# Spatial Variation in Higher Education Financing and the Supply of College Graduates

John Kennan*
University of Wisconsin-Madison and NBER

May 2015

**Abstract**

In the U.S. there are large differences across States in the extent to which college education is subsidized, and there are also large differences across States in the proportion of college graduates in the labor force. State subsidies are apparently motivated in part by the perceived benefits of having a more educated workforce. The paper extends the migration model of Kennan and Walker (2011) to analyze how geographical variation in college education subsidies affects the migration decisions of college graduates. The model is estimated using NLSY data, and used to quantify the sensitivity of migration and college enrollment decisions to differences in expected net lifetime income, focusing on how cross-State differences in public college financing affect the educational composition of the labor force. The main finding is that these differences have substantial effects on college enrollment, with no evidence that these effects are dissipated through migration

## 1 Introduction

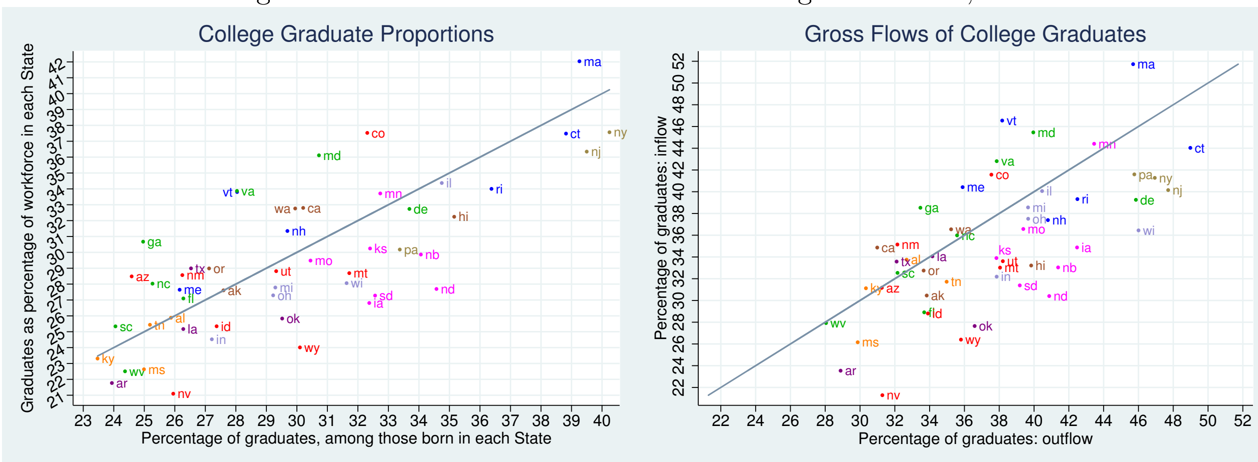
There are substantial differences in subsidies for higher education across States in the U.S. Are these differences related to the proportion of college graduates in each State? If so, why? Do the subsidies change decisions about whether or where to go to college? If State subsidies induce more people to get college degrees, to what extent does this additional human capital tend to remain in the State that provided the subsidy?

There is a considerable amount of previous work on these issues, discussed in Section 3 below. What is distinctive in this paper is that migration is explicitly modeled. Recent work on migration has emphasized that migration involves a sequence of reversible decisions that respond to migration incentives in the face of potentially large migration costs.[1] The results of Kennan and Walker (2011)

*Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706; jkennan@ssc.wisc.edu. Earlier versions of this paper had a different title: "Higher Education Subsidies and Human Capital Mobility." I thank Gadi Barlevy, Eric French, Ahu Gemici, Jim Heckman, Tom Holmes, Lisa Kahn, Maurizio Mazzocco, Derek Neal, Mike Rothschild, Chris Taber, Jean-Marc Robin, Jim Walker, Yoram Weiss and many seminar participants for helpful comments.

[1]See Kennan and Walker (2011),Gemici (2011) and Bishop (2008).

Figure 1: Birth and Work Locations of College Graduates, 2000




indicate that labor supply responds quite strongly to geographical wage differentials and location match effects, in a life-cycle model of expected income maximization. The model is related to earlier work by Keane and Wolpin (1997), who used a dynamic programming model to analyze schooling and early career decisions in a national labor market. Keane and Wolpin (1997) estimated that a \$2000 tuition subsidy would increase college graduation rates by 8.4%. This suggests that variation in tuition rates across States should have big effects on schooling decisions.

This paper considers these effects in a dynamic programming model that allows for migration both before and after acquiring a college degree. In the absence of moving costs, the optimal policy for someone who decides to go to college is to move to the location that provides the cheapest education, and subsequently move to the labor market that pays the highest wage. At the other extreme, if moving costs are very high, the economic incentive to go to college depends only on the local wage premium for college graduates, and estimates based on the idea of a national labor market are likely to be misleading. Thus it is natural to consider college choices and migration jointly in a model that allows for geographical variation in both the costs and benefits of a college degree.

## 2 Geographical Distribution of College Graduates

There are surprisingly big differences across States in the proportion of college graduates among those born in each State, and in the proportion of college graduates among those working in the State. Figure 1 shows the distribution of college graduates aged 25-50 in the 2000 Census, as a proportion of the number of people in this age group working in each State, and as a proportion of the number of workers in this age group who were born in each State. For example, someone who was born in New York is almost twice as likely to be a college graduate as someone born in Kentucky, and someone working in Massachusetts is twice as likely to be a college graduate as someone working in Nevada. Generally, the proportion of college graduates is high in the Northeast, and low in the South.[2]

[2]The colors in Figure 1 (and subsequent figures) represent the nine Census Divisions.

There are also big differences in the proportion of college graduates who stay in the State where they were born. On average, about 45% of all college graduates aged 25-50 work in the State where they were born, but this figure is above 65% for Texas and California, and it is below 25% for Alaska and Wyoming. One might expect that the proportion of college graduates in the flow of in-migrants would be relatively high in States that have relatively few graduates in the native population, and similarly that the proportion of graduates in the flow of out-migrants would high is States that have a high proportion of graduates in the native population. The right panel of Figure 1 show that this is not the case.

States spend substantial amounts of money on higher education, and there are large and persistent differences in these expenditures across States. Figure 2 shows the variation in (nominal) per capita expenditures across States in 1992 and 2012, using data from the Census of Governments. The magnitude of these expenditures suggests that a more highly educated workforce is a major goal of State economic policies, perhaps because of human capital externalities. Thus it is natural to ask whether differences in higher education expenditures help explain the differences in labor force outcomes shown in Figures 1. Figure 3 plots expenditure per student of college age against the proportion of college graduates among those born in each State. There are big variations across States in each of these variables, but there is little apparent relationship between them.

Figure 2: Higher Education Expenditures

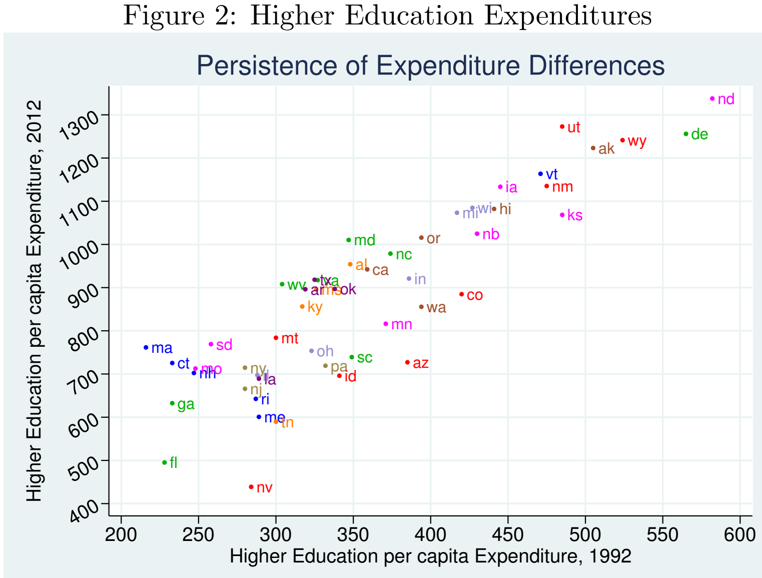
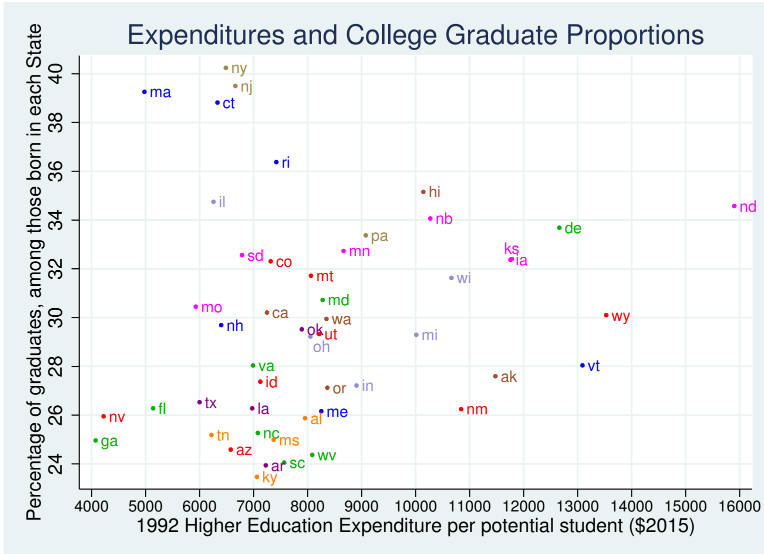


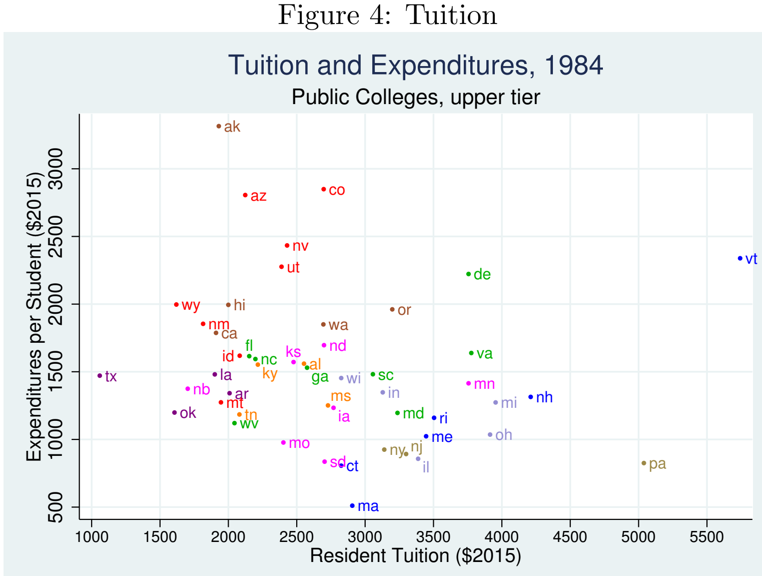
Figure 3: Higher Education Expenditures and Human Capital Distribution




## 2.1 Tuition Differences

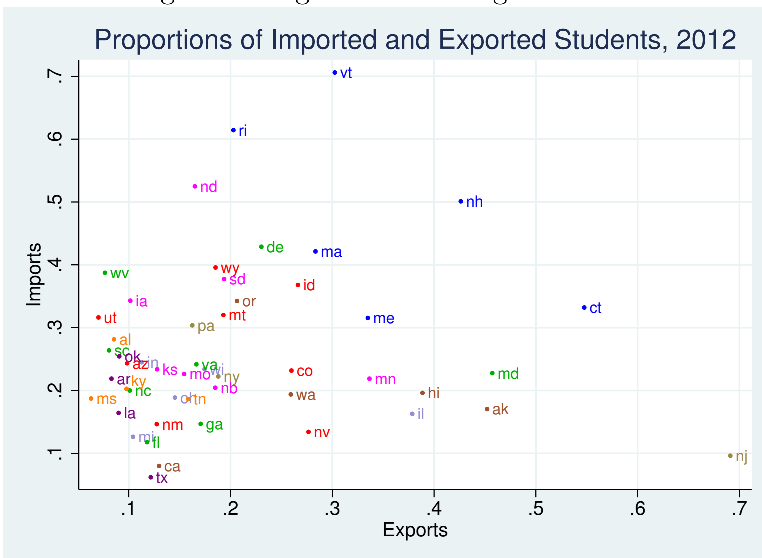
State expenditure on higher education provides a very broad measure of the variation in subsidies, while tuition levels provide a more direct measure of the variation in college costs. Figure 4 shows tuition levels in public universities in 1984 (when the people in the NLSY cohort were aged 19 to 25), plotted against a measure of expenditures per student in these universities. Although differences in tuition levels and expenditures are correlated across States, Figure 4 shows that there is considerable independent variation in these variables.

Figure 4: Tuition




A common assumption in the literature on the relationship between college enrollment and cost

Figure 5: Migration of College Students




is that the relevant measure of tuition is the in-state tuition level, given that most students attend college in their home State. This is a crude approximation. On average, about 20% of college freshmen in 2012 enrolled in an out of State college[3]. Moreover, this proportion varies greatly across States, as shown in Figure 5. At one extreme, the proportion of both imported and exported students was close to 10% for California and Texas.[4] At the other extreme, most of the freshmen in Vermont were not from Vermont, while most students from New Jersey were not studying in New Jersey.

## 2.2 Intergenerational Relationships

One possible explanation for the differences in the proportion of college graduates across States is that there are similar differences across States in the proportion of college graduates in the parents' generation, and there is a strong relationship between the education levels of parents and children. Of course this "explanation" merely shifts the question to the previous generation, but it is still of interest to know whether parental education is enough to account for most of the observed differences in college choices.

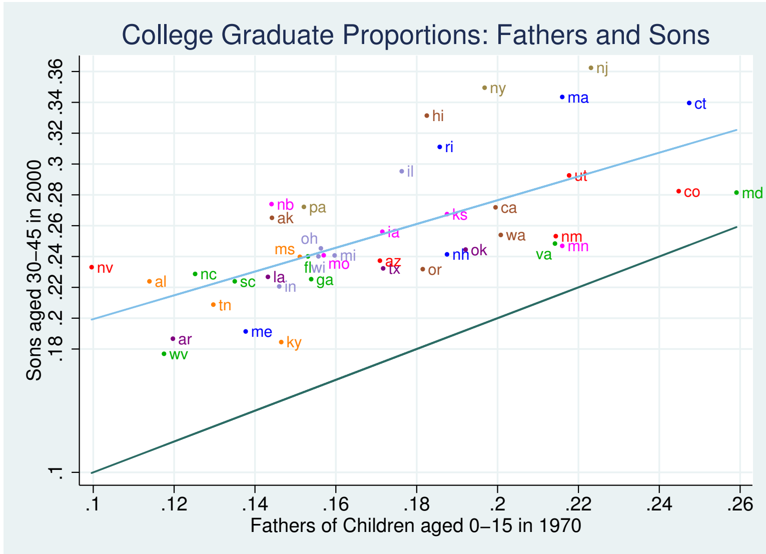
Figure 6 plots the proportion of college graduates by State of birth for men aged 30-45 in the 2000 Census against the proportion of college graduates among the fathers of these men, by State of residence in the 1970 Census. As one might expect, these proportions are quite strongly related: the regression coefficient is .78, and the $R^2$ is .45.[5] The figure includes a 45° line, showing a substantial increase in the proportion of graduates from one generation to the next, and a regression line, showing

[3] See nces.ed.gov/programs/digest/d13/tables/dt13_309.20.asp?current=yes

[4] The proportion of imported students is the number of nonresident students as a fraction of total enrollments in the State, while the proportion of exported students is the number of students from this State attending college out of state, as a proportion of all students from this State.

[5] The inclusion of mother's education levels or of the proportion of fathers who attended college adds almost nothing to this regression.

Figure 6: Intergenerational Relationships




that there is still plenty of inter-State variation in college graduation rates, even after controlling for the proportion of fathers who are college graduates.[6]

# 3 Related Literature

The literature on the effects of State differences in college tuition levels is summarized by Kane (2006, 2007). The "consensus" view is that these effects are substantial – that a \$1,000 reduction in tuition increases college enrollment by something like 5%. Of course a major concern is that the variation in tuition levels across States is not randomly assigned, and there may well be important omitted variables that are correlated with tuition levels.[7] There is no fully satisfactory way to deal with this problem. One approach is to use large changes in the net cost of going to college induced by interventions such as the introduction of the Georgia Hope Scholarship, as in Dynarski (2000), or the elimination of college subsidies for children of disabled or deceased parents, as in Dynarski (2003), or the introduction of the D.C. Tuition Assistance Grant program, as in Kane (2007). Broadly speaking, the results of these studies are not too different from the results of studies that use the cross-section variation of tuition levels over States, suggesting that the endogeneity of tuition levels might not be a major problem. A detailed analysis of this issue would involve an analysis of the political economy of higher education subsidies in general, and of tuition levels in particular. For example, a change in the party controlling the State legislature or the governorship might be associated with a

[6]The interstate differences in the proportions of college graduates in the 1970 Census are determined to a substantial extent by differences in the proportions of high school graduates. For example, 71% of white parents living in Kansas had graduated high school, while in Kentucky only 42% of white parents had graduated high school. In the country as a whole, 23% of the white parents had some college (including college graduates); the figures for Kansas and Kentucky were 26% and 17% . Thus the proportion of high school graduates going to college was actually slightly higher in Kentucky than in Kansas (40.5% vs. 36.9%, the national proportion being 37.5%).

[7]Kane (2006) gives the example of California spending a lot on community colleges while also having low tuition.

change in higher education policies, and the variation induced by such changes might be viewed as plausibly exogenous with respect to college choices, although of course this begs the question of why the political environment changed.

As was shown in Figure 2 above, differences in support for higher education across States are highly persistent in recent years. Goldin and Katz (1999) show that these differences are in fact persistent over a much longer period of time, and they explain why:

> "To sum up, newer states with a high share of well-to-do families and scant presence of private universities in 1900 became the leaders in public higher education by 1930. They remain so today."

As Bound et al. (2004) point out, some of these differences across States might be related to other unmeasured differences in factors affecting college choices. For example, heavy industries requiring a lot of engineers and scientists might be located in places where conditions are favorable in terms of availability of natural resources, but unfavorable in that they happen to be populated by people who are skeptical about the value of higher education. In that case, the business community might push for more investment in public universities, and this would lead to a downward bias in estimates of the response to policy variables. On the other hand, Goldin and Katz (1999) argue that wealthier families are more likely to expect that their children will go to college, and indeed when they use automobiles per capita as a proxy for the level of wealth in the State, they find a positive relationship between wealth and public expenditures on higher education; this would lead to an upward bias in estimates of the response to policy variables. But although bias in one direction or the other cannot be ruled out, it seems reasonable to expect that differences in State policies arising from circumstances that prevailed many years ago would not be strongly related to unmeasured differences in determinants of college choices for recent cohorts (such as the NLSY79 cohort analyzed in this paper).

Card and Lemieux (2001) analyzed changes in college enrollment over the period 1968-1996, using a model of college participation that included tuition levels as one of the explanatory variables. The model includes State fixed effects, and also year fixed effects, so the effect of tuition is identified by differential changes in tuition over time within States – i.e. some States increased their tuition levels more or less quickly than others. The estimated effect of tuition is significant, but considerably smaller than the results in the previous literature (which used cross-section data, so that the effect is identified from differences in tuition levels across States at a point in time).

Card and Krueger (1992) analyzed the effect of school quality using the earnings of men in the 1980 Census, classified according to when they were born, where they were born, and where they worked. An essential feature of this analysis is that the effect of school quality is identified by the presence in the data of people who were born in one State and who worked in another State (within regions, since the model allows for regional effects on the returns to education). This ignores the question of why some people moved and others did not.

Bound et al. (2004) and Groen (2004) sidestep the issue of what causes changes in the number of college graduates in a State, and focus instead on the relationship between the flow of new graduates in a State and the stock of graduates working in that State some time later. They conclude that this relationship is weak, indicating that the scope for State policies designed to affect the educational composition of the labor force is limited.

Keane and Wolpin (2001) estimated a dynamic programming model of college choices, emphasizing the relationship between parental resources, borrowing constraints, and college enrollment (but with no consideration of spatial differences). A major result is that borrowing constraints are binding, and yet they have little influence on college choice. Instead, borrowing constraints affect consumption and work decisions while in college: if borrowing constraints were relaxed, the same people would choose to go to college, but they would work less and consume more while in school.

Aghion et al. (2009) used a set of political instruments to distinguish between arguably exogenous variation in State expenditures on higher education and variation due to differences in wealth or growth rates across States. The model allows for migration, and it considers both innovation and imitation. Higher education investments affect growth in different ways depending on how close a State is to the "technology frontier". Each State is assigned an index measuring distance to the frontier, based on patent data. In States close to the frontier, the estimated effect of spending on research universities is positive, but the estimated effect is negative for States that are far from the frontier. The model that explains this in terms of a tradeoff between using labor to innovate or to imitate.

# 4 A Life-Cycle Model of Expected Income Maximization

The empirical results in Kennan and Walker (2011) indicate that high school graduates migrate across States in response to differences in expected income. This paper analyzes the college choice and migration decisions of high-school graduates, using an extension of the dynamic programming model developed in Kennan and Walker (2011), applied to panel data from the 1979 cohort of the National Longitudinal Survey of Youth. The aim is to quantify the relationship between college choice and migration decisions, on the one hand, and geographical differences in college costs and expected incomes on the other. The model can be used to analyze the extent to which the distribution of human capital across States is influenced by State subsidies for higher education. The basic idea is that people tend to buy their human capital where it is cheap, and move it to where wages are high, but this tendency is substantially affected by moving costs.

Suppose there are $J$ locations, and individual $i$'s income $y_{ij}$ in location $j$ is a random variable with a known distribution. Migration and college enrollment decisions are made so as to maximize the present value of expected lifetime income. Let $x$ be the state vector (which includes the stock of human capital, ability, wage and preference information, current location and age, as discussed below), and let $a$ be the action vector (the location and college enrollment choices). In the basic dynamic discrete choice model[8], the utility flow is specified as $u(x, a) + \zeta_a$, where $\zeta_a$ is a random variable that is assumed to be iid across actions and across periods and independent of the state vector. It is assumed that $\zeta_a$ is drawn from the Type I extreme value distribution. Let $p(x' \mid x, a)$ be

[8] See Rust (1994).

the transition probability from state $x$ to state $x'$, if action $a$ is chosen. The decision problem can be written in recursive form as

$$V(x,\zeta) = \max_a (v(x,a) + \zeta_a)$$

where

$$v(x,a) = u(x,a) + \beta \sum_{x'} p(x' \mid x,a)\bar{v}(x')$$

and

$$\bar{v}(x) = E_\zeta V(x,\zeta)$$

and where $\beta$ is the discount factor, and $E_\zeta$ denotes the expectation with respect to the distribution of the vector $\zeta$ with components $\zeta_a$. Then, using arguments due to McFadden (1974) and Rust (1994), we have

$$\exp(\bar{v}(x)) = \exp(\bar{\gamma}) \sum_{k=1}^{N_a} \exp(v(x,k))$$

where $N_a$ is the number of available actions, and $\bar{\gamma}$ is the Euler constant. Let $\rho(x,a)$ be the probability of choosing $a$, when the state is $x$. Then

$$\rho(x,a) = \exp(v(x,a) + \bar{\gamma} - \bar{v}(x))$$

The function $v$ is computed by value function iteration, assuming a finite horizon, $T$. Age is included as a state variable, with $v \equiv 0$ at age $T+1$, so that successive iterations yield the value functions for a person who is getting younger.

### 4.1 Nested and Sequential Choices

In the basic model, payoff shocks affecting enrollment and migration decisions are drawn independently from the Type I extreme value distribution. This is too restrictive: for example, enrollment choices might be much more predictable than migration choices (or *vice versa*).

Suppose the choices are arranged in an array with $m$ rows corresponding to locations, and $n$ columns corresponding to enrollment decisions. The model associates continuation values $v_{ij}$ with location $i$ and enrollment level $j$, and there are payoff shocks $\zeta_i$ associated with each row, and $\kappa\zeta'_j$ associated with each column, where the shocks are drawn independently from the Type I extreme value distribution, with $\kappa > 0$. Then if row $i$ has been chosen, the column choice is determined by

$$j = \arg\max_k \left(v_{ik} + \zeta_i + \kappa\zeta'_j\right)$$

The probability of choosing column $j$ is

$$\rho(j \mid i) = \frac{\exp\left(\frac{v_{ij}}{\kappa}\right)}{\sum_{k=1}^{n} \exp\left(\frac{v_{ik}}{\kappa}\right)}$$

and the expected value of the row, $\bar{v}_i$, is given by

$$\exp(\bar{v}_i) = \exp(\bar{\gamma}) \left( \sum_{k=1}^{n} \exp\left( \frac{v_{ik}}{\kappa} \right) \right)^{\kappa}$$

If row $i$ is chosen before the column shocks are realized (with the understanding that these shocks will be realized before the column is chosen) then the row choice is determined by

$$i = \arg\max_{s} (\bar{v}_s + \zeta_s)$$

The probability of choosing row $i$ is

$$\rho_0(i) = \frac{\exp(\bar{v}_i)}{\sum_{s=1}^{m} \exp(\bar{v}_s)}$$

and the expected value of the whole array is

$$\begin{aligned}
\bar{v}_0 &= \bar{\gamma} + \log\left( \sum_{s=1}^{m} \exp(\bar{v}_s) \right) \\
&= \bar{\gamma} + \log\left( \sum_{s=1}^{m} \exp(\bar{\gamma}) \left( \sum_{k=1}^{n} \exp\left( \frac{v_{sk}}{\kappa} \right) \right)^{\kappa} \right) \\
&= 2\bar{\gamma} + \log\left( \sum_{i=1}^{m} \left( \sum_{j=1}^{n} \exp\left( \frac{v_{ij}}{\kappa} \right) \right)^{\kappa} \right)
\end{aligned}$$

The choice probabilities are given by

$$\begin{aligned}
\text{Prob}(d_{ij} = 1) &= \rho(j \mid i)\, \rho_0(i) \\
&= \frac{\exp\left(\frac{v_{ij}}{\kappa}\right)}{\sum_{k=1}^{n} \exp\left(\frac{v_{ik}}{\kappa}\right)} \frac{\left( \sum_{k=1}^{n} \exp\left(\frac{v_{ik}}{\kappa}\right) \right)^{\kappa}}{\sum_{s=1}^{m} \left( \sum_{k=1}^{n} \exp\left(\frac{v_{sk}}{\kappa}\right) \right)^{\kappa}}
\end{aligned}$$

If $\kappa = 1$ (or if $m = 1$ or $n = 1$), this reduces to the standard logit formula for the choice probabilities.

The Nested Logit Model discussed by McFadden (1978) gives the same choice probabilities, but with a different interpretation: the continuation value associated with each choice is specified as $v_{ij} + y_{ij}$, where $y_{ij}$ is a generalized extreme value random vector, with joint distribution function

$F(y)$ given by

$$F(y) = \exp\left(-\sum_{i=1}^{m} Y_i\right)$$
$$Y_i^{\frac{1}{\kappa}} = \sum_{j=1}^{n} \exp(-y_{ij})^{\frac{1}{\kappa}}$$

subject to the restriction $0 \leq \kappa \leq 1$ (which ensures that the density function is non-negative).[9] In this interpretation all of the shocks are realized before any choices are made. In the present context, the period length is taken to be a year, and the timing of the location and enrollment choices within the year is necessarily fuzzy, so various interpretations are possible, and each is just a rough approximation of the way that decisions are actually made. The estimated version of the model assumes that location choices are made before enrollment choices (but the reverse ordering gives similar results).

### 4.2 Enrollment Decisions

In simple models of higher education choices, high school graduates choose whether to give up four or five years worth of earnings at high school wages in order to earn a college wage premium for the remaining forty years or so. In practice, the choices are more complicated. While many students enroll in college immediately after finishing high school, and stay in college continuously until they graduate, many others enroll in college after first spending some time in the labor force, or leave college without finishing a degree, either permanently or temporarily, on enroll in two-year colleges, with the possibility of subsequently transferring to a four-year college.[10] Accordingly, the model analyzed here treats college choices as the outcome of a sequence of decisions on whether to enroll in one of several types of college, with uncertainty about whether enrollment will lead to graduation with a degree.

The specification of the model involves the usual tradeoff between realism and computational difficulty; in particular, since there are many locations, and location is an essential state variable, it is necessary to use a relatively coarse specification of the other state variables so that the state space does not become too big. For this reason, there are just three levels of schooling: high school (12 or 13 years of schooling completed), some college (14 or 15 years) and college graduate (16 years or more).

In each period, there is a choice of whether to enroll in college. There are four types of college: community colleges, other public colleges and universities, and private colleges at two quality levels. The college types differ with respect to tuition, State subsidies, financial aid, graduation probabilities, and psychic costs and benefits. Enrollment choices are influenced by ability, parental schooling and family income, represented by permanent state variables, which are restricted to just two values,

[9] See Börsch-Supan (1990)

[10] Agan (2014) presents a detailed description of the various paths taken by college students, using NLSY79 data.

high or low.[11]

### 4.3 Wages

The wage of individual $i$ in location $j$ at age $g$ in year $t$ is specified as

$$w_{ij} = \mu_j(e_i) + \upsilon_{ij}(e_i) + G(e_i, X_i, g_i) + \varepsilon_{ij}(e) + \eta_i$$

where $e$ is schooling level, $\mu_j$ is the mean wage in location $j$ (for each level of schooling), $\upsilon$ is a permanent location match effect, $G(e, X, g)$ represents the effects of observed individual characteristics, $\eta$ is an individual effect that affects wages in the same way in all locations, and $\varepsilon$ is a transient effect. The random variables $\eta$, $\upsilon$ and $\varepsilon$ are assumed to be independently and identically distributed across individuals and locations, with mean zero. It is also assumed that the realizations of $\upsilon$ and $\eta$ are seen by the individual (although $\upsilon_{ij}(e_i)$ is seen by individual $i$ only after moving to location $j$ with education level $e_i$).

The function $G$ is specified as a piecewise-quadratic function of age, with an interaction between ability and education:

$$G(e, b, g) = \begin{cases} \theta_e b + y_e^* - c_e (g - g_e^*)^2 & g \leq g_e^* \\ \theta_e b + y_e^* & g \geq g_e^* \end{cases}$$

where $b$ is measured ability, $y_e^*$ is the peak wage for education level $e$, and $g_e^*$ is age at the peak. Thus both the shape of the age-earnings profile and the ability premium are specified separately for each level of education, with four parameters to be estimated ($\theta_e, y_e^*, c_e,$ and $g_e^*$).

The relationship between wages and actions is governed by the difference between the quality of the match in the current location, measured by $G(e, b, g) + \mu_j(e) + \upsilon_{ij}(e)$, and the prospect of obtaining a better match in another location or at a higher level of schooling. The other components of wages have no bearing on migration or college choice decisions, since they are added to the wage in the same way no matter what decisions are made.

#### 4.3.1 Stochastic Wage Components

Since the realized value of the location match component $\upsilon$ is a state variable, it is convenient to specify this component as a random variable with a discrete distribution, and compute continuation values at the support points of this distribution. For given support points, the best discrete approximation $\hat{F}$ for any distribution $F$ assigns probabilities so as to equate $\hat{F}$ with the average value of $F$ over each interval where $\hat{F}$ is constant. If the support points are variable, they are chosen so that $\hat{F}$ assigns equal probability to each point.[12] Thus if the distribution of the location match component $\upsilon$ were known, the wage prospects associated with a move to State $k$ could be represented by an

[11] Again, binary state variables are used here in order to keep the state space manageable.

[12] See Kennan (2006)

$n$-point distribution with equally weighted support points $\hat{\mu}_k + \hat{\upsilon}(q_r), 1 \le r \le n$, where $\hat{\upsilon}(q_r)$ is the $q_r$ quantile of the distribution of $\upsilon$, with

$$q_r = \frac{2r-1}{2n}$$

for $1 \le r \le n$. The distribution of $\upsilon$ is in fact not known, but it is assumed to be symmetric around zero. Thus for example with $n = 3$, the distribution of $\mu_j + \upsilon_{ij}$ in each State for each education level is approximated by a distribution that puts mass $\frac{1}{3}$ on $\mu_j$ (the median of the distribution of $\mu_j + \upsilon_{ij}$ ), with mass $\frac{1}{3}$ on $\mu_j \pm \upsilon^0$ , where $\upsilon^0$ is a parameter to be estimated.

Measured earnings in the NLSY are highly variable, even after controlling for education and ability. Moreover, while some people have earnings histories that are well approximated by a concave age-earnings profile, others have earnings histories that are quite irregular. In other words, the variability of earnings over time is itself quite variable across individuals. It is important to use a wage components model that is flexible enough to fit these data, in order to obtain reasonable inferences about the relationship between measured earnings and the realized values of the location match component. The fixed effect $\eta$ is assumed to be uniformly and symmetrically distributed around zero, with three points of support, so that there is one parameter to be estimated. The transient component $\varepsilon$ should be drawn from a continuous distribution that is flexible enough to account for the observed variability of earnings. It is assumed that $\varepsilon$ is drawn from a normal distribution with zero mean for each person, but with a variance that differs across people. Specifically, person $i$ initially draws $\sigma_\varepsilon(i)$ from a uniform discrete distribution with two support points (which are parameters to be estimated), and subsequently draws $\varepsilon_{it}$ from a normal distribution with mean zero and standard deviation $\sigma_\varepsilon(i)$, with $\varepsilon_{it}$ drawn independently in each period.

## 4.4 State Variables and Flow Payoffs

Let $\ell = (\ell^0, \ell^1)$ denote the current and previous location, let $\omega$ be a vector recording wage information at these locations, and let $\xi$ denote current enrollment status (with the convention that $\xi = 0$ means that the individual is not enrolled in college, and otherwise $\xi$ represents the college type). The state vector $x$ consists of $\ell$, $\omega$, education level achieved so far, ability, parental education, family income, home location and age.[13] The flow payoff may be written as

$$\tilde{u}_h(x,a) = u_h(x,a) + \zeta_a$$

where $h$ is the home location, and $u_h(x,a)$ represents the payoffs associated with observed states and choices, and $\zeta_a$ represents the unobserved component of payoffs.

The systematic part of the flow payoff is specified as

[13] As in Kennan and Walker (2011), a limited (location) history approximation is used to reduce the size of the state space in a way that takes advantage of the low migration rates seen in the data.

$$u_h(x,j) = \alpha_0(e) + \alpha_1 w\left(g,e,b,\ell^0,\omega,\xi\right) + \alpha_2 Y\left(\ell^0\right) + \alpha^H \chi\left(\ell^0 = h\right) - C_h\left(\ell^0,\xi\right) - \Delta(x,j)$$

Here the first term refers to consumption values associated with different education levels. The second term refers to wage income in the current location (which depends on age, schooling and ability, as discussed above). This is augmented by the amenity variable $Y\left(\ell^0\right)$. The parameter $\alpha^H$ is a premium that allows each individual to have a preference for their home location ($\chi$ denotes an indicator). The cost of attending a college of type $\xi$ in location $\ell$ for a person whose home location is $h$ is denoted by $C_h(\ell,\xi)$. The cost of moving from $\ell^0$ to $\ell^j$ is represented by $\Delta(x,j)$.

### 4.5 College Costs

Aside from consumption values and expected income, all of the variables in the model that affect college choices do so by changing the costs associated with being in college. Earnings while enrolled in college are ignored. The college cost depends on ability, $b$, age, $g$, relative to an initial age $g_0$ which is set to 19. The cost also depends on resident and nonresident tuition rates, $\tau_r(\ell,\xi)$ and $\tau_n(\ell,\xi)$, expenditure on higher education, $y(\ell,\xi)$, financial aid (scholarships), $s(\ell,\xi)$, and on parents' education and family income. Let $d_m$ and $d_f$ be indicators of whether the mother and the father have some college education, and let $y_f$ be an indicator of whether family income is high or low. Let $\Xi$ be the set of upper-tier colleges. The cost of attending a college of type $\xi$ is specified as

$$\begin{aligned} C(\ell,\xi) \quad = \quad & \delta_0(\xi) + \delta_1(\xi)\,\tau(\ell,\xi) - \delta_2(\xi)\,y(\ell,\xi) - \delta_3 b - \delta_4 b\chi_{\xi\in\Xi} - \delta_5 d_m - \delta_6 d_f - \delta_7 y_f + \delta_8(g-g_0) \\ & - \left(\delta_9(\xi) + \delta_{10} b + \delta_{11} d_m + \delta_{12} d_f - \delta_{13} y_f\right) s(\ell,\xi) \end{aligned}$$

where tuition is given by

$$\tau(\ell,\xi) \quad = \quad \chi(\ell = h)\,\tau_r(\ell,\xi) + \chi(\ell \neq h)\,\tau_n(\ell,\xi)$$

(with $\tau_r = \tau_n$ for private colleges).[14] For each college type $\xi$, $\delta_0(\xi)$ measures the disutility of the effort involved in taking college courses (offset by the utility of life as a student); the effort cost depends on ability ($\delta_3$), especially in upper-tier colleges ($\delta_4$).[15] The tuition measures are averages

[14] The sign convention used here is that each parameter is likely to be positive; for example, given measured ability and family income, it is anticipated that parental education is positively associated with a student's academic achievement, in which case the parameters $\delta_{11}$ and $\delta_{12}$ are positive.

[15] In general it is not possible to distinguish between the nonpecuniary costs of college ($\delta_0$) and the nonpecuniary benefits of having a college education ($\alpha_0$). The income coefficient is identified by the migration component of the model. So the proportion who would choose college is known if there is no college cost, and if there is no difference between education levels except that college graduates earn more. Suppose the prediction is that the proportion going to college is 80%, and suppose that only 30% choose college in the data. The model might explain this by saying that going to college is costly. Alternatively, it might be explained by saying that there are nonpecuniary payoffs associated with the different education levels. The specification of costs and returns used here imposes an exclusion restriction

over each college type within a State; it is assumed that the actual net tuition is a linear function of the State average tuition measures, and $\delta_1(\xi)$ represents the slope of this function, for each college type. Similarly, for each college type, the parameter $\delta_2(\xi)$ measures the extent to which higher education expenditures reduce the cost of college, without specifying any particular channel through which this effect operates. The effect of scholarships is also measured separately for each college type, and in addition it depends on ability, parental education, and family income. The point here is that scholarships are largely allocated on the basis of merit or need; a college that has a large scholarship budget is more attractive (given tuition and expenditure levels), but the size of the scholarship budget is obviously more relevant for students who are more likely to be eligible for scholarships.

### 4.6 Moving Costs

Let $D(\ell^0, j)$ be the distance from the current location to location $j$, and let $\mathbb{A}(\ell^0)$ be the set of locations adjacent to $\ell^0$ (where States are adjacent if they share a border). The moving cost is specified as

$$\Delta(x,j) = \left(\gamma_0(e) + \gamma_1 D(\ell^0, j) - \gamma_2 \chi\left(j \in \mathbb{A}(\ell^0)\right) - \gamma_3 \chi\left(j = \ell^1\right) + \gamma_4 g - \gamma_5 n_j\right) \chi\left(j \neq \ell^0\right)$$

Thus the moving cost varies with education. The observed migration rate is much higher for college graduates than for high school graduates, and the model can account for this either through differences in potential income gains or differences in the cost of moving. The moving cost is an affine function of distance (which is measured as the great circle distance between population centroids). Moves to an adjacent location may be less costly (because it is possible to change States while remaining in the same general area). A move to a previous location may also be less costly, relative to moving to a new location. In addition, the cost of moving is allowed to depend on age, $g$. Finally, it may be cheaper to move to a large location, as measured by population size $n_j$.

### 4.7 Transition Probabilities

The state vector can be written as $x = (\tilde{x}, g)$, where $\tilde{x} = \left(e, \ell^0, \ell^1, x_v^0\right)$ and where $x_v^0$ indexes the realization of the location match component of wages in the current location. Let $q(e, \xi)$ denote the probability of advancing from education level $e$ to $e+1$, for someone who is enrolled in a college of type $\xi$, with $q(e, 0) = 0$ for someone who is not enrolled, and let $a = (j, \xi)$. The transition

that distinguishes one from the other: the transition probabilities are more favorable for high-ability people, but the nonpecuniary benefits of having a college education are the same for both types. This assumption is arbitrary. But the main point of the model is not to make these distinctions, but rather to estimate the responses to changes in the policy variables.

probabilities are as follows

$$p\left(x' \mid x\right) = \begin{cases} q\left(e,\xi\right) & \text{if} \quad j = \ell^0, \quad \tilde{x}' = \left(e+1, \ell^0, \ell^1, x_v^0\right), \quad g' = g+1 \\ 1 - q\left(e,\xi\right) & \text{if} \quad j = \ell^0, \quad \tilde{x}' = \left(e, \ell^0, \ell^1, x_v^0\right), \quad g' = g+1 \\ q\left(e,\xi\right) & \text{if} \quad j = \ell^1, \quad \tilde{x}' = \left(e+1, \ell^1, \ell^0, s_v\right), \quad g' = g+1, \quad 1 \leq s_v \leq n_v \\ 1 - q\left(e,\xi\right) & \text{if} \quad j = \ell^1, \quad \tilde{x}' = \left(e, \ell^1, \ell^0, s_v\right), \quad g' = g+1, \quad 1 \leq s_v \leq n_v \\ \frac{q(e,\xi)}{n} & \text{if} \quad j \notin \left\{\ell^0, \ell^1\right\}, \quad \tilde{x}' = (e+1, j, \ell^0, s_v), \quad g' = g+1, \quad 1 \leq s_v \leq n_v \\ \frac{1-q(e,\xi)}{n} & \text{if} \quad j \notin \left\{\ell^0, \ell^1\right\}, \quad \tilde{x}' = (e, j, \ell^0, s_v), \quad g' = g+1, \quad 1 \leq s_v \leq n_v \\ 0 & \text{otherwise} \end{cases}$$

# 5 Empirical Results

## 5.1 Data

The primary data source is the National Longitudinal Survey of Youth 1979 Cohort (NLSY79); data from the Census of Population are used to estimate State mean wages and parental income and education distributions, and data from the Integrated Postsecondary Education Data System (IPEDS) are used to measure tuition and college expenditures and financial aid. The NLSY79 conducted annual interviews from 1979 through 1994, and changed to a biennial schedule in 1994. The location of each respondent is recorded at the date of each interview, and migration is measured by the change in location from one interview to the next. Only the migration information from 1979 through 1994 is used here, but wage information is available (biennially) through 2009, and this is used in order to obtain better estimates of the lifetime wage profile.

In order to obtain a relatively homogeneous sample, only white non-Hispanic male high school graduates (or GED recipients) are included, using only the years after schooling is completed; the analysis begins at age 19. The (unbalanced) sample includes 12,895 annual observations on 1,281 men. Summary statistics on college enrollment for this sample are shown in Table 1.

Wages are measured as total wage and salary income, plus farm and business income, adjusted for cost of living differences across States (using the ACCRA Cost of Living Index). The State effects $\{\mu_j(e)\}$ are obtained from 1990 Census data, using median wage regressions with age and State dummies, applied to white males who have recently entered the labor force (so as to avoid selection effects due to migration).

### 5.1.1 Tuition and Subsidies

In the model, each State has one representative college of each type[16], and all of these colleges are available choices for everyone[17]. Tuition rates were estimated by computing enrollment-weighted

[16] There are a few exceptions: there are no private colleges in Wyoming (aside from Wyoming Technical Institute, a for-profit operation of dubious repute), and there are no upper-tier private colleges in Montana, Nevada and South Dakota. Thus these alternatives are excluded from the choice set in the dynamic programming model.

[17] This does not involve an assumption that every high school graduate is free to choose Harvard. There are 43 colleges in Massachusetts that are classified as upper-tier (including Harvard), and the assumption is that every high school graduate can get into at least one of these colleges.

**Table 1: College Enrollment, NLSY**

| | | |
|---|---|---|
| **Enrollment Counts** | | |
| Public low | 469 | 17% |
| Public high | 1,497 | 53% |
| Private low | 138 | 5% |
| Private high | 737 | 26% |
| Subtotal | 2,841 | |
| Average years enrolled | 3.7 | |
| Not enrolled | 10,054 | |
| Total (person-years) | 12,895 | |
| **Ever enrolled in college** | | |
| No | 523 | 41% |
| In-State only | 565 | 44% |
| Out-of-State only | 98 | 8% |
| Both | 95 | 7% |
| Total (persons) | 1,281 | |

averages of "sticker prices" for each college type, using IPEDS data for 1984. Students attending college in their home State are assumed to pay tuition at the resident rate, while others pay the non-resident rate (allowing for a few reciprocity agreements across States)[18]. The home State is defined as the State in which the individual last went to high school.

State subsidies to higher education might affect either the cost or the quality of education. For example, given the level of tuition, the cost of attending college is lower if there is a college within commuting distance, and the cost of finishing college is higher if graduation is delayed due to bottlenecks in required courses. From the point of view of an individual student, an increase in tuition paid by other students has much the same effect as an increase in subsidies, in the sense that it increases the resources available for instruction and student support services. But because tuition also acts as a price, it seems more informative to model the effect of direct subsidies, holding tuition constant. This means that the effect of tuition should not be interpreted as a movement along a demand curve, since a college that charges high tuition, holding subsidies constant, can use the additional tuition revenue to improve the quality of the product, or to reduce other components of college costs.

Subsidy measures were constructed by adding federal, State and local appropriations and grants over all public colleges in the 1984 IPEDS file, by State, and by college level, the lower level being defined as community colleges, and the upper level as all other public colleges.[19] Similarly, the financial aid variables measure total expenditures on scholarships, by State and college level[20]. Since

[18]Minnesota has tuition reciprocity agreements with Wisconsin and with North and South Dakota; there is a similar agreement between Oregon and Washington State.

[19]These data can be found at nces.ed.gov/ipeds/datacenter/Default.aspx

[20]These data were obtained from the IPEDS finance file for 1984 (nces.ed.gov/ipeds/datacenter/data/F1984_Data_Stata.zip); the expenditure variable includes expenditures on Instruction, Research, Public service, Academic support (excluding libraries), Student services, Institution support, and Educational Mandatory Transfers. The financial aid variable

Table 2: **Wage Differentials and College Costs**

| | Mean | S.D. | Min | Max |
|---|---|---|---|---|
| **Earnings** ($1983) | | | | |
| High School (age 20) | 7,856 | 871 | 5,824 | 10,196 |
| Some College (age 22) | 9,966 | 982 | 7,451 | 11,809 |
| College Graduate (age 24) | 13,984 | 1,271 | 9,345 | 16,174 |
| **Tuition** | | | | |
| Public, low, Resident | 663 | 280 | 86 | 1,422 |
| Public, low, Nonresident | 1,830 | 738 | 555 | 3,742 |
| Public, low, Resident | 1,224 | 398 | 471 | 2,553 |
| Public, high, Nonesident | 3,166 | 903 | 1,532 | 6,181 |
| Private, Low | 3,767 | 927 | 1,438 | 5,749 |
| Private, High | 5,197 | 1,765 | 1,518 | 9,166 |
| **Expenditure** (per potential student) | | | | |
| Public, low | 111 | 92 | 13 | 402 |
| Public, high | 679 | 252 | 227 | 1,474 |
| Private, Low | 54 | 53 | 2 | 311 |
| Private, High | 218 | 224 | 2 | 898 |
| **Financial Aid** (per potential student) | | | | |
| Public, low | 13.0 | 8.8 | 2.8 | 41.0 |
| Public, high | 51.6 | 24.8 | 16.3 | 149.3 |
| Private, Low | 13.1 | 10.9 | 0.7 | 59.2 |
| Private, High | 33.2 | 30.5 | 0.1 | 136.3 |

these expenditure aggregates involve populations of very different sizes, the expenditure and financial aid figures are divided by the number of potential students, measured as the number of high school graduates in the State aged 22-36 in the 1990 Census. Summary statistics are shown in Table 2

### 5.1.2 College Choices

As is well known, there is a very strong relationship between college choices and parental education levels. For the sample used here, this relationship is summarized in Table 3, for low-ability and high-ability students, where the ability measure is an indicator of whether the AFQT percentile score is above or below the median in the full sample (which is 63).

For example, if both parents went to college, there is a 52% chance that their sons will graduate from college, and this rises to 64% if the son is in the top half of the distribution of AFQT scores. There is also a strong relationship between AFQT scores and college choices, but note that sons whose parents went to college are much more likely to have high AFQT scores.

Transition rates for the NLSY sample are shown in Table 4. These are treated as transition probabilities, and held fixed when the the model is estimated.

includes Scholarships (unrestricted) and Scholarships (restricted).

Table 3: **Ability, Parents' Education and Schooling**

| **Neither** Parent went to College | | | | |
|---|---|---|---|---|
| | High School | Some College | College | Total |
| Years | 12-13 | 14-15 | 16+ | |
| Low Ability | 375<br>**84.8%** | 33<br>7.5% | 34<br>7.7% | 442<br>62.3% |
| High Ability | 128<br>47.8% | 56<br>20.9% | 84<br>31.3% | 268<br>37.7% |
| Total | 503<br>70.8% | 89<br>12.5% | 118<br>16.6% | 710 |
| **Both** Parents went to College | | | | |
| Low Ability | 41<br>51.9% | 19<br>24.1% | 19<br>24.1% | 79<br>29.7% |
| High Ability | 24<br>12.8% | 44<br>23.5% | 119<br>**63.6%** | 187<br>70.3% |
| Total | 65<br>24.4% | 63<br>23.7% | 138<br>51.9% | 266 |

Table 4: **College Transition Rates**

| | | | **Low AFQT** | | **High AFQT** | |
|---|---|---|---|---|---|---|
| | **Initial Grade** | | 12-13 | 14-15 | 12-13 | 14-15 |
| | $e$ | | 0 | 1 | 0 | 1 |
| | **Next Grade** | | 14-15 | 16 | 14-15 | 16 |
| | $e$ | | 1 | 2 | 1 | 2 |
| | | $\xi$ | | | | |
| Public | Lower-Tier | 1 | 24.4% | 12.8% | 32.9% | 6.3% |
| Public | Upper-Tier | 2 | 45.1% | 35.6% | 56.7% | 34.6% |
| Private | Lower-Tier | 3 | 44.4% | 18.2% | 62.9% | 41.7% |
| Private | Upper-Tier | 4 | 41.3% | 29.5% | 57.5% | 35.7% |

## 5.2 College Choices and Migration

Table 5 gives the main empirical results. The parameters of the wage process are estimated separately, using the most recent data (including the biennial interviews)[21]; these parameters, which are shown in the right panel of Table 5, are treated as known when estimating the other parameters governing college choice and migration decisions[22]. The estimates of the parameters governing migration decisions are similar to the estimates in Kennan and Walker (2011). The estimated income coefficient in this model reflects both migration and college choice decisions; as in the migration model, the effect is highly significant. Ability and parental education levels have strong effects on college costs (as would be expected, given the data in Table 3). The estimated moving costs are decreasing in the level of education, reflecting the positive relationship between education and migration rates in the data. The value of $\kappa$ indicates that migration decisions are substantially less predictable the enrollment decisions.

[21]The wage unit is $\$10,000$ (at 1983prices).

[22]Surprisingly, the direct effect of the (binary) AFQT score is very weak (conditional on the education level). Ability is of course strongly correlated with earnings, but the estimated earnings process attributes this almost entirely to a strong relationship between ability and educational attainment.

Table 5: **College Location Choice and Migration, White Males**

| Utility Parameters | | | |
|---|---|---|---|
| | | $\hat{\theta}$ | $\hat{\sigma}_\theta$ |
| Moving Cost: HS | $\gamma_0$ (1) | 4.427 | 0.273 |
| Moving Cost: SC | $\gamma_0$ (2) | 3.864 | 0.294 |
| Moving Cost: CG | $\gamma_0$ (3) | 3.555 | 0.304 |
| Distance | $\gamma_1$ | 0.353 | 0.064 |
| Adjacent Location | $\gamma_2$ | 0.930 | 0.084 |
| Home Premium | $\alpha^H$ | 0.130 | 0.007 |
| Previous Location | $\gamma_3$ | 2.267 | 0.115 |
| Age effect (moving cost) | $\gamma_4$ | 0.094 | 0.011 |
| Population | $\gamma_5$ | 0.864 | 0.065 |
| Climate | $\alpha_2$ | 0.010 | 0.003 |
| Income | $\alpha_1$ | 0.100 | 0.010 |
| Disutility, college: Pub lo | $\delta_0$ (1) | 0.422 | 0.089 |
| Disutility, college: Pub hi | $\delta_0$ (2) | -0.146 | 0.131 |
| Disutility, college: Pvt lo | $\delta_0$ (3) | 0.956 | 0.183 |
| Disutility, college: Pvt hi | $\delta_0$ (4) | 0.380 | 0.143 |
| Age effect (college cost) | $\delta_8$ | 0.131 | 0.013 |
| Nonpecuniary value, SC | $\alpha_0$ (1) | -0.175 | 0.025 |
| Nonpecuniary value, CG | $\alpha_0$ (2) | -0.256 | 0.045 |
| Mother's education | $\delta_5$ | 0.003 | 0.023 |
| Father's education | $\delta_6$ | 0.175 | 0.029 |
| Family Income | $\delta_7$ | 0.084 | 0.022 |
| Ability effect on cost | $\delta_3$ | 0.461 | 0.054 |
| Ability×upper tier | $\delta_4$ | 0.237 | 0.036 |
| Spend/Student: Pub lo | $\delta_2$ (1) | 14.774 | 2.050 |
| Spend/Student: Pub hi | $\delta_2$ (2) | 0.141 | 0.450 |
| Spend/Student: Pvt lo | $\delta_2$ (3) | -6.499 | 13.772 |
| Spend/Student: Pvt hi | $\delta_2$ (4) | 4.962 | 1.770 |
| Tuition: Public lo | $\delta_1$ (1) | 1.021 | 0.255 |
| Tuition: Public hi | $\delta_1$ (2) | 1.295 | 0.162 |
| Tuition: Private lo | $\delta_1$ (3) | -0.783 | 0.276 |
| Tuition: Private hi | $\delta_1$ (4) | 0.057 | 0.135 |
| Financial Aid: Pub lo | $\delta_9$ (1) | 5.480 | 19.544 |
| Financial Aid: Pub hi | $\delta_9$ (2) | -1.935 | 5.880 |
| Financial Aid: Pvt lo | $\delta_9$ (3) | 88.781 | 54.527 |
| Financial Aid: Pvt hi | $\delta_9$ (4) | -14.863 | 12.836 |
| Ability×aid | $\delta_{10}$ | 27.379 | 5.206 |
| Mother-ed×aid | $\delta_{11}$ | 26.920 | 5.128 |
| Father-ed×aid | $\delta_{12}$ | -1.936 | 4.374 |
| Family income×aid | $\delta_{13}$ | -2.437 | 4.063 |
| Enroll/migrate shocks | $\kappa$ | 0.337 | 0.033 |
| Loglikelihood | | -22770.3 | |

| | Wage Parameters | | | | | |
|---|---|---|---|---|---|---|
| | High School | | Some College | | College | |
| | $\hat{\theta}$ | $\hat{\sigma}_\theta$ | $\hat{\theta}$ | $\hat{\sigma}_\theta$ | $\hat{\theta}$ | $\hat{\sigma}_\theta$ |
| Peak Wage | 1.752 | 0.013 | 2.393 | 0.047 | 2.666 | 0.061 |
| Age at Peak | 38.579 | 0.350 | 47.282 | 1.404 | 52.832 | 1.370 |
| Curvature | 1.185 | 0.046 | 0.850 | 0.088 | 0.954 | 0.075 |
| AFQT | 0.097 | 0.021 | -0.188 | 0.043 | 0.083 | 0.036 |
| Location match | 0.389 | 0.008 | 0.772 | 0.021 | 0.815 | 0.015 |
| Transient s.d. 1 | 0.524 | 0.002 | 0.676 | 0.006 | 0.672 | 0.005 |
| Transient s.d. 2 | 2.005 | 0.009 | 2.900 | 0.036 | 3.639 | 0.033 |
| Individual Effect | | 1.011 | | | 0.010 | |

For public colleges, higher tuition has a strong negative effect on enrollment, and expenditure per (potential) student has a positive effect, and these effects are stronger for community colleges that for other public colleges. There is considerable variation in tuition levels for private colleges, but since this variation is not determined at the State level, the effect of differences in private college tuition cannot easily be inferred from location choices, as is done here for public colleges.

## 6 Effects of Changes in Tuition Levels and State Expenditures

The results in Table 5 indicate that college enrollment decisions are affected by tuition and expenditure and financial aid levels, while expected income differences affect both enrollment and migration decisions. This raises the question of whether changes in State policies regarding tuition and expenditures have long-term effects on the educational composition of the State's labor force, as opposed to transient effects that are undone by migration, as suggested by Bound et al. (2004). The main point of the model is that it can answer questions of this kind.

Suppose for example that Michigan reduces tuition, or increases expenditures. The effects of such changes are presumably small for high school graduates in Alaska or Louisiana, but perhaps not so small for students from Michigan, or neighboring States. Moreover, the effects depend on individual characteristics. The model has 800 types, classified by State, and by binary measures of ability, family income, and parental education. In order to estimate the effects of changes in college costs (or wages) it is necessary to use the value functions for each of these types to compute the responses for each type, and then construct a suitably weighted average over types. The main complication here is that parental education and family income vary considerably across States. In order to deal with this, data from the 1970 Census were used to identify households with children aged $5-13$ (corresponding to the ages of the individuals in the NLSY data), and the family income and parental education data for these children were then tabulated, by State. Results of these tabulations for large states are shown in Table 6, and the cross-State dispersion in parental education and family income is shown in Figure 7.[23]

The evolution of the population distribution in the model is computed by iterating the transition matrix of the Markov chain on the state space. The model specifies choice probabilities $\rho(a,x)$, where $x$ is the state vector, and $a$ is the choice variable; the next state $x'$ is then determined by the transition probabilities $q(x,a,x')$. There is a frequency distribution $p(x)$ over current states, and the model implies a transition matrix $T$ from $p(x)$ to $p'(x)$ given by

$$T(p)(x) = \sum_{t\in X} p(t) \sum_{a\in A} \rho(a,t)\, q(t,a,x)$$

The effects of changes in the policy variables are computed by first iterating the transition matrix implied by the values of the policy variables used in the estimation, and then doing the same thing

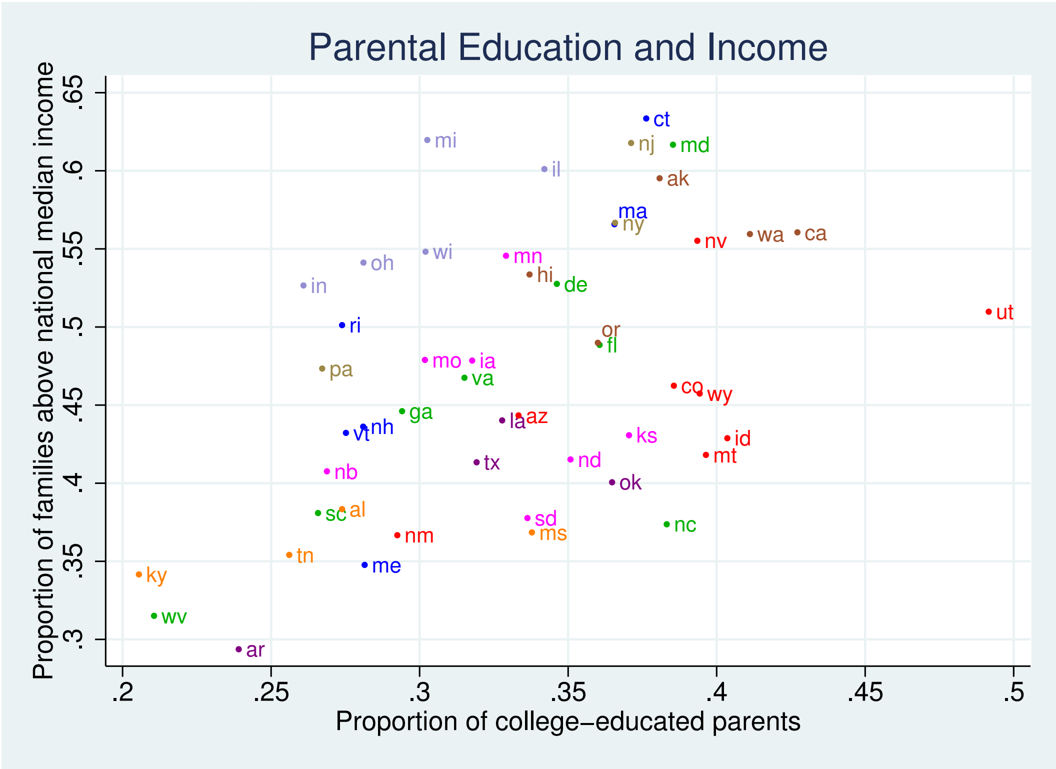
[23]It is assumed that the proportion of high-ability types is the same in all States.

Table 6: **Parental Education: Large States**

| **Parental Education and Family Income by State** | | | | | |
|---|---|---|---|---|---|
| Parents of NLSY79 Cohort | | | | | |
| | Neither | Father only | Mother only | Both | High Income |
| California | 57.3% | 17.7% | 8.0% | 17.1% | 56.1% |
| Florida | 63.9% | 15.3% | 7.2% | 13.6% | 48.8% |
| Illinois | 65.8% | 14.1% | 6.1% | 14.0% | 60.1% |
| Massachusetts | 63.4% | 15.7% | 6.7% | 14.1% | 56.6% |
| Michigan | 69.7% | 13.4% | 5.6% | 11.3% | 62.0% |
| New York | 63.4% | 15.8% | 6.1% | 14.6% | 56.7% |
| Pennsylvania | 73.3% | 13.3% | 4.2% | 9.2% | 47.3% |
| Texas | 68.1% | 13.6% | 5.2% | 13.1% | 41.3% |
| Wisconsin | 69.8% | 12.3% | 6.3% | 11.6% | 54.8% |
| U.S. | 66.9% | 13.9% | 6.1% | 13.0% | 50.0% |

Note: The first four columns show the proportion of parents who attended college. The last column shows the proportion of households with income above the national median.

Figure 7: Parental Education and Family Income



for the new values of the policy variables, and comparing the population distributions.[24].

Table 7: **Effects of Policy Changes: Michigan**

| | | Population at Age 36 | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Current Location** | | | **Home Location** | | |
| **Increase (20%)** | | Graduates | Some College | High School | Graduates | Some College | High School |
| **Tuition** | Michigan | **-6.5%** | **0.0%** | **3.9%** | **-7.3%** | **-0.1%** | **4.1%** |
| | not Michigan | -0.09% | 0.01% | 0.04% | -0.02% | 0.02% | 0.01% |
| **Expenditure** | Michigan | **0.8%** | **1.3%** | **-0.9%** | **0.9%** | **1.5%** | **-1.0%** |
| | not Michigan | 0.010% | 0.017% | -0.009% | 0.001% | 0.004% | -0.002% |
| **Decrease (20%)** | | | | | | | |
| **Tuition** | Michigan | **7.0%** | **-0.2%** | **-4.1%** | **7.8%** | **0.0%** | **-4.3%** |
| | not Michigan | 0.10% | -0.01% | -0.04% | 0.03% | -0.02% | -0.01% |
| **Expenditure** | Michigan | **-0.7%** | **-1.2%** | **0.8%** | **-0.8%** | **-1.3%** | **0.9%** |
| | not Michigan | -0.010% | -0.016% | 0.008% | -0.001% | -0.004% | 0.002% |

Some illustrative results are shown in Table 7, taking Michigan as an example; results for some other large States are shown in Tables 8 and 9. The population distributions over locations and educational attainment are compared at age 36. The tables show (roughly symmetric) changes in the proportions of college-educated men, classified alternatively by current location and by home location. The tuition effects are substantially larger than the expenditure effects, although expenditure changes do seem to influence the number of people who have some college education, without much effect on the number of graduates (which is consistent with the coefficient estimates in Table 5). The most striking result is that, contrary to the findings in Bound et al. (2004), changes in the policy variables have substantial long-term effects on the educational composition of the local population many years after the fact, despite some leakage due to migration.

The tuition effects shown in Table 8 are quite large. The estimated effects vary considerably across States, and this is also true for the expenditure effects. To some extent this variation arises because a 20% change in tuition or expenditure levels corresponds to different dollar amounts, depending on the initial level. But there is still considerable cross-State variation even if the changes are rescaled to represent equal dollar amounts in each State.

The contrast between the findings here and the results in Bound et al. (2004) provides a nice illustration of the difference between structural and so-called "reduced form" empirical models. Bound et al. (2004) found a fairly weak relationship between flows of new college graduates and subsequent stocks of graduates in the labor force at the State level, suggesting that there is relatively little scope for State policies that aim to increase the proportion of college graduates in the State labor force by investing more in the State's public colleges. The interpretation of this finding is that increases in the flow of college graduates generated by tuition reductions or expenditure increases in the State do not have much effect on long-term stocks because workers are mobile (and college graduates are

[24]Note that there is no need to simulate actual choices, so there is no simulation error in these calculations (aside from rounding errors arising from repeated multiplication of large probability matrices that have some very small elements associated with very unlikely choices).

Table 8: **Effects of Tuition Reductions**(20%)

| | Population at Age 36 | | | | | |
|---|---|---|---|---|---|---|
| | Current Location | | | Home Location | | |
| | Graduates | Some College | High School | Graduates | Some College | High School |
| California | 2.6% | -0.9% | -2.1% | 2.7% | -1.0% | -2.3% |
| Florida | 3.0% | -0.3% | -2.2% | 3.6% | -0.3% | -2.6% |
| Illinois | 5.1% | -0.4% | -3.4% | 6.0% | -0.3% | -3.7% |
| Massachusetts | 2.4% | -0.4% | -1.8% | 2.6% | -0.3% | -2.0% |
| Michigan | 7.0% | -0.2% | -4.1% | 7.8% | 0.0% | -4.3% |
| New York | 4.1% | -0.3% | -3.3% | 4.5% | -0.2% | -3.6% |
| Pennsylvania | 7.5% | 0.1% | -3.8% | 9.1% | 0.3% | -4.1% |
| Texas | 1.9% | -0.1% | -1.5% | 2.1% | -0.1% | -1.6% |
| Wisconsin | 5.1% | -0.1% | -3.2% | 5.7% | 0.0% | -3.4% |

Table 9: **Effects of Increases in Expenditures**(20%)

| | Population at Age 36 | | | | | |
|---|---|---|---|---|---|---|
| | Current Location | | | Home Location | | |
| | Graduates | Some College | High School | Graduates | Some College | High School |
| California | 1.7% | 5.3% | -3.9% | 1.9% | 5.8% | -4.3% |
| Florida | 2.4% | 6.2% | -4.0% | 3.1% | 7.5% | -5.0% |
| Illinois | 0.6% | 1.2% | -0.8% | 0.7% | 1.3% | -0.9% |
| Massachusetts | 0.1% | 0.3% | -0.2% | 0.2% | 0.4% | -0.2% |
| Michigan | **0.8%** | **1.3%** | **-0.9%** | **0.9%** | **1.5%** | **-1.0%** |
| New York | 0.4% | 1.2% | -0.7% | 0.5% | 1.3% | -0.8% |
| Pennsylvania | 0.3% | 0.3% | -0.2% | 0.3% | 0.3% | -0.2% |
| Texas | 0.3% | 0.1% | -0.3% | 0.3% | 0.1% | -0.3% |
| Wisconsin | 0.3% | 0.1% | -0.2% | 0.3% | 0.1% | -0.2% |

particularly mobile). The problem with this interpretation is that there is no analysis of what caused the flow increase, and there are good reasons to expect that the proportion of the flow increase that "sticks" in the State is not invariant with respect to alternative causes of the increase. One example is that if the increased flow of new graduates was generated by attracting a lot of students from other States, then it is likely that many of these students would return home or move elsewhere after graduation, whereas an increase in the number of students from this State would be associated with a strong tendency to remain in the State after graduation. In contrast, the structural model considers specific policy changes, keeping track of the effects of these changes on the choices made by individuals who differ in various respects, and in particular allowing for migration decisions that are strongly affected by individuals' home locations. This gives much sharper conclusions, especially with respect to the leakage of college graduates due to migration. Indeed, the structural results indicate that the leakage due to migration is negligible.

# 7 Conclusion

The data indicate that there are strong economic incentives to migrate from low-wage to high-wage locations. Using a dynamic programming model of expected income maximization to quantify these incentives, it is found that they do in fact generate sizable supply responses in NLSY data. There are also big differences across States in the extent to which higher education is subsidized, and these State subsidies are apparently motivated to a large extent by a perceived interest in having a highly educated labor force. Given the finding that workers respond to migration incentives, it might be expected that State subsidies would have the intended effect, in the sense that States that provide more generous subsidies induce more people to go to college. It is then reasonable to conclude that even if some of these people subsequently move elsewhere, the costs of migration are such that most people will choose to stay, so that subsidies increase the level of human capital in the local labor force. The evidence presented here suggests that more generous subsidies actually do have significant effects on college enrollments, particularly when the subsidies take the form of tuition reductions. The strongest effects are found for community colleges, which are financed to a large extent by subsidies at the local level in many States. The model is used to quantify the effects of changes in subsidies and tuition in individual States. The main result is that the effects on the educational level of the local labor force are apparently long-lasting: the extent to which they are dissipated through migration is very slight.

## References


Agan, A. Y. (2014). Disaggregating the returns to college. Unpublished, Princeton University.

Aghion, P., Boustan, L., Hoxby, C., and Vandenbussche, J. (2009). The causal impact of education on economic growth: Evidence from the United States. Unpublished, Harvard University.

Bishop, K. C. (2008). A dynamic model of location choice and hedonic valuation. Unpublished, Washington University in St. Louis.

Börsch-Supan, A. (1990). On the compatibility of nested logit models with utility maximization. *Journal of Econometrics*, 43(3):373–388.

Bound, J., Groen, J., Kezdi, G., and Turner, S. (2004). Trade in university training: cross-state variation in the production and stock of college-educated labor. *Journal of Econometrics*, 121(1-2):143 – 173.

Card, D. and Krueger, A. B. (1992). Does school quality matter? returns to education and the characteristics of public schools in the united states. *The Journal of Political Economy*, 100(1):1–40.

Card, D. and Lemieux, T. (2001). Dropout and enrollment trends in the postwar period: What went wrong in the 1970s? In Gruber, J., editor, *An Economic Analysis of Risky Behavior Among Youth*, pages 439–482. University of Chicago Press, Chicago.

Dynarski, S. (2000). Hope for whom? Financial aid for the middle class and its impact on college attendance. *National Tax Journal*, 53(3 Part 2):629–662.

Dynarski, S. M. (2003). Does aid matter? Measuring the effect of student aid on college attendance and completion. *The American Economic Review*, 93(1):279–288.

Gemici, A. (2011). Family migration and labor market outcomes. Unpublished, New York University.

Goldin, C. and Katz, L. F. (1999). The shaping of higher education: The formative years in the united states, 1890 to 1940. *Journal of Economic Perspectives*, 13(1):37–62.

Groen, J. A. (2004). The effect of college location on migration of college-educated labor. *Journal of Econometrics*, 121(1-2):125 – 142. Higher education (Annals issue).

Kane, T. J. (2006). Public intervention in post-secondary education. In Hanushek, E. and Welch, F., editors, *Handbook of the Economics of Education*, volume 2, pages 1369 – 1401. Elsevier Science, North-Holland, Amsterdam, New York and Oxford.

Kane, T. J. (2007). Evaluating the impact of the D.C. tuition assistance grant program. *Journal of Human Resources*, XLII(3):555–582.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, 105(3):473–522.

Keane, M. P. and Wolpin, K. I. (2001). The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, 42(4):1051–1103.

Kennan, J. (2006). A note on discrete approximations of continuous distributions. Unpublished, University of Wisconsin–Madison.

Kennan, J. and Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, 79(1):211–251.

McFadden, D. D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*. Academic Press.

McFadden, D. L. (1978). Modelling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. S. and Weibull, J., editors, *Spatial Interaction Theory and Planning Models*, chapter 3, pages 75–96. North Holland: Amsterdam.

Rust, J. P. (1994). Structural estimation of markov decision processes. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, pages 3081–3143. Elsevier Science, North-Holland, Amsterdam, New York and Oxford.