⁴ These are generated using CPS data beginning in 1973 and are updated annually.

²⁴ These data are available to download from <http://www.unionstats.com/>.