Trade in university training: cross-state variation in the production and stock of college-educated labor

John Bound[a,e,*], Jeffrey Groen[b], Gábor Kézdi[c,f], Sarah Turner[d,e]

[a] *Department of Economics, University of Michigan, 611 S. Tappan Street, Ann Arbor,
MI 48109-1220, USA*
[b] *Department of Labor Economics, Cornell University, 357 Ives Hall, Ithaca, NY 14853-3901, USA*
[c] *Department of Econometrics, Budapest University of Economics, Fovam ter 8,
1093 Budapest, Hungary*
[d] *Ruffner Hall, University of Virginia, Charlottesville, VA 22904-4260, USA*
[e] *National Bureau of Economic Research, Cambridge, MA, USA*
[f] *Institute of Economics, Hungarian Academy of Sciences, Budapest, Hungary*

Abstract

The question of this analysis is how the production of college graduates at the state level affects the stock of college-educated workers in the state. The potential mobility of skilled workers implies that the number of college students graduating in an area need not affect the number of college graduates living in the area. However, the production of relatively large numbers of college graduates in a state may lead to increases in the employment of university-trained manpower if industries expand production of goods and services that use college-educated workers intensively. We find at best only a modest link between the production and stock of baccalaureate degree recipients.
ⓒ 2003 Published by Elsevier B.V.

* Corresponding author. Department of Economics, University of Michigan, 611 S. Tappan Street,
Ann Arbor, MI 48109-1220, USA. Tel.: +1-734-647-2319; fax: +1-734-764-2769.
E-mail addresses: jbound@umich.edu (J. Bound), jg277@cornell.edu (J. Groen), kezdi@econ.core.hu
(G. Kézdi), sturner@virginia.edu (S. Turner).

0. Introduction

In the United States, higher education draws heavily on the resources of state and local governments through direct subsidies and indirect subsidies in the form of exemption from taxation. A rationale often given for why states invest in the education of their residents is that states enjoy some of the returns from such investments—the more highly educated a workforce, the more productive it is. What is more, highly educated workforces may contribute to regional economic growth by attracting new business. In fact, there is increasing evidence that the overall skill level of an area's workforce has fundamental effects on the local economy. [1] However, given the mobility of the labor force in general (Long, 1988; Bartik, 1991; Blanchard and Katz, 1992) and of college-educated labor in particular (Long, 1988; Bound and Holzer, 2000), there may be little correspondence between the number of college students graduating in a state and the number of college graduates living in the area. The question addressed in this analysis is whether the production of higher education in a state affects the stock of human capital in a state.

This analysis contributes to the understanding of how policies affecting the "supply side" or the production of college-educated workers compare to other incentives influencing the location choice of college-educated workers. Framing this analysis at the state level reflects the observation that it is state policymakers who determine the level of institutional subsidy for higher education and the associated tuition rates at public colleges and universities.

Our work is also important in understanding how local labor markets adjust to supply shocks. While labor economists have typically emphasized the importance of migration as the means by which local areas respond to supply shocks (e.g., Borjas et al., 1997), changes in output mix emphasized by trade economists are also potentially an important mechanism for the adjustment of local labor markets (Hanson and Slaughter, 2002). As far as we know, we are the first to attempt to quantify the relative importance of these two factors.

The first section of the paper outlines a model of the relationship between changes in collegiate degrees conferred in a state (flows) and the supply of college-educated workers in an area (stocks). The second section outlines our empirical strategy and describes the data available to analyze this relationship. The third section describes the level and variation in the flow and stock measures across states and over time and then presents estimates of the relationship between flows and stocks over the long run and in response to transitory shocks.

A central finding of this paper is that the relative flow of degrees conferred within a state has only a modest effect on the relative stock of university-educated workers within the state. As such, states have only limited capacity to influence the human

---

[1] Cities with well-educated workforces tend to grow faster than do cities with less well-educated workforces, with such differences persisting over time (Glaeser et al., 1995; Glendon, 1998). Moreover, wages of both well- and less-well-educated workers tend to be positively associated with the educational attainment of a city's workforce (Rauch, 1993; Moretti, 2004). The evidence on whether this association is causal is mixed (Moretti, 2004; Acemoglu and Angrist, 2001).

capital levels in their workforces by investing in higher education degree outputs. For university graduates likely to be employed in sectors producing goods and services traded across states, there is a positive association between the production and use of college-educated labor. In this regard, the presence of college graduates in a state works, to some extent, to attract industries that are intensive in such labor. Yet, labor markets for services produced and consumed locally produce quite different behavior and the case of medical degree recipients illustrates this point. In the health care sector, where output is more likely to be produced and consumed locally, there is virtually no relationship between state of degree production and employment. Although the production of MD degrees tends to be concentrated in large, densely populated states, MDs disperse across the country after degree completion.

1. Conceptual framework

A simple conceptual model aids in the interpretation of our empirical estimates. The purpose of the model is twofold. First, the model places the state-level degree outcomes of higher education in the context of supply and demand in the labor market, thereby guiding the economic interpretation of our estimates. The theoretical model also helps to illustrate the sources of bias likely to arise in estimating the effect of degree flows on stocks, which is important in the absence of credible instruments for our flow variables. (A more detailed discussion together with a more general version of the model can be found in Appendix 2 of Bound et al., 2001, henceforth BGKT.)[2]

Of primary interest is the effect of an exogenous change in the share of people graduating from college in a state (what we have referred to as flows) on the share of the population or the workforce with college degrees in a state (what we have referred to as stocks).

Fig. 1 presents the model graphically, with the horizontal axis indicating the supply of college-educated labor relative to high school-educated labor within the state and the vertical axis indicating relative wages for college-educated labor within the state. The F curve represents the relative share of the state's population receiving a college education. Without post-college migration, this would be the supply of college-educated labor relative to supply of high school-educated labor in the state. The S curve incorporates migration. Under infinitely elastic migration, S would be horizontal at the national wage ratio, w. Panel 1 shows the case of imperfect but nonzero mobility, which gives a more elastic S curve than F curve. The two curves cross at the relative wage level for which there is no net migration. For wages above this point there is net immigration of college-educated labor and S lies to the right of F; for wages below this level there is net emigration of college-educated labor and S lies to the left of F.

D represents the relative long-run demand schedule for college-educated labor within the state. Shifts in the relative supply of college-educated labor in a state lead to

---

[2] The model we present captures changes within a state in a partial equilibrium context. In this regard, wages outside the state are assumed to be exogenous and not affected by migration. We have confirmed that the qualitative implications of the model are robust to the partial equilibrium assumptions using a parameterized general equilibrium model of two equally large states.
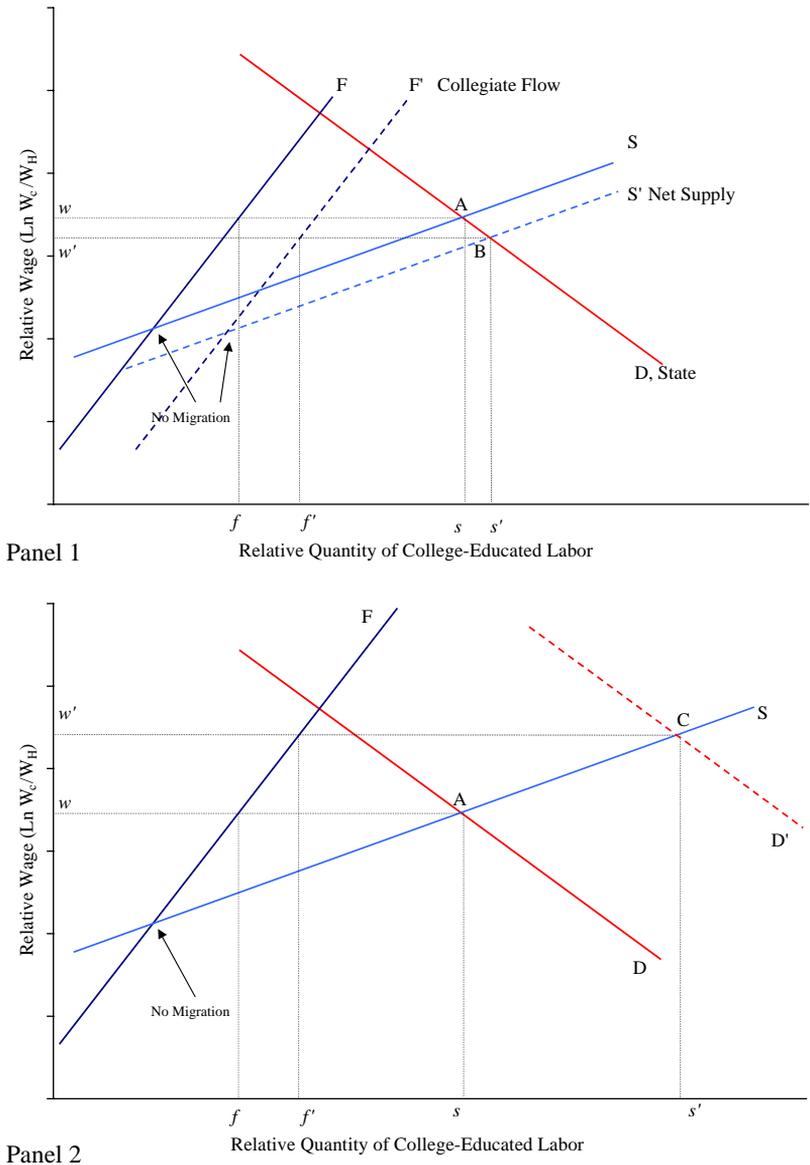
Fig. 1. State-level adjustments to changes in flows.

adjustments in production, with shifts toward industries and technologies intensive in college-educated workers when this factor is relatively plentiful. For this reason, within a state, we expect relative demand to be quite elastic. Indeed, in the textbook Heckscher–Ohlin model, relative demand curves are horizontal. In the top panel of

Fig. 1, the initial equilibrium occurs at point A. In the case represented, the state is a net importer of college-educated labor: equilibrium flows are smaller than stocks. We expect some states to be net exporters and others to be net importers.

An exogenous increase in the relative supply of college graduates in the state is indicated as a shift in the schedule of flows in the top panel of Fig. 1 from F to F'. The shift in F induces a shift in the net relative supply of college-educated labor in the state from S to S', and the equilibrium shifts from point A to B. The shift in S is likely to be somewhat smaller than the shift in F if, at the going wage rates, college graduates are more likely than high school graduates to migrate out of state. At the same time, we have drawn the curves under the simplifying assumption that the shift in F (and S) does not induce a shift in D—that is, assuming an increase in the relative flow of college graduates in the state does not directly affect the demand for college-educated labor. With the shift in the schedule of college graduates relative to high school graduates from F to F' and the shift in the schedule of the relative supply of college-educated labor (the stock) from S to S', equilibrium flows will shift from f to f' while equilibrium stocks will shift from s to s'.

Of interest is how the shift in relative flows (the shift from f to f') affects the change in the relative stock of college-educated workers in a state (measured by the adjustment in stocks from s to s'), in equilibrium. Algebraically, we have three equations:

$$f = \gamma w + \xi,$$

$$s = \theta w + f,$$

$$d = -\sigma w + \zeta,$$

where γ represents the within-state relative supply elasticity, θ represents the migration elasticity, and σ represents the elasticity of substitution between college- and high school-educated labor, and ξ and ζ represent degree flow and demand shifters, respectively. These three equations determine equilibrium values for w, f, and s, and we are interested in the comparative static results describing the effect of an exogenous shift in degree flows ($d\xi$) on w, f, and s:

$$dw = \frac{-1}{\sigma + \theta + \gamma} d\xi,$$

$$df = \frac{\sigma + \theta}{\sigma + \theta + \gamma} d\xi,$$

$$ds = \frac{\sigma}{\sigma + \theta + \gamma} d\xi.$$

In this regard, we are estimating how an exogenous shift in flows translates into a shift in stocks:

$$\beta \equiv \frac{ds/d\xi}{df/d\xi} = \frac{\sigma}{\sigma + \theta} \leqslant 1.$$

More mobility, represented by larger values of θ, dampens the effect of flows on stocks. At one extreme, no mobility ($\theta = 0$) leads to a one-to-one mapping between

changes in flows and changes in stocks. At the other extreme, frictionless mobility ($\theta = \infty$) leads to a zero effect of flows on stocks. In contrast, the larger the within-state elasticity of substitution between college- and high school-educated labor (σ), the larger will be the effect of changes in flows on stocks.

The model outlined above is quite general, and it is natural to consider how the parameters are likely to differ across markets for different types of skills. For baccalaureate degree recipients and types of skilled labor likely to be employed in the production of goods and services traded across geographic areas, we expect a substantial elasticity of substitution, reflecting changing output levels in response to changes in the availability of college-educated labor.[3] In contrast, for types of skilled labor that work exclusively or almost exclusively in the non-traded sector of the economy (e.g., physicians, nurses, school teachers), we expect the within-state demand curves to be relatively steep, since there is less room for the reallocation of labor across sectors. In this case, we also expect the relationship between the production and use of this type of skilled labor to be small.[4]

We are interested in determining the effect of *exogenous* differences in degree flows on stocks of university-educated workers at the state level. However, substantial variation across states in local labor market conditions may confound the measurement of the effect of variation in flows on the long-term concentration of college-educated labor in a state. To see this, consider the case where there is an exogenous shift in the relative demand for college educated labor: $d\zeta > 0$ and $d\xi = 0$, as illustrated in the bottom panel of Fig. 1. The equilibrium in flows, stocks, and wages represented by point A is identical to that in the top panel. A shift in the demand function is represented by D', which leads to movements along the flow and supply functions, with the new equilibrium established at point C. While demand-side variation also leads to a positive association between flows and stocks, the magnitude is quite different than in the initial case, with the change in stocks now exceeding the change in flows. What is more, the within-state equilibrium wage level rises rather than falls, as w' is greater than w in the bottom panel, while w' is less than w in the top panel. More generally, variation across states in the demand for college-educated labor tends to bias the parameter estimate upwards, with the magnitude of this bias depending on the relative magnitudes of the exogenous cross-state variation in supply of college-educated labor and demand for college-educated labor (see BGKT, Appendix 2).

---

[3] The point is not that all or even most college graduates work in the traded goods sector, but that enough do so that shifts in the share of a state's workforce that is college educated can be accommodated by sectoral shifts.

[4] In our empirical work, we focus on medical doctors because the American Medical Association maintains good data on the location of physicians and the institution from which they received their training. Most physicians are involved in direct patient care; in our data, 2.8 percent of active physicians identify their primary activity as research and another 1.5 percent identify their primary activity as teaching. It has been argued that physicians are capable of inducing a demand for their services, implying a perfectly elastic demand curve at the extreme. However, the empirical evidence on this point (see, for example, Newhouse et al., 1982) suggests that the location-specific demand for physicians is quite inelastic. Furthermore, in cases like medicine where the local supply elasticity is likely to be small, one might expect that employers and schools would work together to create institutions that facilitate geographic mobility, further weakening the relationship between flows and stocks.

The interpretation of our results requires some understanding of the source of the cross-state variation in stocks and flows. Variation in degree flows across states may occur because some states have a "comparative advantage" in the production of higher education, with differences derived from historical forces affecting the location choice of colleges more than a century ago, proximity to population centers, or willingness of voters to support higher education. Variation in demand for college-educated labor across states may occur because some states have a "comparative advantage" in the production of goods and services intensive in college-educated labor. The nation's political and financial capitals (Washington, DC, and New York City) might be examples of this kind of phenomenon.

Variation across states in the comparative advantage in the production of higher education, but not in the use of college-educated labor, would result in the states that produce a relatively large number of college-graduates per capita also using college graduates most intensively. While market forces would induce those trained in high degree production states to emigrate, the rank order of the measures of flows and stocks by state would not change. This dynamic would lead to a negative association with relative wages for both stocks and flows. In contrast, variation across states in the comparative advantage in the production of goods and services intensive in the employment of college-educated labor, but not in the production of college-educated labor, would also yield a high rank-order correlation between states' production and use of college-educated labor. Yet, in this case, we would expect to find a positive correlation with relative wages for both the production and the use of college-educated labor; however, causation would run from the labor market to the education market.

As we explain below, what we observe is that some of the states with high concentrations of college-educated workers also produce a disproportionate share of college graduates, while others import college graduates. Likewise, some of the states that produce a disproportionate share of college graduates also have a disproportionate share in their work forces, while others export college graduates. This is consistent with the notion that there is cross-state variation in the comparative advantage in both the production and use of college graduates.

2. Empirical strategy and data

2.1. Estimating equations

In placing this model in an empirical context, we analyze the association between cumulative per-capita flows of degrees awarded for birth cohort g in state j and the per-capita stocks for the same cohort in the same state in some subsequent year t by estimating the following equation:

$$\ln \frac{Stock_{jgt}}{Population_{jgt}} = \alpha_{gt} + \beta \ln \frac{Flow_{jgt}}{Population_{jg(g+r)}} + \varepsilon_{jgt}. \tag{1}$$

The independent variable is the total flow accruing to a cohort relative to the size of the cohort in the state around some modal age r, where this age reflects the typical

age of degree completion.[5] The dependent variable is the stock of degree recipients measured years after degree conferral for each cohort relative to the population in the state.[6] We present estimates for BA and MD degrees, as these two degree types are economically important and are well-identified in data on degree production and the educational attainment of the population.[7]

In addition to the cross-sectional analysis we investigate how changes in cohort-specific flows translate to changes in cohort-specific stocks. We look at changes over the intervals 1960–1970, 1970–1980, and 1980–1990. Here, the focus is on differences in the measures of flows and stocks over 10-year intervals defined for people of the same age referenced by birth cohort g and $g - 10$ in a state j. Again, we present the relationship in an elasticity form:

$$\Delta \ln \frac{Stock_{jgt}}{Population_{jgt}} = \alpha_{gt} + \beta \Delta \ln \frac{Flow_{jgt}}{Population_{jg(g+r)}} + \varepsilon_{jgt}, \qquad (2)$$

where Δ means differences between 1960 and 1970, etc. More specifically, for a variable x_{jgt}, the 10-year difference Δx_{jgt} is defined as

$$\Delta x_{jgt} = x_{j,g,t} - x_{j,g-10,t-10}.$$

This differenced specification captures medium-run dynamic effects rather than the long-run differences measured in the cross-sectional specification. These specifications have the advantage of eliminating state-specific fixed effects. Our intent in the differenced specifications is to estimate the extent to which idiosyncratic changes in a state's college degree output have sustained effects on the concentration of college-educated workers in the population. Of course, changes in degree flows may still be endogenous, and, as such, our estimates of the effect of changes in flows on changes in stocks are likely to exaggerate causal effects.

It is natural to assume that the medium-run impact of a shift in flows on stocks will be larger than the long-run impact. After all, the long-run supply (migration) elasticity will be larger the medium-run elasticity. However, when considering the relative magnitude of medium- and long-run equilibrium shifts, it is important to bear in mind that the magnitude of these shifts will depend on both demand and supply

---

[5] It is worth noting that flows measure the number of individuals in cohort g who eventually receive a degree in state j, regardless of whether the degree recipients were residents of the state at the time. In the U.S. most, but by no means all, undergraduate students attend college in their home state. Regardless, we are interested in the location of the degree-granting institution, not the residence status of individuals receiving the degree.

[6] Given our motivation, it might seem as if our dependent variable should be the share of the workforce, not the population, with a college degree. Empirically it makes virtually no difference whether or not we restrict attention to the workforce. Because we restrict our attention to cohorts that are working age (25–54) at the point of observation, cross-state variation in the fraction of the workforce with a college education is driven by the variation in the fraction of the population with a college education. Relative employment rates simply do not vary that much.

[7] The unit of the static model (1) is the state-cohort cell. As discussed in more detail with the presentation of the empirical results, the inclusion of year effects means that variation across states is what identifies our estimates. Throughout the text, we follow the common practice of using BA to refer to the broad class of baccalaureate or bachelor's level degrees, including the AB, BA, and BS degrees. Neither the Census data nor the Department of Education data collections distinguish among these degree labels.

parameters. In particular, if capital were relatively immobile (across sectors and across space) in the short run, but mobile in the long run, then it is entirely possible that the short-run impact of labor supply shifts might be small even if the long-run impact were substantial.

2.2. Data[8]

The data used in this analysis are from the decennial census surveys and annual institutional surveys of degrees awarded by colleges and universities conducted by the Department of Education. For BA degrees, we use institutional data on degrees awarded per year and the national distribution of degree recipients by age to estimate the number of degrees awarded to each birth cohort at the state level. To obtain measures of per-capita flows for BA degree recipients, we divide these imputed cohort-specific flows by the population in the state at age 22, calculated from widely available tabulations of the age distribution from the Census Bureau.

Because reliable state-level data on the age distribution of college graduates are not available, national data on the age distribution of BA degree recipients are employed in the calculation of the BA flow measures at the state-cohort level.[9] The use of national data in place of state-level measures undoubtedly introduces some measurement error. The bias caused by this error is complicated and reflects a combination of sampling error in the estimated age distribution, which is likely to be classical in form, and the unobserved cross-state variation in the age distribution at degree receipt, which is likely to be non-classical in form. Because the state-specific flows are substantially stable across time, small errors attributing annual flows to specific birth cohorts are unlikely to have any significant effect on our cross-sectional estimates. However, it seemed quite plausible that this problem could have an appreciable effect on our dynamic estimates. To gauge the magnitude of this problem, we conducted a number of simulations and they suggest that the magnitude of the bias introduced by the imputation error is relatively small (see Appendix B).

The data for MD degree recipients is from a database maintained by the American Medical Association (AMA) that records age and other demographic characteristics, institution of degree receipt, and professional employment location. Because we observe this universe in 1980 and 1991, we are able to make comparisons over time as well as across states. Our ability to organize the MD information by birth cohort mitigates some of the measurement problems associated with the timing of degree receipt for this group.

---

[8]Considerable additional detail on the data sources and the construction of the measures used in the analysis is available in Appendix A.

[9] Using micro data from the October Current Population Survey (CPS), we estimated the age distribution of college seniors and then estimated the number of college degrees conferred to individuals in each birth cohort each year. Micro data from the October CPS are available from 1969 to the present. For years prior to 1969 we assumed the age distribution of college seniors was similar to the age distribution in 1969. While it is possible to identify all states after 1977 in the CPS, cell sizes are generally too small to permit reliable state-level calculation of age at degree receipt.

To estimate the per-capita stock of BA college graduates in a state, we use micro data from the decennial census for years 1960, 1970, 1980, and 1990. We calculate the share of BA recipients in an age group relative to the population size as our stock measure. The 1990 census provides an advantage over previous decennial files for this analysis because educational attainment is presented as degree levels rather than years of completed education. For earlier census years (1960–1980), we make the standard assumption that 16 years of completed education equates to the receipt of a baccalaureate-level degree.[10] For MDs, we use AMA data on degree receipt to measure the numerator and census data to measure cohort size in the denominator.

3. Empirical stock–flow analysis

3.1. Concentration of flows and stocks

The starting point for the empirical analysis is the consideration of the concentration of degree flows and stocks across states and the population. We begin by considering those who received degrees between 1966 and 1985; for BA graduates this reflects the 27–46 age group in 1990, and for MDs the 32–51 age group. The mean degree flow and stock measures, presented in the first column of Table 1, are indicative of degree receipt, with BA degree recipients nearly 75 times more prevalent than MDs.

Colleges in the plains and northeast states are particularly strong producers of BAs. States such as New Hampshire, Vermont, and Massachusetts have nearly twice the per-capita flow as states like Georgia, South Carolina, and California at the BA level.[11] There is appreciably more variation across states in the production of degrees of MD degrees than in the production of BA degrees, as indicated by the lower cross-state coefficient of variation in BA flows. At one extreme, states that are not densely populated such as Montana, Idaho, and Wyoming do not record any institutions awarding MD degrees. At the other extreme, states such as New York, Illinois, and Iowa report relatively high production of MD degrees.

---

[10] The 1990 Census identifies both the state in which people live and, for those that work, the state in which they work. Earlier Census enumerations either do not identify state of work, or do so for only a subset of the sample. For consistency sake, all results we report are based on state of residence. We did, however, replicate our 1990 cross-sectional results classifying individuals according to the state in which they work. Switching to state of work made virtually no difference to any of our results.

We also limit the analysis to the 48 continental states, as data for Alaska and Hawaii are often difficult to obtain in early years and the obvious differences in geographic integration may lead to somewhat different dynamics. We do not include DC because the unusual political and industrial structure of this area often leaves this case an outlier.

[11] Recall that these measures reflect the state in which the degrees are obtained, not the state of student residence. Since the significant majority of undergraduates attend college in their home state, a measure based on the state of residence of students would produce a very similar picture. That said, some of the states that produce a larger number of BAs such as Vermont and Utah are net importers of out-of-state students, while others such as New Jersey, Connecticut, and California, on net, send some students out of state for college.

Table 1
Stock and flow summary statistics, 1966–1985 degree cohorts

		# per 1000	Cross-state CV	Analysis of variance		
				State	Cohort	Within
BA	Flow	256.09	1.1	0.77	0.12	0.11
	Stock	243.29	0.8	0.69	0.21	0.10
MD	Flow	3.47	2.7	0.87	0.05	0.08
	Stock	4.33	0.8	0.53	0.26	0.21

Notes: "Flow" data represent the number of degree recipients from a state divided by the age-appropriate population. Measures of the number of BA degrees awarded at the cohort-state level were estimated from year-state level institutional data and the national yearly age distribution of new college graduates. Cohort-state level numbers of MD degrees come from the AMA database. "Stock" data are the number of degree recipients living in the state in the appropriate age group, divided by the total population of the age group in the state. The stock data are based on 1990 for BAs and 1991 for MDs. The BA data represent individuals 27–46 years old in 1990, while the MD data represent individuals born between 1939 and 1958. The data are restricted to individuals receiving their degrees from or residing in one of the 48 continental states. CV is the partial coefficient of variation reflecting between-state variation in flows and stocks. The entries under "Analysis of Variance" represent proportions of total variance.

Table 1 also presents the analysis of variance for the stock and flow measures for BA and MD degrees. Decomposing the observed variance for the two decades of state-level observations reveals that the bulk of the variation is consistently across states. For example, at the BA degree level about 77 percent of the observed variation in flows is across states. Such persistence in the difference in the production of BA degrees awarded across states points to the presence of long-run differences. These cross-state differences have been in place since at least the pre-World War II point of observation. The dispersion across states of BA degrees awarded in 1929 is remarkably similar to the more recent distribution of flows, and the correlation between the two is 0.5 though the absolute level of dispersion across states is appreciably lower today.

3.2. Cross-sectional stock–flow analysis

The analytic question of interest is the measurement of the effect of variation in degrees conferred on the stock of college-educated labor across states. At issue in the interpretation of the cross-sectional relationships between degree flows and stocks is the extent to which the observed variation in the flow measure is exogenous, rather than reflecting essentially permanent differences in labor demand across states.

Overall, the distribution and scale of colleges and universities across states reflect a range of factors including the founding of private colleges in the 18th and 19th centuries, the willingness of local populations to support public expenditures on higher education, the introduction of federal support through the land-grant colleges, and the industrial composition of a state (Goldin and Katz, 1999). Some of these factors would seem largely exogenous to state labor markets, while others clearly are not. States with

Table 2
Cross-sectional stock-flow elasticities, 1966–85 degree cohorts

	Elasticity Ln–Ln (1)	Linear (2)	Elasticity from (2) (3)
BA	0.34 (0.08)	0.30 (0.07)	0.32 (0.08)
MD		0.08 (0.04)	0.07 (0.03)

Notes: See the notes to Table 1 for the definitions of "flow" and "stock". The regressions include data for 48 continental states. The unit of analysis is a state (48) × cohort (20) cell. The stock and flow variables correspond to the degree type listed in the first column. The regressions include cohort-specific fixed effects and correspond to the specification in Eq. (1) in the text. Standard errors are calculated using the method of Huber–White and allow for arbitrary clustering at the state level. Stock variables are associated with the 1990 census year for BAs and 1991 for MDs.

industries that have historically hired a disproportionate share of college graduates may be those that have invested in producing a supply to match the demand, thereby contributing to an upward bias in the cross-sectional estimates.

Table 2 presents OLS estimates in elasticity form of the cross-sectional link between flows and stocks, represented by Eq. (1). [12] Results indicate a modest association between flows and stocks for BAs, with an elasticity of 0.3. Plainly, states with relatively high production of undergraduate students also tend to have relatively high concentrations of the university educated in their working-age populations. Yet, this relationship is appreciably less than 1:1 and points quite clearly to the importance of migration as a mechanism for mitigating the impact of supply shocks on local economies. Calculating stock–flow relations for different age ranges and at different points in time (results not shown) underscores the persistence of the basic result. [13]

Comparisons across degree types highlight the quite different labor markets faced by university-educated labor with different types of training. The cross-sectional relationship between the production of MD degrees and the representation of MDs in the population is remarkably weak, with an elasticity estimate very close to zero. Taking the point estimates at face value, an increase of 10 percent in the medical degrees awarded in a state would be expected to produce an increase in the concentration of

---

[12] The inclusion of cohort-year dummies implies that the parameter is identified from cross-state variation. Standard errors are estimated allowing for arbitrary heteroskedasticity and state-level clustering in the errors. As can be seen in Table 1, most of the variation in flows is accounted for by permanent cross-state differences. As a result, we would have lost very little in terms of precision by estimating a regression of state-specific average stocks on average flows. Indeed, such regressions yield parameter and standard error estimates that are virtually identical to the ones we report.

[13] Results are unweighted, treating each of the 48 states equally. Weighting by average population size in a state or the square root of land area makes little difference to point estimates, but lowers the estimated precision of the estimates somewhat (Dickens, 1990). Conceptually, we prefer the analysis that puts each state on an equal footing.

Fig. 2. Stocks and flows of degrees awarded relative to cohort size, 1966–85 degree cohorts.

physicians in the state of less than 1 percent. The weak link between the flow and stock of MDs is consistent with an inelastic demand for MD services, reflecting employment concentrated in the non-traded sector.

   The graphical presentation of flows and stocks in Fig. 2 helps to sharpen our understanding of these estimates. Each panel represents the stock–flow relationship averaged over the 1966–85 degree cohorts, with the diagonal line distinguishing net importers (above) and net exporters (below). For the stock and flow of BA degrees, states such as California and Connecticut are BA importers while other states like Utah and Vermont

consistently export baccalaureate-trained personnel. [14] The picture for MDs is striking in the lack of association between flows and stocks.

While we estimate an elasticity of stocks with respect to flows that is significant and appreciably less than one, these estimates may exaggerate the effect of degree flows on stocks. Instrumental variables (IV) estimation provides a strategy to isolate the causal effect of the production of college-educated workers on the long-term stock. Because we are considering relatively permanent differences across states, variation across states in demographics and historical dimensions of the higher education industry are potential factors affecting degree production that are exogenous to contemporary developments in the labor market. One potential instrument is the per-capita flow of BA degrees in 1929; a second potential instrument is a measure of historical ethnic diversity, which potentially reflects the willingness to support public expenditure on higher education and we compute following Alesina et al. (1999). The IV estimates (not shown) tend to be somewhat larger than the corresponding OLS estimates (perhaps reflecting the observation that the instruments are plausibly related to differences in industrial structure), though the differences are not statistically significant. While we do not find our IV estimates more credible than our OLS estimates, they do underscore the point that the observed cross-sectional estimates reflect differences persisting throughout the 20th century.

Variation in relative wages across states with the concentration of college-educated workers provides a further indicator of the direction of the relationship between flows and stocks and the degree of mobility in the labor force. Table 3 presents estimates of the regression of relative wages for college graduates on the concentration of college graduates at the state level for different decennial points of observation. The relative wages were adjusted for demographic characteristics. The first column uses the observed concentration of college graduates as the explanatory variable. The second column uses the aggregate of flows (from 1950 to the indicated year) as an instrument for the stock of college graduates and this flow measure captures variation attributable to differences across states in degree production from higher education.

In an integrated labor market in which labor adjusts fully in location to changes in demand, these coefficients would be uniformly indistinguishable from zero. Yet, these estimates are consistently negative (particularly the IV estimates), implying an inverse relationship between flows and relative wages. This result is consistent with a situation in which some states have a comparative advantage in producing college graduates while others have a comparative advantage in the use of college-educated labor. Moreover, our estimates suggest that even in the long run, college-educated labor

---

[14] Here and elsewhere, there is suggestive evidence that states that are small geographically or those that are sparsely populated have a particularly hard time retaining their college graduates. Indeed, this would make considerable intuitive sense. We have tried to formally test for this by including interactions between flows and state size in our regressions. While the coefficients on such interaction terms were usually positive— suggesting that flows do, in fact, have a larger effect on stocks in larger states—magnitudes were generally small, and the coefficients never came close to being statistically significant at conventional levels. These results do not lead us to reject the notion that states differ systematically in their capacity to retain college graduates. Rather, with fundamentally only 48 observations, our sample size is not sufficiently large to detect such subtle differences.

Table 3
Estimates of effect on relative wages of concentration of college-educated labor

	OLS (1)	IV (2)
1960	−0.02 (0.06)	−0.38 (0.31)
1970	−0.06 (0.03)	−0.28 (0.15)
1980	−0.04 (0.02)	−0.16 (0.08)
1990	−0.03 (0.02)	−0.19 (0.09)

Notes: See the notes to Table 1 for the definitions of "flow" and "stock". Data used in these calculations are from the indicated decennial census files. The dependent variable is the regression-adjusted state-specific measure of the difference in the log of college and high school wages. The adjusted average relative wage measures are computed as the return to exactly a BA degree (or 16 years of completed education) from state-specific hourly wage regressions with a full set of controls for demographic and labor force characteristics including race, sex, part-time status, and potential experience specified as a quartic. The independent variable is the log of the ratio of college degree recipients to those with a high school degree or less. In column 2, the average (cohort-specific) per-capita flows over the 1950–1960, 1950–1970, 1950–1980 and 1950–1990 years were used as instruments for the measure of relative supply.

is not perfectly mobile across states.[15] In general, college graduates located in states that employ a large number of college graduates but do not produce a large number tend to receive something of a wage premium, while college graduates residing in states that produce a relatively large number of college graduates per capita tend to earn relatively little.

The coefficients in the second column of Table 3 can be interpreted as the negative inverse of the elasticity of relative demand for college-educated labor. The consistency of these estimates relies on the exogeneity of the instruments. If, however, the reported flow measure is endogenous, the estimated coefficients will tend to underestimate the causal effect of relative supply on relative wages and, as a result, will tend to overestimate this elasticity of demand.[16] Taking these coefficients at face value (i.e., interpreting their absolute value as estimates of the inverse of the elasticity of demand) suggests a within-state relative demand elasticity in the neighborhood of 5. These

[15] There are a variety of plausible reasons for why this might be true. The costs of moving may be sufficiently high for a large enough fraction of the population that the discounted benefits to relocation are not sufficiently high to merit the move. Alternatively, the marginal college graduate may have a preference for living in states that produce a large number of college graduates (this could simply represent a preference for living near friends and relatives). In this case, the earnings differences would represent equalizing differences.

[16] If flows are endogenous, then the regression of stocks on flows will tend to overestimate the causal effect of flows on stocks. Similarly, in this case the regression of relative wages on flows will tend to underestimate the causal effect of flows. The IV estimates are the ratio of these two estimates, and therefore will tend to underestimate the causal effect of stocks on relative wages.

Table 4
Means of difference measures of flow and stock

Time period	Age group	Ln difference of BA flow		Ln difference of BA stock	
		Mean	Std. dev.	Mean	Std. dev.
1960–1970	25–34	0.41	0.21	0.37	0.11
1970–1980	25–34	0.28	0.18	0.45	0.09
1980–1990	25–34	−0.10	0.16	−0.08	0.10
1970–1980	35–44	0.46	0.19	0.41	0.09
1980–1990	35–44	0.27	0.18	0.36	0.08
1980–1990	45–52	0.46	0.19	0.36	0.08

Notes: See the notes to Table 1 for the definitions of "flow" and "stock". Table entries reflect the log difference over the indicated decade of averages of flows nd stocks for the indicated age ranges.

estimates are all substantially larger than comparable estimates using U.S. time series data (Katz and Murphy, 1992), suggesting that considerable reallocation of production across states accounts for cross-state differences in relative stocks.[17] However, it also seems clear that even in the long run, within-state relative demand elasticities are well below infinity. Exogenous, cross-state increases in the supply of college graduates are accommodated by out-migration of college graduates and a drop in relative wages, as well as by reallocation of production across sectors.

3.3. Dynamic stock–flow analysis

Difference estimates capture changes over a relatively short horizon and thus measure something conceptually different from our cross-sectional estimates, which reflect permanent cross-state differences in educational capacity. In this regard, we want to know what happens to the stock of college graduates in a state if the degree output of the state's higher education institutions changes at a rate different from the national norm for a short interval.[18] Table 4 presents the means of the decennial log differences in flows and stocks by age and period of observation. As is well known, overall college attendance expanded dramatically into the early 1970s, accounting for the large and positive changes in flows for those in the 25–34 age group during 1960–1970 and 1970–1980. Decreased labor market returns to college education faced by cohorts

---

[17] Consistent with Hanson and Slaughter (2002), examination of cross-state variation in industrial composition indicates that states with relative concentrations of college graduates also tend to have concentrations of industries that are intensive in college-educated labor. Using 1990 census data we found the correlation between the fraction of a state's workforce that was college educated and the fraction predicted to be college educated based on the three-digit industry composition of the state to be 0.88.

[18] The data support this interpretation, as there is not uniformity in the correlation of changes in flows. States that increased relative flows between 1960 and 1970 were not identical to those with relative increases between 1970 and 1980, though there is a positive relationship between the 1970–1980 change and the 1980–1990 change. Overall, none of these relationships among flows is very strong nor is there evidence that they persist over time.
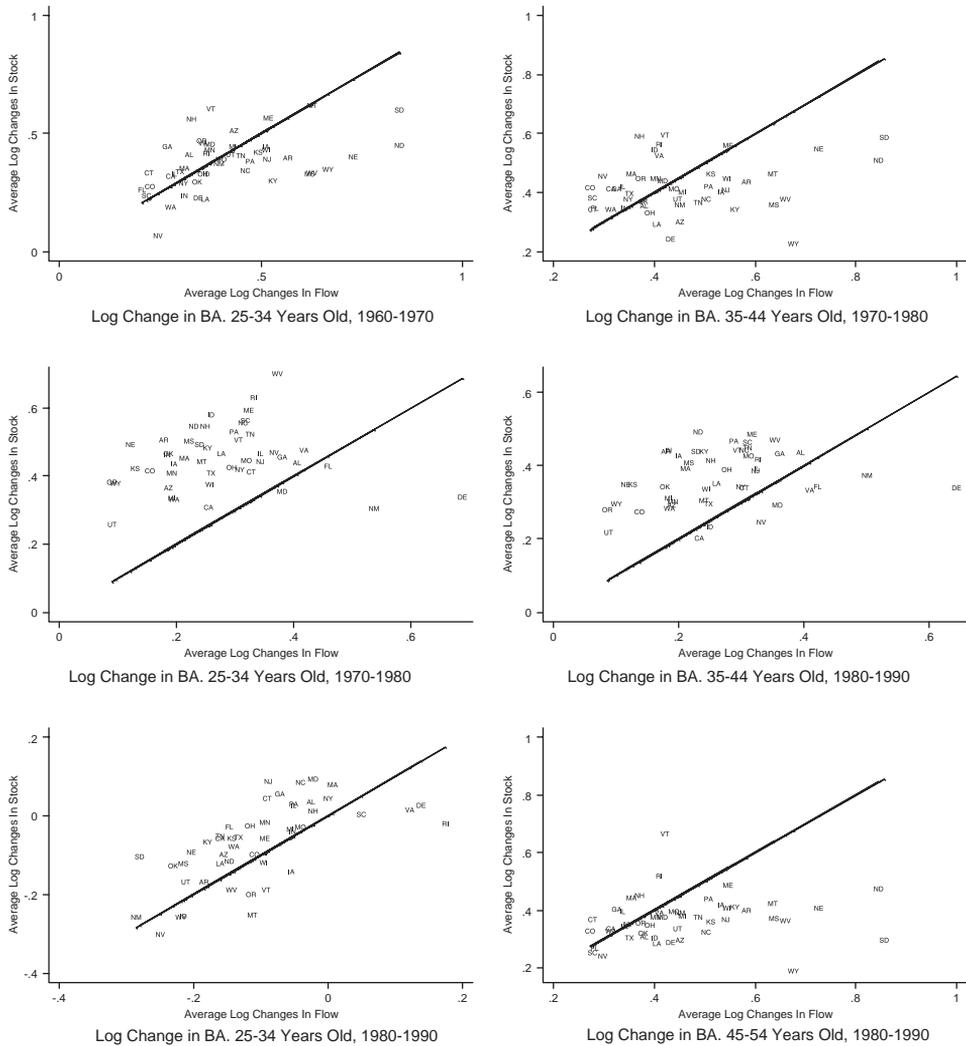
Fig. 3. Changes in flows and stocks relative to cohort size, BA degrees.

making educational investments in the mid- and late-1970s contributed to the decline in flows for the 25–34 age group during 1980–1990.

While the variation within states over time in degrees awarded is an appreciably smaller share of the total variance than the cross-sectional differences (see the analysis-of-variance numbers in Table 1), there is still significant cross-state variation in the change in the flows from one decade to the next, as shown in Table 4 and Fig. 3. Thus, for example, while average per-capita flows for 25–34 year olds increased by roughly 25 percent between 1970 and 1980, this growth varied considerably. It was close to 10 percent for states such as Oregon, Utah, Wyoming, and Nebraska; between

Table 5
Dynamic stock–flow estimates, BA degrees, 1960–1990

	Age group		
	25–34 (1)	35–44 (2)	45–52 (3)
1980–1990	0.44 (0.06)	0.22 (0.08)	0.08 (0.07)
1970–1980	0.13 (0.10)	0.22 (0.08)	
1960–1970	0.36 (0.09)		

Notes: See the notes to Table 1 for the definitions of "flow" and "stock". The regressions include data from the 48 continental states. The regressions are specified with differenced observations for each age and include age fixed effects, as presented in Eq. (2). The unit of analysis is the decennial difference measured at the state (48) and single-year age level (10 or 8 ages, as noted). Standard errors are calculated using the method of Huber–White and allow for arbitrary clustering at the state level. Since our flow data start in 1950, it is not possible to construct flows for cohorts older than 32 in 1960, 42 in 1970, or 52 in 1980. Thus, the observed age range is 25–32 for the 1960–1970 difference, 35–42 for the 1970–1980 difference, and 45–52 for the 1980–1990 difference.

36 and 46 percent for Florida, Nevada, Alabama, and Virginia; more than 50 percent for New Mexico; and about 70 percent for Delaware.

In terms of the changes over time, most of the movement is in the public sector,[19] where there was significant growth both at research universities and among public comprehensive institutions. The over-time analysis eliminates permanent cross-state differences, including long-run differences in state demand for college-educated workers, which contribute to bias in the estimates of the effects of flows on stocks. Nevertheless, the change over time in per-capita flows could still be endogenous to state-specific changes in the demand for college-educated labor. When thinking about how serious an issue this is, it is important to understand that the variation at issue represents cross-state differences in the growth of flows from one decade to the next. Growth in one decade is typically not followed by growth in the next. Although all states experienced an increase in the fraction of their college-aged population attending and finishing college between 1950 and 1970, the timing and magnitude of these increases varied, suggesting exogenous factors at work.[20]

Table 5 presents estimates with the decennial change in stock regressed on the decennial change in flow for different age groups. These dynamic estimates, reflecting the difference presentation from Eq. (2), use variations over time within states to

---

[19] Within states over the 1966–1985 time period, there is roughly three times more variation in per-capita flows in the public sector than there is in the private sector.

[20] State-level politics and the extent to which governors have exhibited strong leadership in the higher education sector explain some of the variation within states over time in degree production. See BGKT for a further discussion of this point.

identify the effect of flows on stocks. Estimates for relatively recent college graduates—those that are 25–34 years old as of the census years—are shown in the first column. For these cohorts, the difference estimates show significant effects of flows on stocks (0.36–0.44) for the 1960–1970 and 1980–1990 intervals, while the estimate for 1970–80 is somewhat weaker.

The scatter plots in Fig. 3 reveal several outliers. For the 1960–1970 decade, the Dakotas and Nebraska witnessed particularly dramatic increases in degrees conferred relative to the size of the college-age cohorts during the 1950s. For the 1970–80 decade, Delaware and New Mexico are notable for their relative expansion in college graduates during the 1960s, an increase that yielded relative little growth in the proportion of college-educated workers living in these states. Removing these two states from our calculations produces results for this cohort that are much more in line with results for other cohorts and that show statistically and quantitatively large associations between changes in flows and changes in stocks [0.28 (0.08)].[21] However, outliers usually contain valuable information. In Delaware and New Mexico, despite enormous increases in the per-capita share of college graduates, the stocks did not increase some years down the line. These outliers seem to confirm our tabulations, showing that flows have at best a moderate effect on stocks.

Table 5 also presents results for older age groups that would typically have graduated from college more than 10 years prior to the year in which we observe them (columns 2 and 3). These results indicate that the relationship between flows and stocks tends to diminish somewhat as cohorts age, with the elasticity declining to 0.22 (0.08) for those in the 35–44 age group and to 0.08 (0.07) for those in the 45–52 age group. When thinking about this diaspora of college graduates, it is important to bear in mind that the growth in flows in one decade is not generally followed by increases in flows in following decades. Thus, the impact of a change in flows on stocks two to three decades later is conceptually distinct from the long-run impact of variation in flows (i.e., the kind of quantity we were attempting to estimate using the cross-state variation in flows).

The graphs in Fig. 3 illustrate the dynamics underlying the regression results. States above the 45° line are cases in which the change in the stocks of college-educated workers exceeded the change in flows of college graduates, which means these states increased net imports of college graduates. In turn, states below the line tended to shift to exporting college graduates. The cross-state patterns vary considerably over time, with the shifts observed in the 1960–1970 period markedly larger than those observed for 1980–1990.

To address the concern that these first difference estimates do not necessarily reflect the effect of flow shifts on stocks, the optimal fix would be to use exogenous factors that have changed over time as instruments for changes in flows in our difference specification. However, state policy variables such as tuition and appropriations do not have a strong enough effect on college completion to provide us with IV estimates of

---

[21] For those in the 25–34 age group, difference estimates for other cohorts include 0.30 (0.09) for 1960–1970 and 0.41 (0.06) for 1980–1990 for regressions limited to 46 states and excluding DC, Delaware, and New Mexico.

Table 6
Dynamic stock–flow estimates, MD degrees, 1980–1991

Age group	Linear (1)	Elasticity from (1) (2)
35–44	0.22 (0.06)	0.17 (0.04)
45–54	0.25 (0.07)	0.16 (0.05)

Notes: See the notes to Table 1 for the definitions of "flow" and "stock". The regressions include data from the 48 continental states. The regressions are specified with differenced observations for each single year of age, with the inclusion of age fixed effects, and correspond to the specification presented in Eq. (2). The unit of analysis is the 1980–1991 difference measured at the state (48) and single-year age (10) level. Standard errors are calculated using the method of Huber–White and allow for arbitrary clustering at the state level.

any reasonable degree of precision. As an alternative, we have also tried controlling for various factors such as changes in cohort size[22] and labor demand that might confound the effect of changes in flows on changes in stocks. Adding cohort size as a covariate reduces the coefficient slightly and including direct measures of demand captured by the employment level produces negligible changes in the point estimates from the original specification (see BGKT, Table 5).

Table 6 presents dynamic estimates in the MD-degree market for age groups 35–44 and 45–54 over the 1980–1991 interval. The results indicate a positive and significant relationship between changes in flows and changes in stocks. While this result is some-what unexpected in light of the insignificant cross-sectional result, scatter plots of the data presented in Fig. 4 reveal a clear and compelling story. States that had the largest changes in flows tended to be states like West Virginia and South Dakota that may have been underserved in medical care at the beginning of the interval. As such, the addition of medical schools to two West Virginia universities could have been a policy remedy to increase the supply of doctors in the state.[23] Both institutions have mission statements that address the need to provide physicians and medical personnel for underserved areas and make explicit reference to recruiting students from rural West Virginia and placing graduates in clinical practices to improve health care in West Virginia. It is likely that the medium-term effects of changing the production of MDs within a state, captured by these difference regressions, may be relatively large as the additional MDs produced in a state like West Virginia include many people who are from West Virginia and have a preference for remaining in the state. Still, the absolute magnitudes of the coefficients are small (0.2) and indicate that for each 10 additional physicians trained in the state, only about 2 will remain in the state's population.

---

[22] There is a strong negative relationship between state-specific cohort size and educational attainment (Card and Lemieux, 2000; Bound and Turner, 2003).

[23] These were the West Virginia University School of Medicine (part of the Robert C. Byrd Health Sciences Center) and the Joan C. Edwards School of Medicine at Marshall University. West Virginia University awarded its first MD in 1962 and Marshall University established its medical school in 1977.
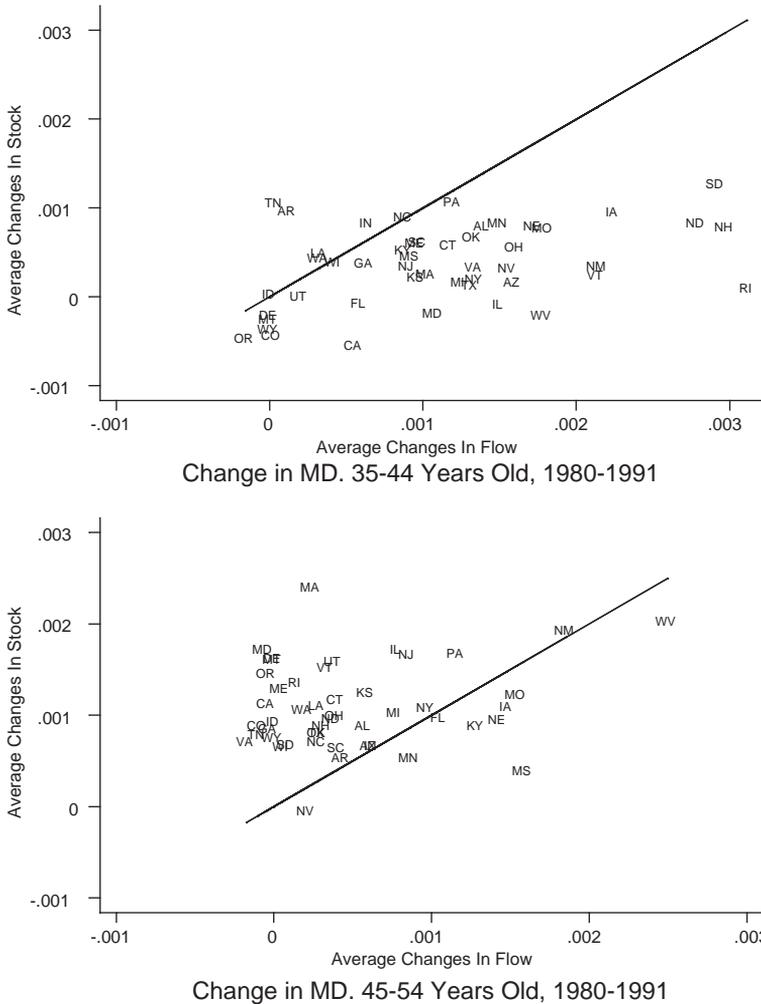
Change in MD. 35-44 Years Old, 1980-1991

Change in MD. 45-54 Years Old, 1980-1991

Fig. 4. Changes in flows and stocks relative to cohort size, MD degrees.

4. Conclusion

The empirical evidence in this analysis points to a modest relationship between degree production in the education market and the concentration of college-educated workers in a state's population in the long run. For BA degrees, we estimate the long-term elasticity between stocks and flows to be no greater than 0.3. In contrast, for MDs we find little relationship between where physicians are trained and where they practice. The contrast between our results for BAs and our results for MDs suggests that the nature of demand in the labor market is a substantial determinant of the stock–flow relationship, with the relatively inelastic demand for MDs within states contributing to the wide dispersion of MD graduates across states and the relatively

weak link between flows and stocks. For MDs, dynamic estimates are somewhat larger than long-run estimates (though still smaller than the corresponding estimates for BAs) and analysis of specific cases points to the influence of public policies intended to expand the supply of MDs in underserved areas.

Our results indicate that state policymakers have only a modest capacity to influence the human capital levels of their populations by investing in higher education degree outputs. Although, presumably, states have a number of reasons to invest in higher education, the mobility of college-educated labor reduces incentives to make such investments (Justman and Thisse, 1997). Indeed, some evidence suggests that the degree of mobility affects states' investment behavior. States with a higher out-migration of college graduates tend to have lower per-capita education expenditures (Clotfelter, 1976; Strathman, 1994). What is more, the relative magnitude of the subsidies states provide to medical students appears to be quite low relative to the subsidies provided to undergraduate education. [24] The fact that physicians are quite unlikely to remain in the state in which they were trained provides one explanation for this phenomenon.

Our estimates are also suggestive of how state economies adjust to supply shocks in the U.S. The labor economics literature (e.g., Blanchard and Katz, 1992; Borjas et al., 1997) has argued for the importance of migration as a means that states have of adjusting to macroeconomic shocks. In contrast, trade economists (e.g., Davis et al., 1997; Hanson and Slaughter, 2002) have emphasized the important of output mix shifts. Our results suggest that, at least in the U.S., migration does have a central role in mitigating the effect of labor supply shifts, but clearly other adjustment processes are also at work. For BA degree recipients, who are relatively likely to work in the traded goods sector, the within-state relative demand elasticity for college-educated labor seems to be substantially higher than the one for the country as a whole, suggesting that supply shifts induce the reallocation of production across sectors. What is more, even in the long run, wage differentials between college and high school graduates seem to be relatively small in states that produce a large number of BAs on a per-capita basis, suggesting that college-educated labor is less than perfectly mobile across states.

Acknowledgements

---

[24] Given the joint nature of production, it is difficult to estimate directly state subsides to either undergraduate or graduate education. The difference between private and public institutions in the tuition charged to medical students is much smaller than the difference private and public institutions in the tuition charged to undergraduate students. For universities offering both BA and MD degrees, private universities charge on average a tuition of $25,655 in MD programs and $18,235 in undergraduate programs. At public universities, the average tuition charged to medical students is $10,398 for residents and $22,353 for non-residents, while the average tuition charged to undergraduates is $3,140 for residents and $8,668 for non-residents. (These data are from academic year 1996–97. Data for tuition and fees in MD programs are from the Association of American Medical Colleges and the corresponding undergraduate data are from the IPEDS survey.)

Appendix A. Data appendix

The primary sources of data for this analysis are: the decennial census files, the decennial census publications, population estimates by the Census Bureau, the October Current Population Survey (CPS) files, institutional surveys of degrees awarded, and the American Medical Association (AMA) Physician Professional Data. Table 7 lists specific references.

In the paper, we analyze two types of degrees: BAs and MDs. The first part of the appendix is organized by degree type. Steps that apply to both types are discussed with BA degrees.

A.1. BA degrees

A.1.1. Flows

We wish to compute the *per capita flow* of college graduates for each state and birth cohort, and the *per capita stock* of college graduates for each state and cohort for the census years 1960, 1970, 1980, and 1990.

A major measurement problem arises from the fact that, in our baseline data, the *stock* variable is defined for birth cohorts, while the *flow* refers to the year of graduation. Therefore, we had to estimate *flows* for birth cohorts from the degree-year data. We do not observe the flow of BA degrees awarded to members of birth cohorts directly, but only the sum of degrees awarded to cohorts of different ages. While age 22 is the modal age group for BA recipients, not all BA degree recipients are this age.

First, in each year, we estimated the number of college degrees conferred for different ages. Using micro data from the October CPS, we estimated the age distribution of college seniors and then estimated the number of college degrees conferred to individuals in each birth cohort each year. Micro data from the October CPS are available from 1969 to the present. For years prior to 1969, we assumed the age distribution of college seniors was similar to the age distribution in 1969. Since the age distribution corresponds to the age distribution in October—while most individuals receive their degrees in June—we estimated the June age distribution by assuming that half of each cohort was a year younger at that time (basically, assuming uniform distribution of month of birth).

To estimate the number of degrees conferred by cohort in each state, we multiplied the age distribution of BA degrees conferred by the annual number of degrees awarded in each state as reported in institutional surveys of colleges and universities. We then assigned the different year-by-age numbers to the corresponding cohorts (defined by year of birth) in each state. Thus, for example, if we estimated that 40 percent of

Table 7
Data sources

Source	Use
Census of Population and Housing U.S. Department of Commerce, Census Bureau 1960 Public-Use Sample: 1 Percent Sample 1970 Public-Use Sample: 1 Percent Sample, 15 Percent State Questionnaire 1970 Public-Use Sample: 1 Percent Sample, 5 Percent State Questionnaire 1980 Public-Use Sample: 5 Percent Sample 1990 Public-Use Sample: 5 Percent Sample	Census data provide stock measures of educational attainment by state.
1950, 1960, and 1970 Census State Volumes State Population Estimates, 1950–1970 U.S. Census Bureau	Population estimates by state and single year of age, in 1950, 1960, and 1970. Used for estimating cohort size by state at age 22 and age 25–29, for years 1950–1970.
State Population Estimates by Single Year of Age, 1971–1990 U.S. Census Bureau	Estimates of cohort size by state at age 22 and age 25–29, for years 1971–1990.
Current Population Survey, October, 1969–1990 U.S. Census Bureau and Bureau of Labor Statistics	Used to estimate the age distribution of college seniors, 1969–1990.
Earned Degrees Conferred, Annual, 1950–1966 U.S. Department of Education	Data on BA degrees conferred, collected annually through institutional surveys.
Higher Education General Information Survey: Earned Degrees, 1967–1968 to 1985–1986 *Integrated Postsecondary Education Data System:* Earned Degrees, 1986–1987 to present U.S. Department of Education, National Center for Education Statistics	
Physician Professional Data, 1980 and 1991 year-end files American Medical Association	Data contain information on date of birth, medical school, and current state of residence for physicians practicing in the U.S. We use these data to construct stocks and flows by state and birth cohort.

graduating seniors in 1980 were 22, we would assign 40 percent of the degrees conferred in each state in 1980 to the cohort born in 1958. [25]

Third, we divided this absolute number by the size of the cohort in the state. Since the modal age of college completion is 22, we used estimates of the size of the cohort at

---

[25] Since we did not have degrees conferred data for years prior to 1950, our per capita flow estimates are truncated for those cohorts that turned 22 in the early 1950s. The (extrapolated) age distribution of the graduates suggests that approximately 70 percent of a cohort graduated at the age of 22 or before. This means that 30 percent of those who turned 22 in 1950 received their degrees before 1950, the time period for which we have no data. For similar reasons, we do not have degrees conferred data for 20 percent of the cohort that turned 22 in 1951, for 15 percent of those that turned 22 in 1952, and so on.

age 22 for this purpose. The U.S. Census Bureau reports population estimates by state and single year of age for 1970 and later. For the years prior to 1970, we estimated the share of 22-year-olds in the total population in a state by a weighted average of the corresponding birth cohorts from the two closest census figures, the weights being inversely proportional to the distance from the given census year. [26] We then used these shares and the total population of the states in the corresponding years to estimate the number of 22-year-olds.

The per capita flow data is the ratio of these two estimates: the estimated number of degrees conferred for a given birth cohort in a given year, divided by the size of that cohort in the state when they were 22 years old.

In the analysis, we use stock variables referring to different years: 1960, 1970, 1980, and 1990. When analyzing flows with these different stocks, one does not want to include in the flows those college graduates who received their degrees after the year of the stock. Therefore, were carried out the estimation of the flows four times, each corresponding to one stock variable.

A.1.2. Stocks

To estimate the per-capita stock of college graduates in a state we used micro data from the decennial census for the years 1960, 1970, 1980, and 1990. For 1960, the largest sample available represents a 1 percent random sample of the population. For 1970, the largest sample for which state of residence is available represents a 2 percent random sample of the population. For 1980 and 1990, the samples we used represent 5 percent of the population. In 1990, the census asked about the highest degree received by an individual. We assume that all those who identified themselves as having a Bachelor's, Master's, Professional, or Doctorate degree were college graduates. For earlier years, the census asked how many years of college a person had completed. For these years, we assumed that anyone who completed 4 or more years of college was a college graduate.

The 1980 and 1990 censuses allow one to identify state of work as well as state of residence. We did separate tabulations using data organized by state of residence, but found it made very little difference whether we identified individuals by their state of residence or their state of work.

A.2. MD degrees

For MD degree recipients, we constructed degree flows from the AMA Physician Professional Data. This is a comprehensive source of information on U.S. physicians, including both members and non-members of the AMA. The file includes information on date of birth and medical school for each physician. For physicians trained at U.S. medical schools, we used the year-end files for 1980 and 1991 to construct an estimate of degree flows by state and year of birth. We matched these degree flows—

---

[26] For example, the share of 22-year-olds in the population of a given state in 1963 was estimated by 0.3 times the share of the 29-year-olds in the 1970 census plus 0.7 times the share of the 19-year-olds in the 1960 census.

by state and year of birth—with population in the 5-year age group 25–29. (About 80 percent of the physicians in the 1991 AMA file received their MD between the ages of 25 and 29.)

We calculated the population age 25–29 in each state and year using Census Bureau data and followed the imputation strategy discussed above under BA degrees in the years prior to 1970 when single-year-age tabulations were unavailable.

To estimate the per-capita stock of MDs in 1980 and 1991, we tabulated the number of physicians by state and year of birth. For our stock measure, we dropped those who are not actively practicing medicine and those in residencies. To put these stocks in per-capita terms, we divide them by the population in 1980 or 1991, by state and year of birth.

A.3. Other variables

Birth cohort size: Data on the size of a birth cohort for each state from 1928 to 1970 (these are the cohorts that would have been age 22 between 1950 and 1992) were entered from vital statistics data distributed by the National Center for Health Statistics. The original data came from birth registrations.

BAs in 1929: Counts are from Table 4a, "Summary of degrees conferred in 1929–30," of *Biennial Survey of Education*, 1928–1930.

Racial disparity index: Calculated to replicate the variable used by Alesina et al. (1999). In particular, we computed the index for a state as: $Q_{jt} = 1 - \Sigma_k s_{jkt}^2$, where s_{jkt} is the share of the kth racial group in the jth state in year t. Following Alesina et al. (1999), we categorized individuals as White, Black, American Indian, Asian, or other. While we computed the index for all census years between 1960 and 1990, the results mentioned in the text are based on the 1960 values of the index. With five groups, the index has a possible range from 0.0 to 0.8. In fact, in our data the index ranges from 0.003 for Vermont to 0.49 for Mississippi, with an average value of 0.17.

Relative wages: The adjusted average relative wage measures are computed as the return to exactly a BA degree (or 16 years of completed education) from state-specific hourly wage regressions with a full set of controls for demographic and labor force characteristics including race, sex, part-time status, and potential experience specified as a quartic.

Appendix B. Assessing the effects of measurement error in the BA flow variable

Measurement error may arise because institutional data recorded by year of college graduation rather than year of birth are used in the calculation of the BA flow variable. At issue is the magnitude of the bias due to measurement error in BA flows (number of degrees conferred by state and birth cohort).

If the age distribution of degrees conferred by year of graduation and state were known, transforming institutional data by year of degree to measures based on birth cohorts would be straightforward. Two problems hinder this strategy. First, state-level measures of the age distribution are not available over a long time horizon, though national estimates of the age distribution at degree receipt based on calculations from

the Current Population Survey (CPS) are available. What is more, since the source
of these measures is a sample, these estimated averages are also measured with error.
This second problem might be thought of as classical measurement error and, on its
own, would lead to attenuation in parameter estimates. However, the overall effect of
using national averages on the magnitude and direction of the bias in the parame-
ter estimates is more complicated, depending on both the measurement error caused
by sampling variation and the error attributable to substituting national measures for
state-level variation in the age distribution at degree receipt. Simulation evidence serves
to frame the magnitude of the problem.

In this appendix, we consider the over-time equations of log stocks on log flows; see
Eq. (2) in the text.[27] The introduction of measurement error has a smaller effect on
the cross-sectional estimates, because there is substantial stability in state-specific flows
across time. As a result the mis-assignment of flows to cohorts is likely to have virtually
no effect on our cross-sectional estimates. On the other hand, such mis-assignment
could, potentially, have a large impact on our over-time estimates.

B.1. The problem

Let F indicate the flow variable measured by degrees conferred and let S indicate
the stock variable, the number of people with at least a BA. The unit of observation is
state, year of birth, and year of measurement, with these levels indicated by subscripts
j, g, and t, respectively. The three subscripts are always presented in the order jgt. A
dot instead of a particular subscript means the sum over all possible values of it. For
example, $F_{jg.}$ means total number of degrees conferred in state j among those who
were born in year g.

Following the dynamic estimates in the text, consider estimating the effect of flows
on stocks:

$$\Delta \ln S_{jgt} = \alpha_{gt} + \beta_t \Delta \ln F_{jgt} + \varepsilon_{jgt},$$

where

$$\Delta \ln S_{jgt} = \ln S_{jgt} - \ln S_{j(g-10)(t-10)},$$

$$\Delta \ln F_{jgt} = \ln F_{jgt} - \ln F_{j(g-10)(t-10)}.$$

We want to estimate

$$\beta = \frac{\sum_g Cov_j(\Delta \ln S_{jgt}, \Delta \ln F_{jgt})}{\sum_g Var_j(\Delta \ln F_{jgt})},$$

where Cov_j and Var_j mean covariance and variance across states.

---

[27] For notational simplicity, we consider this problem abstracting from the population variables (the de-
nominators of the dependent and independent variables). There is little reason to suspect mis-measurement
in the population variables and, in practice, the denominators have very little effect on the estimates.

F_{jgt} can be written as the sum of degrees conferred by graduation year (from some minimum age to year t), multiplied by the fraction of those who were born in year g:

$$F_{jgt} = \sum_{a=a0}^{t-g} F_{jg(g+a)} = \sum_{a=a0}^{t-g} \frac{F_{jg(g+a)}}{F_{j.(g+a)}} F_{j.(g+a)},$$

where a_0 is the lowest possible age of getting a BA degree.

Let

$$W_{jgs} \equiv \frac{F_{jgs}}{F_{j.s}}, \quad \bar{W}_{.gs} \equiv \frac{F_{.gs}}{F_{..s}} = \frac{\sum_j F_{jgs}}{\sum_j F_{j.s}}, \quad \delta_{jgs} \equiv W_{jgs} - \bar{W}_{.gs}.$$

Then

$$F_{jgt} = \sum_{a=a0}^{t-g} F_{jg(g+a)} = \sum_{a=a0}^{t-g} W_{jg(g+a)}F_{j.(g+a)} = \sum_{a=a0}^{t-g} (\bar{W}_{.g(g+a)} + \delta_{jg(g+a)})F_{j.(g+a)},$$

where δ_{jgs} is the deviation in state j from the national average $\bar{W}_{.gs}$. Our problem is we observe $F_{j.s}$ and $\bar{W}_{.gs}$ but not δ_{jgs}. Moreover, we measure $\bar{W}_{.gs}$ with a classical error $\omega_{.gs} \sim iid\mathrm{N}(0, \sigma_{2\omega})$ because it comes from survey samples (CPS). The variable we use for estimating our model is

$$\tilde{F}_{jgt} \equiv \sum_{a=a0}^{t-g} (\bar{W}_{.g(g+a)} + \omega_{.g(g+a)})F_{j.(g+a)}.$$

So that instead of β we end up estimating

$$\tilde{\beta} = \frac{\sum_g Cov_j(\Delta \ln S_{jgt}, \Delta \ln \tilde{F}_{jgt})}{\sum_g Var_j(\Delta \ln \tilde{F}_{jgt})}.$$

B.2. Simulation strategy

The results of simulations indicate that while ω increases the variance of the right-hand side variable in the classical way, using $\bar{W}_{.gs}$ instead of $\bar{W}_{.gs} + \delta_{.gs}$ actually decreases it. The latter effect, however, depends largely on the time-series properties of δ: the more serially correlated it is the smaller the artificial decrease in the variance. Moreover, δ may well be correlated with either F or S (across states), which induces further complications.

The sizes of the CPS samples provide good estimates for σ_ω^2. Therefore, we can carry out simulations once we specify a data generating process for δ and calibrate its parameters. Two aspects are important: serial correlation and cross-state correlation with the F and S series. We have census-based estimates for W_{jg1970} and W_{jg1980}. Assuming that those are true values without measurement error (the census samples are very large), we can calculate δ_{jg1970} and δ_{jg1980} in such a way that preserves their actual cross-state correlation with F and S for those years. Unfortunately, two points in time do not convey much information about the time-series properties of a series. A further complication arises because δ has to satisfy some additivity constraints. Moreover, it turns out that neither fitting a linear trend on the two points nor fitting

Table 8
Simulation results: average bias of β in the over-time models

Age	Year	Point estimates[a]	Average bias (biased minus simulated "true" estimate) Specification 1[b]			Specification 2[c]		
			Total	Due to ω[d]	Due to $\bar{W}$[d]	Total	Due to ω[d]	Due to $\bar{W}$[d]
25–34	1960–70	0.21	0.00	−0.04	0.04	−0.04	−0.04	0.00
25–34	1970–80	0.10	−0.01	−0.01	−0.01	−0.01	−0.01	0.00
25–34	1980–90	0.60	0.04	−0.05	0.09	0.07	−0.05	0.13
35–44	1970–80	0.26	−0.01	−0.02	0.01	0.10	−0.02	0.12
35–44	1980–90	0.20	−0.05	−0.02	−0.03	0.09	−0.02	0.11
45–54	1980–90	0.06	0.00	−0.01	0.01	0.02	−0.01	0.03[c]

Notes: The figures are average values from 100 Monte Carlo replications.

[a]The point estimates are the biased estimates from the specification used in the simulations. They are slightly different from the original estimates in Table 5 because the latter are based on variables divided by population.

[b]Specification 1: δ is fixed to its 1980 value; $\omega \sim iid\mathrm{N}(0, \sigma_\omega^2)$, $\sigma_\omega^2 = \bar{W}(1 - \bar{W})/(\text{sample size})$. For the definition of δ and ω, see the text above.

[c]Specification 2: δ is fixed to its 1980 value in 1980 and after, fixed to its 1970 value in 1970 and before, and changes in a linear fashion between 1970 and 1980; $\omega \sim iid\mathrm{N}(0, \sigma_\omega^2)$, $\sigma_\omega^2 = \bar{W}(1 - \bar{W})/(\text{sample size})$.

[d]"Due to $\bar{W}$" means bias due to using national averages instead of state-specific age distributions and is based on the assumption that there is no measurement error in $\bar{W}$ (results from runs with $\omega = 0$ everywhere). "Due to ω" means bias due to sampling error in the measures of national averages and is measured as "Total" minus "Due to $\bar{W}$".

an AR(1) process are good specifications because they yield nonsensical values for the 1950s. Instead, we have specified several data generating processes through the available points of δ_{jg1970} and δ_{jg1980}.

Results from two main specifications are presented below. In specification 1, δ is fixed to its 1980 value. In specification 2, δ follows a piecewise linear trend: it is fixed to its 1980 value in 1980 and after, fixed to its 1970 value in 1970 and before, and changes in a linear fashion between 1970 and 1980. We have experimented with adding *iid* disturbances to the trends (the standard error was calibrated to be half of the absolute deviations between the 1980 and 1970 values). They gave results identical to the deterministic specifications up to the fourth decimal point. Also, we have experimented with fixing δ to its 1970 value instead of 1980. The results from these two were essentially the same. Throughout the exercise, ω was calibrated as an *iid* Normal random variable, with variance $\bar{W}(1 - \bar{W})/(\textit{sample size})$.

In each simulation run, we generated δ and ω series, simulated the "true" W using the available national average and δ and ω, and simulated the "true" birth cohort specific flow variables from W and the graduation year-specific flows. Then we estimated the model. Then we took the national average of the simulated "true" W variables, generated the flow variable using the $\bar{W}$, and estimated the model again. These two estimates were then compared to get the bias. Table 8 shows the results.

Our estimate of the bias arising from the measurement error in the national average estimate is always negative, as expected, and is between 1 and 5 percentage points. Our estimate of the bias arising from using the national average rather than state-specific age distributions varies by year, age group, and the method we use to impute δ's for the intercensal years. More often than not, our simulations suggest that using national average data tends to inflate rather than attenuate the estimates. The estimated bias due to the national average varies between -3 and $+13$ percentage points. The result of the two sources is a bias most likely to be between -5 and $+10$ percentage points. More often than not, our simulations suggest that the use of the national average data tends to inflate rather than attenuate the estimates. There is no suggestion from this simulation evidence that our OLS estimates are severely attenuated.

References

Acemoglu, D., Angrist, J., 2001. How large are the social returns to education? Evidence from compulsory schooling laws. In: Bernanke, B., Rogoff, K. (Eds.), NBER Macroeconomics Annual 2000. MIT Press, Cambridge, MA, pp. 9–59.

Alesina, A., Baqir, R., Easterly, W., 1999. Public goods and ethnic divisions. Quarterly Journal of Economics 114, 1243–1284.

Bartik, T., 1991. Who Benefits from State and Local Economic Development Policies? W.E. Upjohn Institute, Kalamazoo, MI.

Blanchard, O., Katz, L., 1992. Regional evolutions. Brookings Papers on Economic Activity 1, 1–61.

Borjas, G.J., Freeman, R., Katz, L., 1997. How much do immigration and trade affect labor market outcomes. Brookings Papers on Economic Activity 1, 1–67.

Bound, J., Holzer, H., 2000. Demand shifts, population adjustments, and labor market outcomes during the 1980s. Journal of Labor Economics 18, 20–54.

Bound, J., Turner, S., 2003. Cohort crowding: how resources affect collegiate attainment. Working Paper, University of Michigan.

Bound, J., Groen, J., Kézdi, G., Turner, S., 2001. Trade in university training: cross-state variation in the production and use of college-educated labor, National Bureau of Economic Research Working Paper No. 8555.

Card, D., Lemieux, T. 2000. Dropout and enrollment trends in the post-war period: what went wrong in the 1970s?. In: Gruber, J. (Ed.), An Economic Analysis of Risky Behavior Among Youth. University of Chicago Press, Chicago, pp. 439–482.

Clotfelter, C., 1976. Public spending for higher education: an empirical test of two hypotheses. Public Finance 31, 174–194.

Davis, D., Weinstein, D.E., Bradford, S.C., Shimpo, K., 1997. Using international and Japanese regional data to determine when the factor abundance theory of trade works. American Economic Review 87, 421–446.

Dickens, W.T., 1990. Error components in grouped data: is it ever worth weighting?. Review of Economics and Statistics 72, 328–333.

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

Glendon, S., 1998. Urban life cycles. Working Paper, Harvard University.

Goldin, C., Katz, L., 1999. The shaping of higher education: the formative years in the United States, 1890 to 1940. Journal of Economic Perspectives 13, 37–62.

Hanson, G., Slaughter, M., 2002. Labor-market adjustment in open economies: evidence from US States. Journal of International Economics 57, 3–29.

Justman, M., Thisse, J., 1997. Implications of the mobility of skilled labor for local public funding of higher education. Economics Letters 55, 409–412.

Katz, L., Murphy, K., 1992. Changes in relative wages, 1963–1987: supply and demand factors. Quarterly Journal of Economics 107, 35–78.

Long, L.E., 1988. Migration and Residential Mobility in the United States. Russell Sage Foundation, New York.

Moretti, E., 2004. Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. Journal of Econometrics, this issue.

Newhouse, J.P., Williams, A.P., Bennett, B.W., Schwartz, W.B., 1982. Does the geographical distribution of physicians reflect market failure? Bell Journal of Economics 13, 493–505.

Rauch, J.E., 1993. Productivity gains from geographic concentration of human capital: evidence from the cities. Journal of Urban Economics 34, 380–400.

Strathman, J.G., 1994. Migration, benefit spillovers and state support of higher education. Urban Studies 31, 913–920.