Econometrica, Vol. 70, No. 6 (November, 2002), 2367–2420

MOBILITY AND THE RETURN TO EDUCATION: TESTING A ROY MODEL WITH MULTIPLE MARKETS

BY GORDON B. DAHL[1]

Self-selected migration presents one potential explanation for why observed returns to a college education in local labor markets vary widely even though U.S. workers are highly mobile. To assess the impact of self-selection on estimated returns, this paper first develops a Roy model of mobility and earnings where workers choose in which of the 50 states (plus the District of Columbia) to live and work. Available estimation methods are either infeasible for a selection model with so many alternatives or place potentially severe restrictions on earnings and the selection process. This paper develops an alternative econometric methodology that combines Lee's (1983) parametric maximum order statistic approach to reduce the dimensionality of the error terms with more recent work on semiparametric estimation of selection models (e.g., Ahn and Powell (1993)). The resulting semiparametric correction is easy to implement and can be adapted to a variety of other polychotomous choice problems. The empirical work, which uses 1990 U.S. Census data, confirms the role of comparative advantage in mobility decisions. The results suggest that self-selection of higher educated individuals to states with higher returns to education generally leads to *upward* biases in OLS estimates of the returns to education in state-specific labor markets. While the estimated returns to a college education are significantly biased, correcting for the bias does not narrow the range of returns across states. Consistent with the finding that the corrected return to a college education differs across the U.S., the relative state-to-state migration flows of college- versus high school-educated individuals respond strongly to differences in the return to education and amenities across states.

KEYWORDS: Selection bias, polychotomous choice, Roy model, return to education, migration.

1. INTRODUCTION

ESTIMATING SIMPLE HUMAN CAPITAL regressions for workers currently living in each state in the U.S. reveals a wide variation in measured returns to a college education, ranging from 22 percent in Wyoming to 52 percent in Texas.[2] Another empirical observation about the U.S. labor market is the high mobility rate of its workers. The U.S. Census reveals that in 1990 four percent of white males age

[1] I thank Orley Ashenfelter, Mark Bils, David Dahl, Bo Honoré, Shakeeb Khan, David Lee, Lance Lochner, James Powell, an editor, and two anonymous referees for valuable comments. In addition, I am particularly grateful to David Card for many helpful discussions and suggestions. I also thank seminar participants at Brandeis University, Brigham Young University, Dartmouth College, George Mason University, Princeton University, Stanford University, SUNY Stonybrook, UC Berkeley, UC Davis, University of Chicago Harris School, University of Rochester, and UT Austin.

[2] Returns to a college education relative to a high school education, controlling for potential experience, marital status, and residence in an SMSA, estimated on a subset of white males who were working full-time and age 25 to 34 in the 1990 U.S. Census. Throughout the paper, the District of Columbia is treated as if it were a state.

25 to 34 moved across state lines, with the cumulative effect that almost a third of this group no longer resided in the state in which they were born. The wide variation in college returns and the high rate of interstate migration leads naturally to the question of why returns to schooling are not equalized across states. Self-selected migration presents one potential explanation for the observed variability. If workers chose where to live and work based on comparative advantage, then the estimated returns to college in any given state could be biased upward or downward.

To understand the effects of self-selected migration and obtain unbiased estimates of the return to college, I develop a multi-market Roy (1951) model of mobility and earnings. Instead of workers choosing occupations as in Roy's paper, this paper formulates a model where individuals choose in which of the 50 states in the U.S. (plus the District of Columbia) to live and work. In the same spirit as Roy's model, different geographical areas are modeled as having different earnings and different amenity benefits for workers with different schooling levels. Therefore, self-selected migration causes the observed return to education for current residents of a state to differ from the return we would expect a randomly chosen individual to earn.

Estimating a Roy model without imposing severe restrictions on the selection process presents many challenges, especially when there are a large number of alternative choices. Easily implementable techniques for dichotomous and polychotomous selection models were first developed in a parametric framework.[3] As parametric techniques have come under closer scrutiny, efforts have been made to relax these distributional assumptions. For dichotomous choice models, various semiparametric methods that avoid specifying the joint distribution of the error terms in the outcome equation (for example, an earnings equation) and the single selection equation have been proposed.[4] However, the challenges inherent in polychotomous choice models have largely prevented parallel advances in Roy models with many alternatives.[5]

This paper proposes a new semiparametric methodology to correct for sample selection bias in polychotomous choice models. I start with Lee's (1983) insight

[3] For dichotomous choice models, the earliest method to receive widespread use was Heckman's (1979) two-step procedure, which assumes joint normality of the error terms in the outcome equation and the selection equation. Under this assumption, the addition of a simple expression representing the conditional mean of the selected residuals to the outcome equation of interest will control for selectivity bias. Later work fruitfully developed parametric approaches to deal with polychotomous choices. Hay (1980) and Dubin and McFadden (1984) generalized Heckman's method to a multinomial context, while Lee (1982, 1983) transformed univariate order statistics to construct a simpler two-stage estimator.

[4] One of the earliest semiparametric approaches can be found in Heckman (1981). Most recent approaches take advantage of a latent index framework to characterize the conditional mean of the error term in the outcome equation (see, for example, Manski (1985), Newey, Powell, and Walker (1990), Cosslett (1991), Klein and Spady (1993), Ahn and Powell (1993)). A survey of methods to deal with sample selection bias can be found in Vella (1998).

[5] Notable exceptions are Ichimura and Lee (1991) and Lee (1995), who extend the semiparametric estimation of single-index models to a multiple-index context.

that the maximum order statistic can summarize the error terms of a multi-choice selection model with a single random variable. Lee's approach assumes that with multiple alternatives, the choice that matters is the first-best, or observed choice. I then combine Lee's idea with newer research on semiparametric estimation of single-index models (e.g., Ahn and Powell (1993)), with the result that the sample selection correction takes the form of an unknown function of the first-best selection probability.[6] I extend this approach to a multiple-index framework, where the bias correction is an unknown function of a small number of selection probabilities. As a supplement to this new methodology, I classify similar individuals into cells to get a simple distribution-free estimate of the selection probabilities. By using cell means for the selection probabilities, I avoid imposing the undesirable "independence of irrelevant alternatives" property inherent in the conditional logit model. The resulting two-step semiparametric correction avoids the need to specify the joint distribution of the error terms in the outcome and selection equations and is easy to implement.

In the empirical section of the paper, I estimate a Roy model of mobility and earnings and test for the presence of self-selection. The analysis, which uses 1990 U.S. Census data, proceeds in three steps. First, I estimate the selection probabilities semiparametrically by grouping individuals with the same discrete characteristics together and taking cell means for the different migration paths. I find considerable variation in state-to-state migration flows for individuals with different levels of education. These flows indicate that comparative advantage in earnings and differences in tastes by education level potentially play an important role in mobility decisions. In the second step, I use these migration probabilities in the correction functions for each state to get consistent estimates of the return to education. The correction functions from the fifty-one wage equations generally enter significantly, confirming the presence of self-selection. The results suggest that self-selection of higher educated individuals to states with higher returns to education generally leads to *upward* biases in the return to a college education, in many cases by 10 to 20 percent. However, the variation between states in returns does not narrow, suggesting that state-specific amenities or other nonwage variables play important roles in the migration decisions of individuals with different levels of education. In the final step, I test the responsiveness of migration flows to differences in the return to education and amenities across states. I find the relative mobility rate of college-educated to high school-educated men from state to state is strongly correlated with amenity differences and the relative gaps in returns to college across the different local labor markets.

2. AN EXTENDED ROY MODEL OF MOBILITY AND EARNINGS

Roy's 1951 paper, "Some Thoughts on the Distribution of Earnings," discusses the effects of self-selection into different occupations in a surprisingly modern

[6] It should be noted that Heckman (1981) and Heckman and Robb (1985, 1986) first proposed the general approach developed in Ahn and Powell (1993).

way. In the paper, which does not include a single equation, he outlines a simple model of selection based on comparative advantage and investigates the resulting effects on the distribution of earnings in different occupations. Roy's general framework has been applied to a variety of labor market settings, including female labor force participation (Gronau (1974), Heckman (1974)), union versus nonunion employment (Lee (1978)), choice of schooling (Willis and Rosen (1979)), internal and international migration (Nakosteen and Zimmer (1980), Borjas (1987), Borjas, Bronars, and Trejo (1992)), training program participation (Ashenfelter and Card (1985), Ham and LaLonde (1996)), occupational choice (Dolton, Makepeace, and van der Klaauw (1989)), and choice of industry (Heckman and Sedlacek (1990)). In each application, the researchers replace the choice of "occupation" in Roy's original paper with a parallel choice of which market or sector to enter.

In this section I develop a Roy model for the choice of where to live and work. Earnings in different areas vary by schooling level and each individual follows the migration path that maximizes utility. As in Roy's simple model, the pursuit of comparative advantage potentially causes the observed return to education in an area to differ from its true population mean. I introduce three extensions to the simple Roy model: (i) multiple markets or sectors, (ii) choices based on utility maximization, and (iii) an unspecified distribution of latent skills.[7]

2.1. *A Model of Mobility and Earnings*

To formalize ideas, consider a country with N distinct geographic areas and think of individuals as living for two periods. In the first period, individuals are born and do not work, while in the second period individuals work. Individuals are randomly assigned to a geographic area at birth in the first period, but choose where they would like to live and work for the second period of their lives. I refer to the state in which an individual is born as the "birth" state, and the state in which an individual chooses to work as the "residence" state. While each area would have the same distribution of individual skills in the absence of migration, self-selected migration potentially alters the skill distribution across states.

This paper focuses on the returns to a particular measure of skill—education, and on a specific set of areas—the 50 states of the U.S. plus the District of Columbia. However, the following extended Roy model and the estimation procedure proposed in this paper could be adapted to a variety of settings with different measures of skill or different definitions of sectors. Consider individuals who have already made their migration decisions and begun working. The population earnings function for individual i if he works in state k is given by

$$y_{ik} = \alpha_k + x_i'\delta_k + s_i\beta_k + u_{ik} \qquad (k = 1, \ldots, N), \tag{1}$$

[7] Roy's model (i) considers only two "occupations" or sectors, (ii) is based on income maximization, and (iii) assumes lognormality of latent skills. Heckman and Honoré (1990) discuss the empirical content of Roy's original model as well as extensions to it.

where y_{ik} is log earnings, α_k is a state-specific constant, x_i is a vector of individual characteristics, s_i measures level of schooling, and u_{ik} is an error term. Of course, an individual's earnings are not observed for all states, but only for the single state in which he chooses to live and work. In the self-selected sample for state k, the error term u_{ik} does not necessarily have zero mean conditional on x_i and s_i, and ordinary least squares regression potentially yields biased estimates of δ_k and β_k. The schooling coefficients, or "returns to education," for the 51 states corrected for migration-induced selection bias comprise the focus of this paper.

In the absence of mobility, the earnings functions described in (1) can differ in each of the N states since the productivity of different skills may vary from state to state. For example, the return to education in different states may vary due to differences in natural resources or varying skill needs of local employers. However, earnings in a state do not depend on a resident's state of birth, a restriction that will play a key role in identification later in the paper. As an example of this specification, two individuals with identical characteristics, living and working in the same state, but born in different states, will earn the same amount. If skills and other individual characteristics are measured perfectly, this formulation is a natural way to describe labor markets: earnings depend on skills, and not on nonproductive characteristics such as state of birth.[8]

Movement between states is based on utility maximization, where utility is a function of earnings and tastes. For expositional purposes, I assume the utility of individual i, born in state j, considering a move to state k consists of an additively separable function of earnings and a person-specific taste factor:

$$V_{ijk} = y_{ik} + t_{ijk} \qquad (k = 1, \ldots, N), \tag{2}$$

where V_{ijk} indexes utility, y_{ik} is log earnings, and t_{ijk} is a vector indexing taste for moving from state j to state k. This taste vector represents the nonwage determinants that enter the utility functions. As such, it includes any fixed costs of moving, amenity differences between states, and any other nonwage or psychic costs and benefits associated with moving from one state to another. For simplicity, I assume that individuals possess accurate expectations about individual-specific earnings and tastes.[9]

[8] The exclusion restriction that an individual's birth state does not influence earnings may not be entirely convincing as an instrument. If skills are not adequately described by the variables in the wage equation, an individual's state of birth may contain additional information about the earnings process. For example, differences in school quality from state to state may affect the measured return to education (see Card and Krueger (1992)). Equation (1) could be relaxed to allow state of birth to affect earnings; identification would then be achieved through demographic variables not appearing in the wage equations or by restricting the return to education to not depend on the interaction between state of birth and state of residence. In the latter case, cross-equation restrictions would then be necessary for estimation, an approach not pursued in this paper.

[9] The additive separability of the earnings and tastes residuals is not required for the estimation procedure developed in this paper. Likewise, adding in uncertainty so that migration is based on expected utility maximization does not change the main insights of the model or the applicability of the estimation method that follows. Rather, the role of these assumptions is to simplify the discussion of earnings and tastes throughout the paper.

The deviation of an individual's earnings if they were to work in state k from the average for the entire population (including individuals who do not actually work in state k) is

$$y_{ik} - E[y_{ik}|x_i, s_i] = u_{ik} \qquad (k = 1, \ldots, N). \tag{3}$$

Define a similar equation for the deviation of an individual's taste for moving from state j to state k from the population average, so that

$$t_{ijk} - E[t_{ijk}|z_i] = w_{ijk} \qquad (k = 1, \ldots, N), \tag{4}$$

where z_i is a vector of individual characteristics and w_{ijk} is an error term for individual deviations from mean tastes. Notice that I allow the value for mean tastes to be a function of both state of birth j and state of residence k, whereas I restrict mean earnings in (3) to be a function only of state of residence. Tastes for moving from state j to state k potentially include an overwhelming number of variables. For example, t_{ijk} could include the costs of moving from j to k, the difference in climate between j and k, the difference in state tax rates between j and k, or any other nonwage differences between the two states.

The expression for V_{ijk} can now be written in terms of the population mean and an error component specific to the individual:

$$V_{ijk} = V_{jk} + e_{ijk} \qquad (k = 1, \ldots, N), \tag{5}$$

where $V_{jk} = E[y_{ik}|x_i, s_i] + E[t_{ijk}|z_i]$ and $e_{ijk} = u_{ik} + w_{ijk}$. In the selection literature V_{jk} is often called the subutility function.

Individuals follow the migration path that maximizes their utility, so that individual i chooses to move from state j to state k according to

$$\begin{aligned} M_{ijk} &= 1 \quad \text{if and only if} \quad V_{ijk} = \max(V_{ij1}, \ldots, V_{ijN}), \\ &= 0 \quad \text{otherwise}, \end{aligned}$$

where M_{ijk} is an indicator for whether individual i actually moves from state j to state k. The selection equations can alternatively be written as

$$\begin{aligned} M_{ijk} &= 1 \quad \text{if and only if} \quad V_{jk} + e_{ijk} \geq V_{jm} + e_{ijm} \quad \forall m, \\ &= 0 \quad \text{otherwise}. \end{aligned} \tag{6}$$

Utility depends on the specific migration path j to k; that is, the utility for an individual depends not only on the state of residence, but also on the state of birth. Assume the set $\{V_{ij1}, \ldots, V_{ijN}\}$ has a unique maximum and the error terms from the N selection criteria in (6) have a joint distribution that has finite moments and depends on a finite dimensional parameter set.

In this model, an individual can only live and work in one state; therefore, earnings for an individual are not observed in every state. The selection rule is

$$y_{ik} \text{ observed} \quad \text{if and only if} \quad M_{ijk} = 1, \tag{7}$$

so that earnings are observed for an individual's utility maximizing choice. Earnings are observed only if all N selection equations in (6) are satisfied simultaneously. Equations (1)–(7) describe an extended Roy model of earnings and mobility. Note that individuals currently living in state k are not a random sample of the population, and in general

$$
\begin{aligned}
(8) \qquad E[u_{ik}|y_{ik} \text{ observed}] &= E[u_{ik}|M_{ijk} = 1] \\
&= E[u_{ik}|e_{ijm} - e_{ijk} \leq V_{jk} - V_{jm}, \forall m] \\
&\neq 0.
\end{aligned}
$$

I refer to $E[u_{ik}|M_{ijk} = 1]$ as the selectivity bias for observation i. If this conditional expectation is correlated with x_i or s_i, OLS regression of observed y_{ik} on x_i and s_i will result in biased estimates. The direction and size of the bias for an individual depends on the joint distribution of u_{ik} and the error terms from the N migration equations, $e_{ij1} - e_{ijk}, \ldots, e_{ijN} - e_{ijk}$. Since $e_{ijk} = u_{ik} + w_{ijk}$, the bias depends on the correlation of residual earnings across areas as well as on the relationship between residual earnings and residual tastes.

Unlike the case where there are only two markets and no taste variables, no general statements can be made a priori about when the expected selectivity bias is always positive or always negative for different states and skill levels. Since all N selection equations must be satisfied simultaneously for an individual to move to state k, the selectivity bias will in general vary across individuals observed to be born in the same area j and living in the same area k but with different values for $e_{ij1}, \ldots, e_{ijN}$.

2.2. *Challenges to Estimating a Roy Model with Many Sectors*

The estimation of a Roy model with high dimensionality presents many challenges. With a two sector model, the usual approach specifies the joint distribution of the error terms in the outcome equation and the single selection equation to be bivariate normal. While a few researchers have modified and applied this approach for three and even four choices, estimation quickly becomes intractable as the number of choices increases. Therefore, for polychotomous choice models, most previous work makes distributional assumptions that greatly simplify the form of the selectivity bias.

The most popular procedure for estimating the selection equations in a multichoice Roy model has been the conditional logit model or its extension, the nested logit model (McFadden (1974, 1984), Trost and Lee (1984), Falaris (1987)). While this formulation results in convenient expressions for the selection probabilities, it has the undesirable property of "independence of irrelevant alternatives." For two alternative choices perceived by individuals to be similar rather than independent, this model generates a joint probability of selection for the two alternatives that is too high. While the conditional logit model could be relaxed by specifying a nested structure, the researcher must decide which choices

belong in the same nest and which choices are independent, with the restriction that the correlation between choices in the same nest be positive.

Selectivity bias corrections can also be quite sensitive to departures from the true joint distribution of the error terms (Arabmazar and Schmidt (1981, 1982), Goldberger (1983), Mroz (1987)). Estimators relying on joint normality, for example, can perform poorly when the true selectivity effect is nonlinear and nonmonotonic. The bias arising from incorrectly assuming joint normality becomes particularly severe as the amount of truncation in the self-selected sample increases. Such drawbacks have spawned new semiparametric estimation methods for dichotomous choice models, but as mentioned in the introduction, few methods for polychotomous choice models.

A separate challenge inherent in a Roy model based on utility maximization is which variables to include in the subutility function, as well as the functional form of the subutility function, V_{jk}. Estimation schemes based on utility maximization and not just income maximization may be more realistic, but usually require the researcher to model tastes. In this paper, a multitude of variables potentially belong in the taste component of the utility function. A few examples include moving costs, differences in nonwage local labor market characteristics, the relative cost of living, and differences in public services, taxation, climate, and crime rates (Roback (1982, 1988)). To complicate matters, many of the variables that belong in V_{jk} may be unobservable or poorly measured.

In this paper, I attempt to overcome some of the challenges inherent in an extended Roy model. I reduce the dimensionality of the selection criteria without stringent distributional assumptions on the error terms in the outcome and selection equations. This paper also avoids the problems associated with modeling tastes by sidestepping estimation of the underlying parameters of the subutility function.

3. MODELING SELECTION BIAS WITH MULTIPLE CHOICES

In this section, I present a new estimation method and discuss the implications of my approach. While I develop the methodology in the context of a Roy model of mobility, it could be applied to a variety of polychotomous choice settings. My insight is that a reinterpretation of Lee's (1983) maximum order statistic approach combined with more recent work on semiparametric estimation of sample selection models provides a simple estimation procedure that allows flexible modeling of the joint distribution of the error terms in polychotomous choice models.

3.1. *Reducing the Dimensionality of the Joint Distribution of the Error Terms*

I begin by modeling the joint distribution of the error terms in the earnings equation and the selection equations. Let $f_{jk}(u_{ik}, e_{ij1}-e_{ijk}, \ldots, e_{ijN}-e_{ijk})$ denote the joint density function of the error term in the earnings equation (1) and the error terms in the selection criteria (6), and let F_{jk} denote the corresponding

cumulative distribution. Similarly denote the marginal joint density of the selection error as $f^e_{jk}(e_{ij1}-e_{ijk},\ldots,e_{ijN}-e_{ijk})$ for birth state j and residence state k, and let F^e_{jk} denote the cumulative distribution.

3.1.1. *Using the Lee Approach to Selection Correction*

Accounting for the correlation of the error terms from N selection equations with the error term in the earnings equation of interest appears overwhelming. A parametric generalization of Heckman's two-step approach would require a complete specification of $f_{jk}(u_{ik}, e_{ij1}-e_{ijk},\ldots,e_{ijN}-e_{ijk})$, and would involve the integration of an $(N-1)$-fold integral. Lee (1983) suggests reducing the dimensionality of the problem by reframing the N selection equations in (6) in terms of order statistics. Combining equations (6) and (7), the selection rule for state k becomes

$$y_{ik} \text{ observed} \quad \text{if and only if}$$
$$(V_{j1}-V_{jk}+e_{ij1}-e_{ijk},\ldots,V_{jN}-V_{jk}+e_{ijN}-e_{ijk})' \leq \mathbf{0}$$

where $\mathbf{0}$ is an N-dimensional column vector. To understand Lee's approach, note that an equivalent expression is

$$y_{ik} \text{ observed} \quad \text{if and only if} \quad \max_m (V_{jm}-V_{jk}+e_{ijm}-e_{ijk}) \leq 0 \tag{9}$$

where $\max_m(\bullet)$ indicates the maximum over m. Thus any selectivity bias in y_{ik} is driven by the event that the maximum of the collection of random variables $V_{j1}-V_{jk}+e_{ij1}-e_{ijk},\ldots,V_{jN}-V_{jk}+e_{ijN}-e_{ijk}$ is less than or equal to zero. The distribution function H_{jk} of the maximum order statistic, conditional on the subutility function differences, can be expressed as

$$\begin{aligned} H_{jk}(t|V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}) &= Pr\Big[\max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk}) < t|V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}\Big] \\ &= Pr[e_{ij1}-e_{ijk} < V_{j1}-V_{jk}+t,\ldots,e_{ijN}-e_{ijk} < V_{jN}-V_{jk}+t] \\ &= F^e_{jk}(V_{j1}-V_{jk}+t,\ldots,V_{jN}-V_{jk}+t), \end{aligned} \tag{10}$$

which makes clear that H_{jk} (conditional on $V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}$) evaluated at zero is simply the probability of sample selection. Given the equivalent formulation of the selection rule in (9), the cumulative distribution function F_{jk} can now be expressed in the following ways:

$$\begin{aligned} F_{jk}(r, V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}) &= Pr(u_{ik} < r, e_{ij1}-e_{ijk} < V_{j1}-V_{jk},\ldots,e_{ijk}-e_{ijN} < V_{jN}-V_{jk}) \\ &= Pr\Big[u_{ik} < r, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk}) < 0|V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}\Big] \\ &= G_{jk}(r, 0|V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}) \end{aligned} \tag{11}$$

where G_{jk} is a well-defined cumulative joint distribution function for u_{ik} and $\max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})$. Writing (11) in terms of density functions provides another way to express this distributional equivalence:

$$(12)\qquad f_{jk}(u_{ik}, e_{ij1}-e_{ijk}, \ldots, e_{ijN}-e_{ijk}|V_{j1}-V_{jk}, \ldots, V_{jN}-V_{jk}) = g_{jk}\left(u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})|V_{j1}-V_{jk}, \ldots, V_{jN}-V_{jk}\right)$$

where both sides of the equation are explicitly written as conditional densities to emphasize the dependence on the differences in subutility functions. Equation (12) has reduced the dimensionality of the error terms that must be accounted for by expressing an N-variate joint distribution in terms of a bivariate distribution.

3.1.2. *A Reinterpretation of Lee's Approach*

So far, no assumptions have been made to arrive at the distributional equivalence expressed in equation (12). What restrictions will allow the researcher to take advantage of this reduction in dimensionality? As developed more formally in Appendix A, Lee's method constructs a new random variable from the maximum order statistic by making a transformation (to normality, for example). Lee then assumes that this newly created random variable has a joint distribution (bivariate normal, for example) with the error term in the outcome equation of interest. Building on earlier work (Lee (1982)), Lee points out that distributions other than the normal can be used for the transformation, providing a method for generating a large class of models with selectivity. Different transformations allow for a wide variety of joint distributions for the error terms in the outcome equation and the selection equations, regardless of the specific model used to estimate the choice probabilities.

For example, suppose the researcher believes the marginal distribution of the error term in the outcome equation to be normal. Suppose the researcher also uses the conditional multinomial logit model so that the random parts of the utility functions (in the current paper $e_{ij1}-e_{ijk}, \ldots, e_{ijN}-e_{ijk}$) are assumed to be independent and identically distributed with the extreme value distribution. A transformation of the maximum order statistic to normality allows the researcher to assume a joint bivariate normal distribution for the error term in the outcome equation and the transformed maximum order statistic. Then a simple Heckman-type correction will control for selectivity bias, by adding a term that takes the form of the inverse Mill's ratio to the outcome regression function. The choice of the conditional logit model does not dictate the form of the selectivity bias correction, since the researcher can make a transformation consistent with the assumed joint normality.

Reinterpreting Lee, his approach is not just a parametric transformation of the maximum order statistic and an ensuing distributional assumption for this transformed variable and the error term in the outcome equation. Rather, underlying Lee's parameterizations is an implicit assumption; namely, the joint distribution of the error term in the outcome equation and the maximum order

statistic does not depend on the subutility function differences. As noted in (10), the random variables $\max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})$ indexed over i are not identically distributed, since in general the distribution function depends on the subutility function differences, $V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}$. Lee's transformation of the maximum is central not because it creates a new normally distributed variable, but because the same transformation is applied *regardless* of the specific values for $V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}$. Hence, the unstated assumption implicit in Lee's approach for polychotomous choice models is

$$\text{(A-1)}\qquad g_k\Big(u_{ik},\max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})\Big)$$
$$\text{does not depend on}\quad V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}.$$

While I take advantage of Lee's idea to reduce the dimensionality of the error terms, I propose somewhat less restrictive assumptions than assumption (A-1).

3.2. *Using Migration Probabilities as Sufficient Statistics in Single- and Multiple-Index Models*

Lee's approach could be extended with recent semiparametric advances to estimate the parameters of the selection criteria with fewer distributional assumptions. Some of these methods use nonparametric regression to estimate the unknown distribution function of the selection errors and the accompanying regression parameters (for example, Ichimura (1987), Newey, Powell, and Walker (1990), Cosslett (1991), Klein and Spady (1993)). Such an approach would still require modeling the determinants of utility and a correct specification of how these conditioning variables should enter the selection correction function.

To avoid these problems, I pursue an alternative approach motivated by the observation that in single-index selection models, the selectivity bias can be written as a function of the probability of selection given covariates (Heckman and Robb (1985, 1986), Choi (1992), Ahn and Powell (1993)). This form for the correction term follows from the fact that in latent index models, the selected mean of the error term in the outcome equation is an invertible function of the selection probability. Using this fact, Ahn and Powell sidestep estimation of the unknown distribution function of the selection errors. A similar idea extends to multiple-index models. Combining these insights with Lee's approach results in a simple and flexible semiparametric correction for polychotomous selection models.

3.2.1. *Formulation as a Single-Index Model*

The formulation of mobility and earnings in equations (1) and (6) implies the earnings equations can be rewritten as multiple-index, partially-linear models:

$$\text{(13)}\qquad y_{ik}=\alpha_k+x_i'\delta_k+s_i\beta_k+\sum_{j=1}^{N}[M_{ijk}\times\psi_{jk}(V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk})]+v_{ik}\quad(k=1,\ldots,N)$$

where $\psi_{jk}(\bullet) = E[u_{ik}|V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}]$ and v_{ik} is an error term with mean zero in the conditional sample for state k. As a reminder, M_{ijk} is a dummy variable that equals one if individual i was born in state j and currently resides in state k. Equation (13) is called a multiple index model because the control functions ψ_{jk} for each birth state j are unknown functions of the multiple indices $V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}$. Fortunately, the dimensionality of the control functions can be reduced using a modification of Lee's approach.

To take advantage of Lee's insight in a semiparametric framework, I make the following index sufficiency assumption:

$$\text{(A-2)}\qquad g_{jk}\Big(u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})|V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}\Big) = g_{ik}\Big(u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})|p_{ijk}\Big)$$

where p_{ijk} is the probability that individual i moves from state j to state k given the vector $V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}$. The equivalence in (A-2) assumes that $p_{ijk} = p_{ijk}(V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk})$ exhausts all the information about how $V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}$ influences the joint distribution of u_{ik} and $\max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})$ contained in the sample. That is, the conditional distribution of u_{ik} and $\max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})$ can depend on the conditioning variables only through the single index p_{ijk}.

The single index p_{ijk} is the probability of an individual's first-best migration choice, a choice which is observable since the researcher knows where an individual chooses to live and work. This scalar migration probability associated with the maximum order statistic can be written in the following ways:

$$\begin{aligned}\text{(14)}\qquad p_{ijk} &= Pr(M_{ijk}=1|V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk})\\ &= Pr(V_{jk}+e_{ijk}\geq V_{jm}+e_{ijm},\ \forall m)\\ &= F^e_{jk}(V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk})\\ &= H_{jk}(0|V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}).\end{aligned}$$

The researcher must somehow account for the subutility functions to get an estimate of p_{ijk}, since the vector $V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}$ determines an individual's migration choice. Discussion of how to estimate p_{ijk} is postponed until later; for the moment, assume that an appropriate estimator is available.

Using assumption (A-2), equation (12) can be simplified to

$$\text{(15)}\qquad f_{jk}(u_{ik}, e_{ij1}-e_{ijk},\ldots,e_{ijN}-e_{ijk}|V_{j1}-V_{jk},\ldots,V_{jN}-V_{jk}) = g_{jk}\Big(u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})|p_{ijk}\Big)$$

and the earnings equations can now be written as single-index, partially linear models:

$$\text{(16)}\qquad y_{ik} = \alpha_k + x_i'\delta_k + s_i\beta_k + \sum_{j=1}^{N}\{M_{ijk}\times\lambda_{jk}(p_{ijk})\} + \omega_{ik} \qquad (k=1,\ldots,N)$$

where for each birth state j, $\lambda_{jk}(\bullet)$ is an unknown function of the single index p_{ijk} and ω_{ik} is an error term. I refer to the λ_{jk}'s as the selection correction functions for state k. By construction, the error term ω_{ik} has zero mean given the migration probability and the fact that earnings are observed in a state:

$$E[\omega_{ik}|x_i, s_i, p_{ijk}, M_{ijk} = 1] = 0 \qquad (k = 1, \ldots, N).$$

A proof for the result that $\psi_{jk}(V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}) = \lambda_{jk}(p_{ijk})$ if assumption (A-2) holds is provided in Appendix B.

3.2.2. *Extension to a Multiple-Index Framework*

One interpretation for the use of p_{ijk} in the correction function for polychotomous choice models relies on the fact that, subject to an invertibility condition[10]

$$(17) \qquad g_{jk}\Big(u_{ik}, \max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk})|V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}\Big) = g_{jk}\Big(u_{ik}, \max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk})|p_{ij1}, \ldots, p_{ijN}\Big),$$

which simply states that the multiple migration probabilities, $p_{ij1}, \ldots, p_{ijN}$, contain the same information as the differenced subutility functions, $V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}$. This implies the earnings equations can be rewritten as multiple-index, partially-linear models that depend on all N migration probabilities:

$$(18) \qquad y_{ik} = \alpha_k + x_i'\delta_k + s_i\beta_k + \sum_{j=1}^{N}[M_{ijk} \times \mu_{jk}(p_{ij1}, \ldots, p_{ijN})] + \upsilon_{ik} \qquad (k = 1, \ldots, N)$$

where $\mu_{jk}(\bullet) = E[u_{ik}|p_{ij1}, \ldots, p_{ijN}] = E[u_{ik}|V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}]$. Assumption (A-2) simplifies this equivalence by assuming that only the probability of the utility maximizing choice matters for the parameterization of the joint distribution g_{jk}. Hence, (A-2) can also be thought of as an exclusion restriction in that it requires the distribution of u_{ik} and $\max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk})$ given $p_{ij1}, \ldots, p_{ijN}$ to be the same as that given p_{ijk}.

A relaxation of (A-2) allows other probabilities besides the first-best choice probability to also influence the joint distribution g_{jk}. Letting $\vec{q}$ represent a chosen subset of the full set of migration probabilities $\{p_{ij1}, \ldots, p_{ijN}\}$, this less restrictive assumption can be written as

$$(A\text{-}3) \qquad g_{jk}\Big(u_{ik}, \max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk})|V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}\Big) = g_{jk}\Big(u_{ik}, \max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk})|p_{ijk}, \vec{q}\Big).$$

[10] To insure that equation (17) holds locally, the assumptions of the implicit function theorem must be satisfied. The $N \times N$ determinant of the vector of implicit equations $[F^e_{jm}(V_{j1} - V_{jm}, \ldots, V_{jN} - V_{jm}) - p_{ijm}] = 0, m = 1, \ldots, N$, must be nonzero so that the Jacobian is nonzero and a local inverse function exists.

This extension allows the earnings equations to be written as multiple-index, partially linear models, where the bias correction is an unknown function of the first-best migration probability plus a few other chosen probabilities. The proof in Appendix B can easily be adapted to account for these additional probabilities. The difficult task is deciding which other probabilities are important in parameterizing the joint distribution g_{jk}, since only a select number of probabilities can be included before the curse of dimensionality makes estimation infeasible.

It would be natural to include the second, third, or perhaps even fourth best choice probabilities as additional terms in the correction function. Unfortunately, which probabilities correspond to an individual's second through Nth best choices cannot usually be determined since the researcher generally only observes the individual's first best, or actual choice. In the current application, a few probabilities suggest themselves as likely candidates for inclusion in the bias correction functions for state k. One possibility is the "retention" probability; that is, the probability that a person born in state j will choose to remain in state j. Another possibility is to include the highest predicted probability for an individual, excluding the retention probability, namely, $\max_m(p_{ijm})$ $m \neq j$. A final possibility is to include the migration probabilities of states geographically near an individual's birth state.

In the current application, I end up adding the retention probability as another term in the correction functions. I discuss how this term was chosen for inclusion in the next section of the paper. Thus, the maintained distributional assumption for the current application is

$$\text{(A-4)} \qquad g_{jk}\Big(u_{ik}, \max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) | V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}\Big) = g_{jk}\Big(u_{ik}, \max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) | p_{ijk}, p_{ijj}\Big),$$

which implies the earnings equations can be written as

$$\text{(19)} \qquad y_{ik} = \alpha_k + x_i'\delta_k + s_i\beta_k + \sum_{j=1}^{N}\{M_{ijk} \times \lambda_{jk}^*(p_{ijk}, p_{ijj})\} + \omega_{ik}^* \qquad (k = 1, \ldots, N).$$

The correction terms in the wage equations for movers are now unknown functions of two probabilities, p_{ijj} and p_{ijk}. Notice that for stayers the correction terms are a function of a single probability, p_{ikk}, since $j = k$ for individuals who do not move from their birth state.

3.3. *The Index Sufficiency Assumption*

This section explores the restrictions on earnings and tastes imposed by the index sufficiency assumptions. For ease of presentation, I begin by discussing (A-2), which implies the first-best choice probability is sufficient to describe joint distribution of the error terms in (12). In the context of my model, this restriction implies that only the probability of the utility maximizing choice of residence

matters for selectivity bias. The particular identities of the second best through the Nth best migration choices, along with the probabilities that an individual would have chosen those migration paths, convey no information about earnings in the state in which an individual chooses to live. For example, consider two individuals born in the same state who both chose to move to the same state k because it was their expected utility maximizing choice. The fact that one individual's second choice for where to live differs from the other individual's second choice is irrelevant, and cannot affect the error term for earnings in state k, u_{ik}.

Since the selection correction functions for a state depend on the joint distribution of u_{ik} and $e_{ij1}-e_{ijk},\ldots,e_{ijN}-e_{ijk}$, it is informative to consider the bivariate covariances between the error term in the earnings equation and the error term in each of the migration equations:

$$\text{(20)}\qquad \operatorname{cov}(u_{ik}, e_{ij1}-e_{ijk}),\ldots,\operatorname{cov}(u_{ik}, e_{ijN}-e_{ijk}).$$

Each bivariate covariance in (20) can be broken up into four separate covariances. For example, the first term can be expressed as

$$\operatorname{cov}(u_{ik}, e_{ij1}-e_{ijk}) = \operatorname{cov}(u_{ik}, u_{i1}) - \operatorname{var}(u_{ik}) + \operatorname{cov}(u_{ik}, w_{ij1}) - \operatorname{cov}(u_{ik}, w_{ijk}).$$

A fully flexible estimation scheme should allow for a variety of bivariate covariances, and hence permit a rich combination of zero, positive, and negative selection. An example of a very inflexible approach is estimation that disregards selectivity bias altogether, since such an approach implicitly restricts all of the bivariate covariances to be zero. Equation (20) points out that self-selection can potentially be a problem even if unobserved earnings and tastes are not correlated with each other or across states, since in this case each bivariate covariance equals $-\operatorname{var}(u_{ik})$. More generally, if taste variables play no role in biasing the earnings equation and unobserved earnings are equi-correlated across states, the bivariate covariances are all identical to each other. In this case, the uncorrected estimate of β_k is biased up or down depending on whether s_i is positively or negatively correlated with u_{ik} in the conditional sample for state k.[11]

The assumption of index sufficiency places restrictions on the possible bivariate covariances described in equation (20). For example, under the stronger assumption (A-1), if $E[u_{ik}|\max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})]$ is monotonic in $\max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})$ (as it would be if u_{ik} and $\max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})$ had a bivariate normal distribution), then the bivariate covariances must all have the same sign.[12] Under the weaker index sufficiency assumption (A-2), this statement can be relaxed slightly, although the only route for the bivariate covariances

[11] Correlation between x_i and u_{ik} in the conditional sample for state k could also potentially bias the estimate of β_k.

[12] In a discussion of the limitations of Lee's approach, Schmertmann (1994) provides a proof that can easily be adapted to the current setting.

to differ in sign would be as a function of the first-best choice probability p_{ijk}. It should be noted in passing that the index sufficiency assumption does allow for a restrictive form of conditional heteroskedasticity; namely, the variance of the outcome error can depend on the single index (i.e., the first-best migration probability).

What kind of model would satisfy the index sufficiency assumption of (A-2)? For simplicity, suppose unobserved tastes in the selection equations, w_{ijk}, have zero correlation across birth states and residence states and are uncorrelated with u_{ik}. Then consider the following simple specification for the error term in the earnings equation:

$$u_{ik} = a_i + b_{ik} \qquad (k = 1, \ldots, N) \tag{21}$$

where a_i is an individual fixed effect that has mean zero in the population, but is the same for a given individual in all states, and b_{ik} is a state-specific homoskedastic error term for individuals with population mean zero and uncorrelated across states. Let a_i and b_{ik} be uncorrelated, and denote the population variances of a_i and b_{ik} as σ_a^2 and σ_{bk}^2, respectively. Equation (21) is a natural starting point to describe the unobserved component of earnings. For example, a_i could represent an individual's unobserved ability which is identical across states, so that individuals who earn more (or less) than average in a given state also earn more (or less) than average in any other state. The term b_{ik} could represent the component of an individual's unobserved earnings that varies across states as a result of how good a match state k is for a worker. This simple fixed effects model satisfies assumption (A-2).

To understand why this example satisfies the index sufficiency assumption, notice that the bivariate covariances described in equation (20) are identical:

$$\begin{aligned}
\operatorname{cov}(u_{ik}, e_{ijm} - e_{ijk}) &= \operatorname{cov}(u_{ik}, u_{im}) - \operatorname{var}(u_{ik}) \\
&\quad + \operatorname{cov}(u_{ik}, w_{ijm}) - \operatorname{cov}(u_{ik}, w_{ijk}) \qquad (m = 1, \ldots, N) \\
&= \sigma_a^2 - (\sigma_a^2 + \sigma_{bk}^2) + 0 - 0 \qquad (m = 1, \ldots, N) \\
&= -\sigma_{bk}^2 \qquad (m = 1, \ldots, N).
\end{aligned}$$

It follows that the covariance between u_{ik} and $\max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk})$ also equals $-\sigma_{bk}^2$, which does not depend on the individual's second through Nth best choices for the state in which he would like to live. More generally, the bivariate distribution functions for u_{ik} and $e_{ijm} - e_{ijk}$ are identical for all m, and so the joint distribution of u_{ik} and $\max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk})$ does not depend on the subutility function differences, $V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}$. Hence, the index sufficiency assumption is satisfied. Notice that while the joint density of u_{ik} and $\max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk})$ does not depend on $V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}$, the univariate density of $\max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk})$ does depend on $V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}$. A slightly more general example with the same result is the case where u_{ik} and u_{im} are equi-correlated for $m = 1, \ldots, N$, with potentially different variances for u_{ik} in each state k.

The assumption of index sufficiency is less appealing when the bivariate covariances between the error terms in the outcome and selection equations described in equation (20) are not equal. Modifying equation (21), consider the addition of a loading factor in front of the individual fixed effect component of unobserved earnings:

$$u_{ik} = \tau_k a_i + b_{ik} \qquad (k = 1, \ldots, N) \tag{22}$$

where τ_k is the loading factor for state k. If the loading factor equals one for all states, unobserved ability is rewarded equally in all states, and equation (21) results. However, if the loading factor is correlated with the return to education in a state, the joint distribution of u_{ik} and $\max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk})$ in general depends on $V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}$. In the most extreme case, if τ_k equals β_k for each state k, the return to unobserved ability in state-specific labor markets equals the return to observed education. Under the conditional independence assumption, the only route for $V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}$ to affect the joint density g_{jk} is through the first-best choice probability p_{ijk}, which may be too restrictive in many cases.

Unfortunately, testing whether index sufficiency holds is difficult in models with high dimensionality. Under the null hypothesis, only the probability of the first-best choice belongs in the outcome equation. The alternative hypothesis, namely that index sufficiency does not hold, implies that all N migration probabilities belong in the correction function. But the alternative hypothesis generally cannot feasibly be estimated; indeed the rationale for assuming index sufficiency is to circumvent the curse of dimensionality present in equation (13).[13] A partial test that is implementable is to include a few other probabilities in the correction functions, and test whether these terms significantly change the estimated coefficients of interest or significantly improve the fit of the outcome regressions. Hausman tests similar to those described in footnote 25 are useful to see if the estimated schooling coefficients change as additional probabilities are included. Wald tests analogous to those described in footnote 26 can be used to determine significant differences due to the correction functions as additional probabilities are added. If these additional probabilities result in significant differences, the single index assumption should be extended to account for these other probabilities as described in (A-3), and additional tests should be performed. The challenge, of course, is choosing a probability or combination of probabilities that provides a test with sufficient power. After calculating Hausman and Wald tests as described above for various additional sets of probabilities, I end up using the retention probability in addition to the first-best migration probability in the application that follows.

[13] The current application illustrates the curse of dimensionality well. In the empirical application that follows, polynomial expansions will be used to approximate the unknown correction functions. If the series is a second order expansion and all N migration probabilities, $p_{ij1}, \ldots, p_{ijN}$, are included, the number of regressors for the expansion is $\sum_{i=1}^{N}(i+1)$. In the current model with 51 states, 1,377 terms would appear in each correction function.

To assess the index sufficiency assumption and the general approach developed in this paper, I perform a brief Monte Carlo investigation (see Appendix C). The simulations suggest that the estimation method developed in this paper effectively controls for selection bias in a variety of settings. In the baseline simulations, earnings are modeled as in equation (21) where the index sufficiency assumption holds. For these simulations, the estimation method of this paper performs well for models with a single choice as well as for models with many choices. The approach also purges the coefficient estimates of bias under a variety of distributional assumptions for the error terms in the earnings and taste equations. For example, if the errors in the wage equation are skewed or heteroskedastic, the estimated coefficient of interest remains close to the true value on average. When index sufficiency is violated, as it would be with the factor loading specification of equation (22), the estimator only partially corrects for selection bias. The estimation technique works well in moderately large samples, but performs poorly in small samples. Lee's parametric approach as outlined in Appendix A generally produces similar results. One exception is when the error terms in the wage equation are lognormally distributed rather than normally distributed. In this case, Lee's approach yields biased estimates even though the semiparametric approach performs well.

4. IMPLEMENTATION CHOICES

The previous section outlined a simple approach to the difficult problem of modeling selection bias when there are many choices. The main contribution is that an index sufficiency assumption can greatly reduce the dimensionality of the selection correction functions. For example, under assumption (A-2) the selection correction for a state reduces from N, *N-dimensional* control functions to N *univariate* control functions that depend only on the probability of the first-best choice (see equation 16). This methodology could be applied to a variety of Roy models with multiple alternatives.

In this section, I briefly discuss some of the practical estimation choices I make to facilitate estimation of the earnings equations in the empirical work that follows. I first make an assumption to reduce the number of correction functions that enter a state's earnings equation. I then discuss grouping individuals into discrete categories to allow nonparametric estimation of the selection probabilities. Finally, I discuss how to estimate the unknown correction functions using polynomial expansions.

4.1. *Reducing the Number of Correction Functions*

In the empirical application that follows, there are potentially 51 different correction functions for each residence state k, as there is a different correction function for each possible birth state j. Assumption (A4) reduces the dimensionality of each of these 51 corrections from being a function of all 51 probabilities to 51 corrections that are a function of the first-best migration probability, p_{ijk},

and the retention probability, p_{ijj}. One of these correction functions, λ^*_{kk}, is for stayers. It corrects for the selection bias of individuals born in state k who choose to remain in state k. The other 50 functions, λ^*_{jk} $(j = 1, \ldots, 51; j \neq k)$, rid the earnings equation of selection bias for inmigrants to state k from the other 50 states. Even though there are 51 control functions, it should be noted that the "curse of dimensionality" is eliminated by the assumption of index sufficiency. This is because the rate of convergence for the nonparametric control functions is not affected by the number of correction functions included in the regression (Andrews and Whang (1990), Newey (1994b)).

For a Roy model with five or perhaps as many as ten sectors, a separate correction function could be included for each sending sector j. However, in the current application, estimating 51 different functions for each regression equation is impractical. In general, the correction functions in equation (19) depend on a different joint distribution g_{jk} for each origin state j. To reduce the number of correction functions that enter a state's earnings equation, I assume

$$g_{jk} = g_k \quad \forall\, j \neq k. \tag{23}$$

Equation (23) restricts these distributions for "movers" in the following sense: for a given receiving state k, the joint distribution has to be the same for all possible sending states j (where $j \neq k$). In other words, given that two people choose to move to the same state, their selection biases can be characterized by the same distribution, regardless of their states of origin. This restriction is not imposed on "stayers"; that is, individuals who choose to remain in their state of birth (i.e., $j = k$) are allowed to have a different joint distribution compared to inmigrants. Equation (23) implies that λ^*_{jk} equals λ^*_k for all sending states j not equal to k. This restriction will help identify the coefficients in the earnings equation by allowing just two correction functions, one for stayers and one for movers, to enter a state's earnings equation instead of N different functions. It should be noted, however, that this implementation choice is not essential to the general approach outlined in Section 3.[14]

4.2. *Using Cell Migration Flows for the Selection Probabilities*

The formulation of equation (19) assumes the researcher possesses consistent estimates of the relevant migration probabilities.[15] To estimate the selection probabilities in a polychotomous choice model, researchers have mainly used the conditional logit model.[16] Typically, the subutility functions $V_{j1}, \ldots, V_{jk}$ are modeled

[14] This assumption would not need to be made if more data and more cells were added, so that enough variation existed in the migration probabilities for each sending-state correction function λ^*_{jk}. I have explored allowing separate correction functions for different geographical regions of birth, and the empirical results are very similar.

[15] I provide a correction to the standard errors that accounts for the fact that estimates of the migration and retention probabilities are used instead of their true values in footnote 24.

[16] Note that a parametric specification of the selection equations does not preclude using the semiparametric approach outlined for the joint distribution of the error terms in the outcome and selection equations.

by specifying which variables affect mean earnings and mean tastes for a given migration path as well as specifying their functional forms.[17] Potential drawbacks to the conditional logit model are its previously mentioned independence of irrelevant alternatives property and its reliance on an assumed parametric framework.

To simplify estimation of the migration probabilities, I assume that mean earnings and mean tastes are the same for similar types of people. Suppose a vector of variables contains all the relevant attributes about a person's type, so that individuals with the same values for these person-type variables are identically affected by state-to-state differences in the subutility functions. This specification does not require the researcher to explicitly obtain variables for an overwhelming list of state amenities, but only to model an individual's type. Comparative advantage motivates this approach, with the prediction that individuals with different skills and characteristics will follow different migration paths on average.

If the vector describing an individual's type is composed only of discrete variables (as is often the case in labor economics), similar individuals can then be grouped into cells. Assignment into cells is made on the basis of the discrete characteristics such as age, schooling, marital status, the presence of children in the home, race, and sex. The intuition of this specification is simple: people with the same level of education who are the same age and similar in all other relevant characteristics are affected by differences in earnings, moving costs, state taxes, and other state amenities in the same way on average.

With this formulation, the subutility functions depend only on the cell of an individual. The migration probability for an individual belonging to a "cell" can be written as

$$\begin{aligned} p_{ijk} &= Pr(M_{ijk} = 1 | V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk}) \\ &= Pr(M_{ijk} = 1 | \text{cell}). \end{aligned} \tag{24}$$

A similar expression exists for the retention probability, p_{ijj}. An individual's migration probability is simply the fraction of individuals in the same cell who move from j to k. Since the model is fully saturated, note that a conditional logit or multinomial probit model would yield the same probabilities as these cell fractions. The advantage of grouping individuals into cells is that the form of the underlying subutility functions do not have to be specified if appropriately defined similar individuals have the same tastes and earnings on average. Additionally, because of the cell grouping assignment, estimation of the migration probabilities requires no distributional assumptions about the error terms $e_{ij1} - e_{ijk}, \ldots, e_{ijN} - e_{ijk}$. Other researchers using sample proportions to estimate selection probabilities in this way include Dynarski (1987) and Card and Payne (1998).

[17] An exception is Matzkin (1993), who proposes a nonparametric estimation method that does not require a parametric structure for the subutility functions.

4.3. *Estimating the Unknown Correction Functions*

Estimation of the outcome equations requires a suitable method for estimating the unknown correction functions. A variety of nonparametric techniques exist to estimate unknown functions; I employ series expansions for the correction functions (Andrews (1991), Newey (1988, 1997)).[18] Consider the approximation for movers into state k:

$$\lambda_k^*(p_{ijk}, p_{ijj}) \cong \sum_{t=1}^{T} \kappa_k^t b_k^t(p_{ijk}, p_{ijj}) \tag{25}$$

where the functions $b_k^t(\bullet)$ are referred to as the basis functions. Similar approximations exist for the stayers' correction functions. Two common choices for basis functions are the terms of a polynomial or Fourier series. Since both choices yield similar estimates in the current application, I report results using the polynomial expansion. Using series expansions results in a model that is linear in parameters and hence can be estimated by ordinary least squares. The number of basis functions should increase as the sample size increases, improving the accuracy of the approximation. In practice, the number of basis functions must be chosen by the researcher. Newey (1988, 1997) and Andrews (1991) give conditions on the model, basis functions, and number of observations needed so that the coefficients in the outcome equation are $\sqrt{n}$-consistent and asymptotically normal.

Newey (1994a) discusses the asymptotic variance of semiparametric estimators that involve nonparametric estimation of a function. For series estimators, the appropriate correction to the variance can be viewed as the standard parametric correction for estimation of the full set of coefficients in the regression equation (including the coefficients on the basis functions), with a fixed number of expansion terms (see pp. 1368–1372).[19] In other words, consistent estimates for the standard errors of the coefficients in the outcome equation can be read directly from standard regression output if the true migration probabilities are known. Since the true migration probabilities appearing in the basis functions are unknown, in the empirical work that follows I substitute estimated probabilities. I discuss how to correct the standard errors for this extra sampling variability in footnote 24.

The references to Andrews and Newey refer to a continuous variable; hence their results are only applicable when the number of distinct values for the probabilities is large. Consistency of the coefficient estimates requires that the number of unique probabilities entering the basis functions be sufficiently large. As described in the previous section, I divide individuals into cells to calculate migration probabilities. There must be enough cells to allow consistent estimation of the unknown correction function. Just as the number of basis functions should

[18] Estimating the unknown function semiparametrically can be compared to a parametric approach, such as Lee's, which specifies the joint distribution of the error terms and hence the functional form of the correction functions (see Appendix A).

[19] In addition, Newey (1994a, p. 1350) explains that the "method of estimating a function does not affect the asymptotic variance of the estimator."

increase as the sample size increases, one could also partition the dataset into finer cell groupings given a larger dataset. It should be noted, however, that there is a tradeoff between increasing the number of cells and the precision of the estimated migration probabilities. While it would be interesting to know how best to choose the size of cells and the number of basis functions when the correction function is a nuisance term, this question is beyond the scope of the current paper. In the results that follow, I use a relatively large number of cells and adjust the standard errors to account for the sampling variability of the migration probability estimates.

Before turning to the results, notice the proposed approach utilizes existing results and techniques for estimation. The key modeling insight is that an index sufficiency assumption allows for a dimensionality reduction that makes estimation possible using existing semiparametric methods. The approach avoids specifying the joint distribution of the outcome and selection errors and provides flexible estimation of the selection correction function. With only two choices, the index sufficiency assumption is automatically satisfied, and the model is an application of Ahn and Powell using series expansions. Finally, note that the model specifies single equation estimation for each state, which does not take into account any cross-equation restrictions on the coefficients or any cross-equation variance-covariance structure. If restrictions are available, the proposed estimation methodology is less efficient, but still consistent.

5. ESTIMATION

This section corrects and tests for selection bias in the returns to education caused by workers sorting themselves into different states. Estimation proceeds in three steps. First, I estimate the probability that an individual follows a given state-to-state migration path. In the second step, I use the approach developed in Section 3 and the implementation choices discussed in Section 4 to get corrected estimates of the return to college in the wage equations. In the third step, I test the appropriateness of the Roy model of migration and earnings, by estimating how migration flows respond to differences in the corrected returns to education and other amenities.

The model and empirical estimation of this paper considers migration between the fifty-one states and the returns to education in the fifty-one states. Since it is impractical to present detailed information for so many states, this section provides summary results for the fifty-one states and more detailed results for six illustrative states. The six chosen states are California, Florida, Illinois, Kansas, New York, and Texas, states chosen for their geographical and labor market diversity. Due to space limitations, I present results using 1990 Census data; results using 1980 data are generally similar and available from the author on request. Before turning to the three steps of estimation, I first provide a brief description of the data.

5.1. *Data*

This paper uses data from the 5 Percent Public-Use Sample of the 1990 U.S. decennial census (Ruggles and Sobeck (1997)). Since this dataset consists of a 1 in 20 random sample of the entire population, it has enough observations to track state-to-state migration paths fairly accurately. Previous research confirms that age, education, and family structure dramatically affect mobility, and evidence suggests that males and females, and blacks and whites, may migrate for different reasons.[20] To tighten the paper's focus, I restrict the data to white males, age 25 to 34, who were employed full-time.[21] Using this set of individuals not only helps to pinpoint the migration decision, but also controls for much of the variation in the wage equations. In this paper, I define mobility in terms of an individual's state of birth versus current state of residence. That is, an individual is considered to have migrated from state j to state k if he was born in state j and currently lives in state k.[22]

Summary statistics for the entire U.S. as well as the six representatives states appear in Table I. In 1990 around one-third of white males, age 25 to 34 and employed full-time, lived in a different state than where they were born. The table reveals a wide variation in the fraction of inmigrants making up a state's population as well as the fraction of outmigrants who leave their state of birth. For example, 37 percent of California's population were inmigrants in this sample, compared to only 12 percent for New York. The table also lists the percent of individuals with different levels of education by state, and details the fraction of individuals who are married or reside in a standard metropolitan statistical area. Row (11) lists average wages in 1990 dollars, a variable that ranges from a high of $14.38 per hour in California to a low of $10.29 in Kansas. It should be noted that regional price indices are not readily available, so it is difficult to directly compare wages across states. The lack of regional price indices provides another reason to focus on differences in the return to education rather than average wages across states.

5.2. *Step 1: Estimates of the Migration Probabilities*

The first step to correct for self-selection involves estimating the migration probabilities for individuals. As outlined in Section 4, with discrete variables,

[20] For a detailed explanation of the determinants of migration, see Sjaastad (1962), Chiswick (1974), and Robinson and Tomes (1982). For a breakdown of mobility rates by age, education, race, sex, and marital status, see *Geographical Mobility*, U.S. Bureau of the Census, series P-20.

[21] An individual was considered to be employed full-time if in the last year they: (i) were not currently enrolled full time in school, (ii) worked an average of 20 hours or more per week, (iii) worked for pay for at least ten weeks, and (iv) earned an annual salary of at least 2,000 dollars.

[22] An alternative to this definition would be to use the census question that asks which state an individual lived in five years prior to the census. Not surprisingly, the lifetime and five-year mobility definitions yield very similar results. Since the five-year results are estimated with less precision, I utilize the birth-state definition in the results that follow. For simplicity, I exclude individuals who are known to have moved more than once using information from these two measures.

TABLE I
SUMMARY STATISTICS

Variable	U.S.	California	Florida	Illinois	Kansas	New York	Texas
(1) Migrant (%)	31	—	—	—	—	—	—
	(0.1)						
(2) Inmigrant (%)	—	37	69	20	33	12	33
		(0.2)	(0.3)	(0.2)	(0.6)	(0.2)	(0.2)
(3) Outmigrant (%)	—	25	34	31	41	32	22
		(0.2)	(0.4)	(0.3)	(0.6)	(0.2)	(0.2)
(4) Less than High School (%)	13	11	15	9	10	10	15
	(0.1)	(0.1)	(0.2)	(0.2)	(0.4)	(0.2)	(0.2)
(5) High School (%)	38	28	35	37	40	35	33
	(0.1)	(0.2)	(0.3)	(0.3)	(0.6)	(0.2)	(0.2)
(6) Some College (%)	28	35	30	30	31	29	30
	(0.1)	(0.2)	(0.3)	(0.3)	(0.6)	(0.2)	(0.2)
(7) College Graduate (%)	17	19	16	19	16	20	18
	(0.1)	(0.2)	(0.2)	(0.2)	(0.5)	(0.2)	(0.2)
(8) Advanced Degree (%)	4	6	4	6	3	7	5
	(0.0)	(0.1)	(0.1)	(0.1)	(0.2)	(0.1)	(0.1)
(9) Married (%)	63	54	59	63	68	57	68
	(0.1)	(0.2)	(0.3)	(0.3)	(0.6)	(0.3)	(0.2)
(10) Residence in SMSA (%)	64	95	83	70	32	74	72
	(0.1)	(0.1)	(0.2)	(0.3)	(0.6)	(0.2)	(0.2)
(11) Wage	11.93	14.38	11.15	12.78	10.29	13.72	11.27
	(0.01)	(0.04)	(0.05)	(0.05)	(0.07)	(0.05)	(0.04)
(12) Observations	538,953	51,150	24,316	26,792	6,045	38,139	37,846

Note: Standard errors in parentheses.
Source: 1990 U.S. Census data for white males, age 25–34, and employed full-time.

individuals with similar characteristics who have similar costs and benefits for state-to-state migration can be grouped into cells. These cell assignments can then be used to estimate individual migration probabilities. The restrictions on the data set have already eliminated age, race, and sex as factors determining cell assignment; therefore, I assume that educational attainment and family circumstances define the cells for a given birth state.

I first divide the data into two categories: movers, or those who have moved into a state, and stayers, or those who were born in their state of residence. I also assign individuals into one of five education classes: less than high school, high school, some college, college degree, and advanced degree. Within each education class, I divide stayers into 14 mutually exclusive cells. Married stayers are grouped into eight cells based on whether they have a working spouse, children less than five years old, and children 5–18 years old. Non-married stayers are grouped into six cells based on whether they are divorced, and whether they live alone, with extended family (children, parents, grandparents, or siblings), or with a roommate/friend. These assignments result in 70 separate cells for each residence state.

Since there are fewer observations for movers, their groupings are coarser by necessity. I continue to assign individuals into one of the five education classes. Married movers are then divided based on whether children 18 years or younger are present in the home. Nonmarried movers are grouped based on whether they live with extended family. For each birth state, these characteristics divide movers into 20 cells. Since movers can originate from 50 different birth states, this assignment potentially creates 1,000 cells for inmigrants to a state. Not all migration paths will be observed for all cell types, and in practice there will be fewer than 1,000 cells for each residence state.

The fraction of individuals in a cell who migrate from one state to another estimates the probability that any individual in that cell will follow the same migration path. Variation in both stayers' and movers' migration probabilities plays a key role in identifying the education coefficients in the earning equation. For stayers in a given residence state, identification relies on variation in cell probabilities due to differing family circumstances. For movers, there is much less variation in family circumstances due to the coarser groupings, so identification relies more heavily on movers originating from different states who have differing cell probabilities for migration into a residence state. Due to variation in the migration probabilities because of differing family circumstances or different origination states, two individuals can have migration probabilities that are close, but have different levels of education. The variability in education given approximately equal selection bias terms identifies β_k. As a simple example, consider a person with a high school education who has the same migration probability as a person with a college education. A simple differences estimator would yield a consistent estimate of the return to a college education, since the selection bias term is the same for both individuals (Powell (1987, 1998)). The polynomial series expansions for the mover and stayer probabilities perform a similar role in separating out the selection bias.

Table II summarizes the overall variability in the cell migration probabilities for stayers and movers by education class for all 51 states. Looking at stayers first, there is a clear decrease in the cell probabilities as education level increases. For example, the average cell probability for high school dropouts is almost 70 percent, compared to less than 50 percent for individuals with an advanced degree. For movers, around one to three percent of the individuals belonging to a cell follow an average migration path. As expected, the average migration probability increases with education, reflecting the fact that highly educated individuals are more mobile. For both the stayers and the movers, there is wide variation in the cell probabilities within education classes, as is necessary for identification. While educated individuals are more likely to migrate, significant overlap across education classes exists in cell migration probabilities, as evidenced by the 90–10 percentile ranges. This overlap of probabilities plays an important role in estimation. If the cell probabilities for the five education classes did not overlap at all for a given state, the correction term coefficients could not separately be identified from the return to education coefficients in the earnings equation.

TABLE II
SUMMARY OF THE CELL MIGRATION PROBABILITIES

Education	Number of Cells[a]	Mean	Std. Dev.	10th Percentile	90th Percentile
		STAYERS			
Less than High School	616	0.6972	0.1243	0.5417	0.8361
High School Graduate	692	0.6790	0.1422	0.4783	0.8287
Some College	668	0.5997	0.1523	0.4000	0.7686
College Graduate	561	0.5325	0.1626	0.3158	0.7381
Advanced Degree	343	0.4857	0.1668	0.2857	0.7143
		MOVERS			
Less than High School	3923	0.0172	0.0281	0.0018	0.0429
High School Graduate	6090	0.0107	0.0218	0.0009	0.0261
Some College	5879	0.0136	0.0250	0.0012	0.0339
College Graduate	5159	0.0182	0.0311	0.0018	0.0436
Advanced Degree	3048	0.0298	0.0406	0.0038	0.0698

[a]Cells with 10 or fewer observations are excluded.

As expected, Table II confirms that not all migration paths are observed for all types of individuals. In particular, the number of usable cells is markedly smaller for the less than high school and advanced degree categories, since there are fewer individuals in these categories to begin with. It should also be noted that many of the cells for the smaller states do not contain any observations, For example, while California has 70 nonempty cells for stayers and over 800 nonempty cells for movers, Vermont has only 36 usable stayer cells and 190 usable mover cells. Therefore, the result for the smaller states may not be very informative.

5.2.1. *The Transition Matrix*

The estimated migration probabilities can be condensed by education class and grouped into a "transition matrix," where the rows are the birth states and the columns are the residence states. The transition matrix provides a convenient method of summarizing differences in migration paths for the different education classes. Unfortunately, the transition matrix for all fifty-one states is too large to present in the paper since it has 51×51 elements for each of the five education classes. A subset of the 51×51 transition matrix, representing the six states on which I focus, is provided graphically in Figure 1. The figure superimposes the information for each education class into a single matrix of bar graphs. The diagonal elements of the matrices correspond to the fraction of people by education class who stay in their state of birth and the off-diagonal elements correspond to the fraction who move from one state to another.

Looking first at the mobility patterns on the off-diagonal elements, the heights of the bars reveal that the fraction of individuals who follow a given migration path varies widely from one state to the next. California, Florida, and Texas stand out as more popular destinations than Illinois, Kansas, and New York. Although distance between states clearly plays a role, the flow magnitudes are

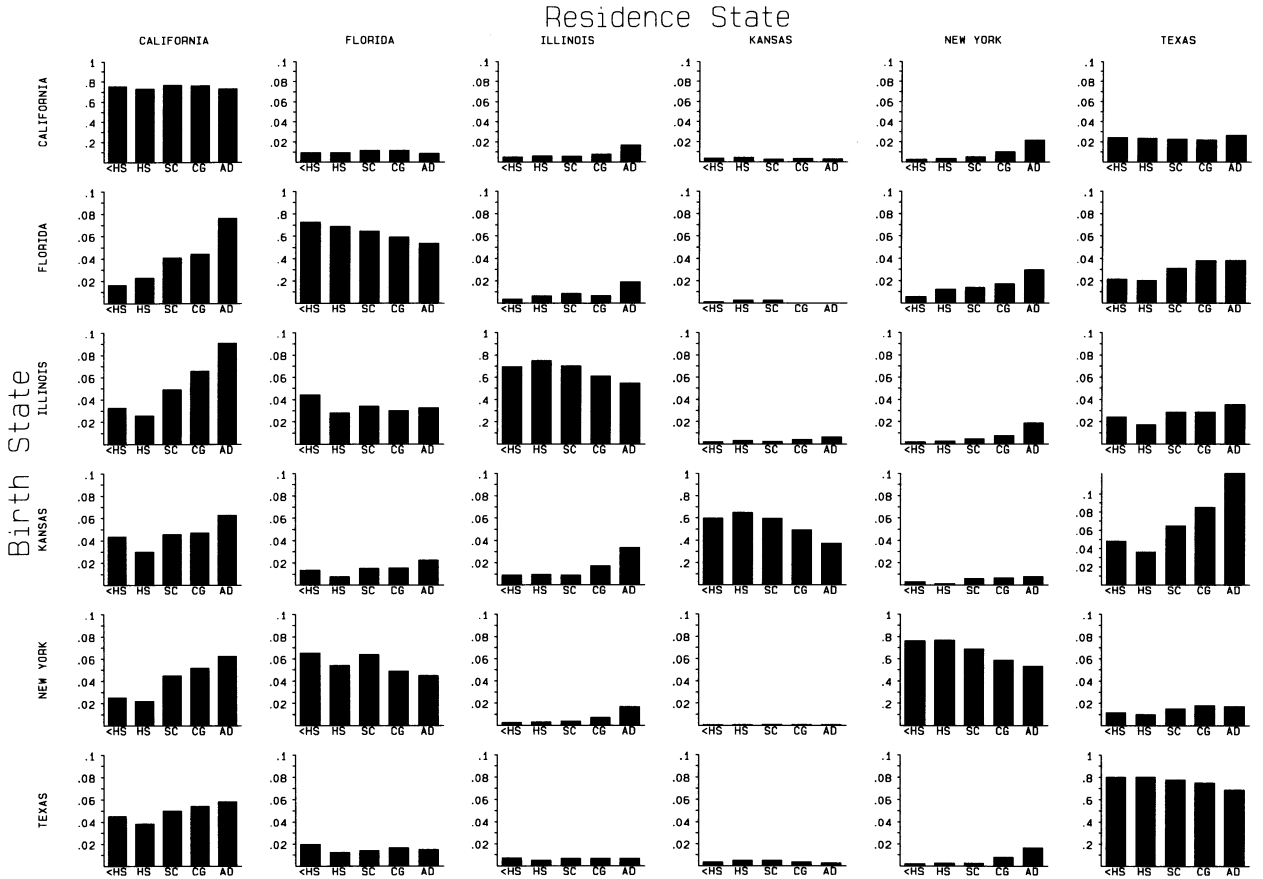

FIGURE 1.—Transition matrix for California, Florida, Illinois, Kansas, New York, and Texas.

not determined solely by geographical proximity. For example, people born in New York were much more likely to migrate to California, even though Illinois was a closer option. The migration flows in Figure 1 are also neither symmetric nor one-sided. For example, migration from Florida to New York is smaller than migration from New York to Florida, and there are considerable migration flows in both directions. While many models of migration predict migration in one direction only, bidirectional migration flows fit naturally into a Roy model where individuals relocate to take advantage of their particular skills and tastes.

Scanning a row horizontally reveals that workers from different education classes have different preferences on average about which potential residence state would suit them best. For example, consider the migration patterns for individuals born in California. For advanced degree holders, New York and Texas appear to be equally attractive destinations. However, for less educated individuals, Texas is five times as popular a destination compared to New York. This implies that for people born in California with little formal education, Texas is a relatively more attractive state for them compared to highly educated individuals, a pattern which supports a Roy model of comparative advantage by education level. Figure 1 also seems to indicate that tastes play a major role in mobility decisions, since comparative advantage in earnings cannot completely explain the observed mobility patterns. For example, the diagonal elements reveal that highly educated individuals are generally more mobile on average, yet there is little connection between education and outmigration in California. This is especially striking since, as will soon be shown, the returns to education in California are about the same as in Florida and New York, and lower than in Texas. These facts point out that tastes must be influencing the mobility decision—otherwise, why would highly educated Californians be so likely to stay?

5.3. *Step 2: Corrected Estimates of the Return to College*

With estimates of the migration probabilities, the earnings equations can now be estimated using the implementation choices outlined in Section 4. The dependent variable of earnings enters as the log of an individual's average hourly wage. The independent variables include potential experience along with its square and cube, a dummy for residence in a standard metropolitan statistical area (SMSA), a marital status dummy, and four education level dummies. The education categories are less than high school, some college, college graduate, and advanced degree, with high school graduate being the omitted category. It should be noted that the regression intercept includes the intercepts for the correction functions as well as any state-specific wage constant, since these cannot be separately identified.

For each of the 51 state regressions, there are separate correction functions for stayers and movers. Using the first-best choice probability as the single index appearing in the correction functions is the natural starting choice for estimation. As mentioned in Section 3, a simple test and extension is to allow a few other probabilities in addition to the first-best choice probability to enter the correction functions. For the mover correction function, it turns out that including the

retention probability in addition to the first-best probability generally results in a significant change in the estimated return to education coefficients and improves the fit of the outcome regressions, particularly for the larger states. The tests carried out to determine significance are a Hausman test for the change in the education coefficients and a Wald test for the difference due to the correction functions. The two tests are analogous to the ones described in footnotes 25 and 26. Other probabilities, such as the highest predicted probability (excluding the retention probability) or the probabilities for nearby states, do not have a similar effect. Therefore, the results presented in the following tables are estimated using the first-best probability (p_{ikk}) for stayers, and the first-best probability (p_{ijk}) plus the retention probability (p_{ijj}) for movers. Other estimates and the results of tests that include other probabilities in the correction function are available from the author on request. Both correction functions are estimated using second order polynomial expansions.[23]

Since estimates of an individual's migration probabilities substitute for the true values in the second-step earnings functions, the estimated coefficients are consistent but the estimated covariance matrix is biased. Naive standard errors are likely to be understated in the second step of a model that does not account for such sampling error (Murphy and Topel (1985)). I correct for the extra sampling variability arising from the imputed migration probabilities to obtain asymptotically correct standard errors.[24] The adjustment turns out to increase

[23] A polynomial of degree three yields comparable results, adding a little explanatory power at the expense of an increase in variance. A polynomial of degree one, however, is apparently not flexible enough. The correction function using a first degree polynomial does not enter the wage equation as significantly and the coefficients on the education variables do not change much compared to the uncorrected estimates.

[24] A feasible estimator of the asymptotically correct covariance matrix is

$$(\widehat{X}'\widehat{X})^{-1}\widehat{X}'\widehat{\Gamma}\widehat{V}(\widehat{P})\widehat{\Gamma}'\widehat{X}(\widehat{X}'\widehat{X})^{-1}+\hat{\sigma}^2(\widehat{X}'\widehat{X})^{-1}$$

where $\widehat{X}$ denotes the matrix of explanatory variables appearing in the wage equations, including the series expansion terms involving the estimated migration probabilities. The first term in the expression accounts for the sampling variability of the estimated probabilities and the second term is the usual covariance matrix (see Murphy and Topel (l985)). The second term could readily be extended to allow for heteroskedasticity using the Huber-White correction as suggested by Ham and Hsiao (1984). Note that the expression for the covariance substitutes in consistent estimates for the unknown parameters. Both $\widehat{\Gamma}$ and $\widehat{V}(\widehat{P})$ are block-diagonal matrices, where the diagonal blocks correspond to subgroups of data belonging to different cells (see equation (24)). Let n_c denote the sample size of cell c. Each block of $\widehat{\Gamma}$ is an $n_c \times 3$ matrix containing the derivatives of the correction functions with respect to the three migration probabilities evaluated at their estimated values. These derivatives are easily calculated for polynomial expansions. Each block of $\widehat{V}(\widehat{P})$ is an estimate of the 3×3 covariance matrix for the estimated mover, retention, and stayer probabilities for a cell. These cell covariance matrices are easily estimated since the migration probabilities are distributed as multinomial random variables. Note that except for the mover and retention probabilities corresponding to the same cell, all of the migration probabilities appearing in a single state's wage equation are estimated off of different samples. Hence, the estimated migration probabilities are uncorrelated across cells. Finally, notice that it is reasonable to assume that the projection errors associated with the migration probabilities are uncorrelated with the error term in the outcome equation since any individual's contribution to an estimated cell migration probability is small.

the estimated standard errors by as much as 20 percent for the imputed regressors appearing in the correction functions, but has a negligible effect for other variables in the earnings equations.

5.3.1. *Estimation Results for the Six States*

Presenting detailed regressions for all fifty-one states is infeasible, so I first provide results for the six selected states and then give summary results for all the states. This paper concentrates most heavily on migration and the return to education, specifically the return to college relative to high school. Of particular interest is whether substantial bias exists in the uncorrected education coefficients because of self-selection. Except for California, the education coefficients in the corrected equations are almost uniformly lower than the uncorrected coefficients in Table III. For example, while the coefficient on "college graduate" does not change much in California, for the other five states the coefficient drops by approximately 10 percent when correcting for self-selected migration. To see whether such decreases were statistically significant, a Hausman-type test was performed.[25] For all six states, the difference in the return to a college degree is significant at the one percent confidence level.

While testing for differences between the corrected and uncorrected coefficients provides a direct test for the presence of selection bias in the return to college, a necessary condition is that the selection correction terms enter the wage equation significantly. This paper uses a Wald test statistic, using the asymptotically correct covariance matrix, to test the impact of the correction terms.[26] The test statistics in row (12) of Table III indicate that the correction function enters significantly at the one percent confidence level for all of the states except Kansas. Although the most general test for selection bias depends jointly on the correction functions for movers and stayers, rows (10) and (11) present the appropriate Wald tests for the two correction functions separately. These subtests indicate that both the movers' and stayers' correction functions play important roles in removing selection bias from the wage equations.

[25] Under the null hypothesis of no selection bias, OLS is efficient and consistent and the selection corrected estimates are consistent but inefficient. Under the alternative, OLS is biased but the selection corrected estimates are still consistent. Therefore, the variance of the difference in the estimates is equal to the difference in the variances (Hausman (1978)). When testing for differences in individual coefficients, one should be aware of a multiple comparisons problem. That is, the covariance matrix for the full set of coefficients appearing in Table III can only have rank 7, since all of the differences arise from the 7 terms comprising the correction function. In this paper, I have chosen to test the difference between the uncorrected and corrected estimates of the return to college.

[26] The appropriate test statistic and its distribution is

$$\eta'[V(P)]^{-1}\eta \sim \chi^2 \quad \text{with degrees of freedom} = \text{rank}(\eta)$$

where η is the vector of coefficients for the terms in the correction function and $V(P)$ is the appropriate block from the covariance matrix of the estimated equation corresponding to the migration probability terms appearing in the correction function. In Table III, the chi-square distribution for row (10) has two degrees of freedom, for row (11) five degrees of freedom, and for row (12) seven degrees of freedom.

TABLE III

ESTIMATED WAGE EQUATIONS FOR CALIFORNIA, FLORIDA, ILLINOIS, KANSAS, NEW YORK, AND TEXAS

	California		Florida		Illinois	
	Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected	Corrected
(1) Less than High School	−0.1597	−0.1489	−0.1527	−0.1520	−0.1710	−0.1898
	(0.0082)	(0.0082)	(0.0101)	(0.0105)	(0.0113)	(0.0116)
(2) Some College	0.1383	0.1505	0.1337	0.1041	0.1165	0.0968
	(0.0059)	(0.0061)	(0.0080)	(0.0087)	(0.0076)	(0.0079)
(3) College Graduate	0.4378	0.4313	0.4485	0.4022	0.3645	0.3272
	(0.0075)	(0.0079)	(0.0106)	(0.0127)	(0.0100)	(0.0117)
(4) Advanced Degree	0.5996	0.5760	0.6407	0.5880	0.5461	0.5059
	(0.0110)	(0.0117)	(0.0172)	(0.0194)	(0.0147)	(0.0178)
(5) Experience	0.0778	0.0745	0.0663	0.0649	0.0580	0.0525
	(0.0075)	(0.0075)	(0.0107)	(0.0107)	(0.0097)	(0.0097)
(6) Experience Squared	−0.0023	−0.0023	−0.0024	−0.0023	−0.0008	−0.0005
	(0.0007)	(0.0007)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
(7) Experience Cubed × 100	−0.0001	0.0001	0.0018	0.0017	−0.0034	−0.0041
	(0.0018)	(0.0018)	(0.0026)	(0.0025)	(0.0026)	(0.0026)
(8) Married	0.1906	0.1438	0.1763	0.1714	0.1925	0.1736
	(0.0047)	(0.0056)	(0.0065)	(0.0070)	(0.0063)	(0.0069)
(9) Residence in SMSA	0.1754	0.1834	0.1146	0.1160	0.2496	0.2521
	(0.0109)	(0.0109)	(0.0084)	(0.0085)	(0.0067)	(0.0067)
(10) Wald test for λ (Movers Only)	—	88.29	—	87.72	—	49.75
		[0.0000]		[0.0000]		[0.000]
(11) Wald test for λ (Stayers Only)	—	1563.56	—	22.23	—	61.65
		[0.0000]		[0.0000]		[0.0000]
(12) Wald test for λ	—	463.56	—	117.57	—	109.96
		[0.0000]		[0.0000]		[0.0000]
(13) *R*-squared	0.1534	0.1606	0.1624	0.1668	0.1856	0.1891
(14) Observations	51,149	51,149	24,315	24,315	26,791	26,791

	Kansas		New York		Texas	
	Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected	Corrected
(1) Less than High School	−0.1887	−0.1933	−0.1958	−0.1985	−0.2023	−0.2046
	(0.0230)	(0.0233)	(0.0099)	(0.0099)	(0.0085)	(0.0085)
(2) Some College	0.0468	0.0349	0.1521	0.1297	0.1614	0.1356
	(0.0153)	(0.0165)	(0.0068)	(0.0074)	(0.0068)	(0.0071)
(3) College Graduate	0.3213	0.2863	0.4310	0.3977	0.5184	0.4697
	(0.0211)	(0.0253)	(0.0085)	(0.0107)	(0.0087)	(0.0098)
(4) Advanced Degree	0.4811	0.4122	0.5898	0.5495	0.6835	0.6130
	(0.0369)	(0.0462)	(0.0118)	(0.0145)	(0.0137)	(0.0153)
(5) Experience	0.0107	0.0110	0.0869	0.0820	0.0834	0.0811
	(0.0267)	(0.0268)	(0.0081)	(0.0081)	(0.0083)	(0.0083)
(6) Experience Squared	0.0028	−0.0028	−0.0041	−0.0038	−0.0029	−0.0027
	(0.0026)	(0.0026)	(0.0007)	(0.0007)	(0.0007)	(0.0007)

TABLE III—*Continued*

	Kansas		New York		Texas	
	Uncorrected	Corrected	Uncorrected	Corrected	Uncorrected	Corrected
(8) Married	0.1790	0.1820	0.1881	0.1748	0.1901	0.2001
	(0.0134)	(0.0135)	(0.0055)	(0.0058)	(0.0057)	(0.0058)
(9) Residence in SMSA	0.2308	0.2296	0.2209	0.2225	0.1234	0.1139
	(0.0135)	(0.0138)	(0.0061)	(0.0061)	(0.0060)	(0.0060)
(10) Wald test for λ	—	7.37	—	85.08	—	43.57
(Movers Only)		[0.1946]		[0.0000]		[0.0000]
(11) Wald test for λ	—	2.31	—	60.67	—	116.81
(Stayers Only)		[0.3156]		[0.0000]		[0.0000]
(12) Wald test for λ	—	8.34	—	132.59	—	110.39
		[0.3037]		[0.0000]		[0.0000]
(13) *R*-squared	0.1574	0.1589	0.1912	0.1938	0.1932	0.1974
(14) Observations	6,044	6,044	38,138	38,138	37,845	37,845

Note: Standard errors in parentheses, *p*-values in brackets; both adjusted for the sampling variability of the estimated migration probabilities appearing in the correction functions (see footnote 24).

To understand how the correction functions can affect the earnings equations, consider the shape of these functions for the state of Texas. Figure 2 plots the value of the correction as a function of the different observed migration probabilities.[27] Looking at the graph for stayers, the correction due to self-selection decreases as the probability of a stayer remaining in his birth state gets larger. In the lower panel of Figure 2, I graph the correction as a function of the first-best migration probability for movers, evaluated at the 20th, 40th, 60th, and 80th percentiles of the retention probability. For movers into Texas, the correction is larger for migrants with low retention probabilities. For a given percentile of the retention probability, the bias correction declines as the migration probability increases.[28] One explanation for the shape of the stayers' and movers' correction functions for Texas is that individuals who move to a state when few others like them do must have better earnings opportunities there than the average individual.[29]

[27] When interpreting the graphs, attention should focus on the shape and relative changes in the functions, since the intercept terms are not identified separately from the intercept in the wage equation.

[28] Around half of the 51 states exhibit a monotonic decline in the stayers' correction functions, with the next most frequent shape being a hump-shaped, inverse U. Similarly, around half of the states have correction functions for movers that roughly resemble Texas', with a wide variety of shapes for the remaining states.

[29] An example unrelated to mobility that exhibits such a monotonic decline is average Scholastic Aptitude Test scores. As the fraction of students in a state taking the SAT increases, the state-wide average performance on the test declines. The reasoning usually given is that as the fraction of test-takers increases, the quality of the marginal student declines (see Dynarski (1987), Card and Payne (1998)).

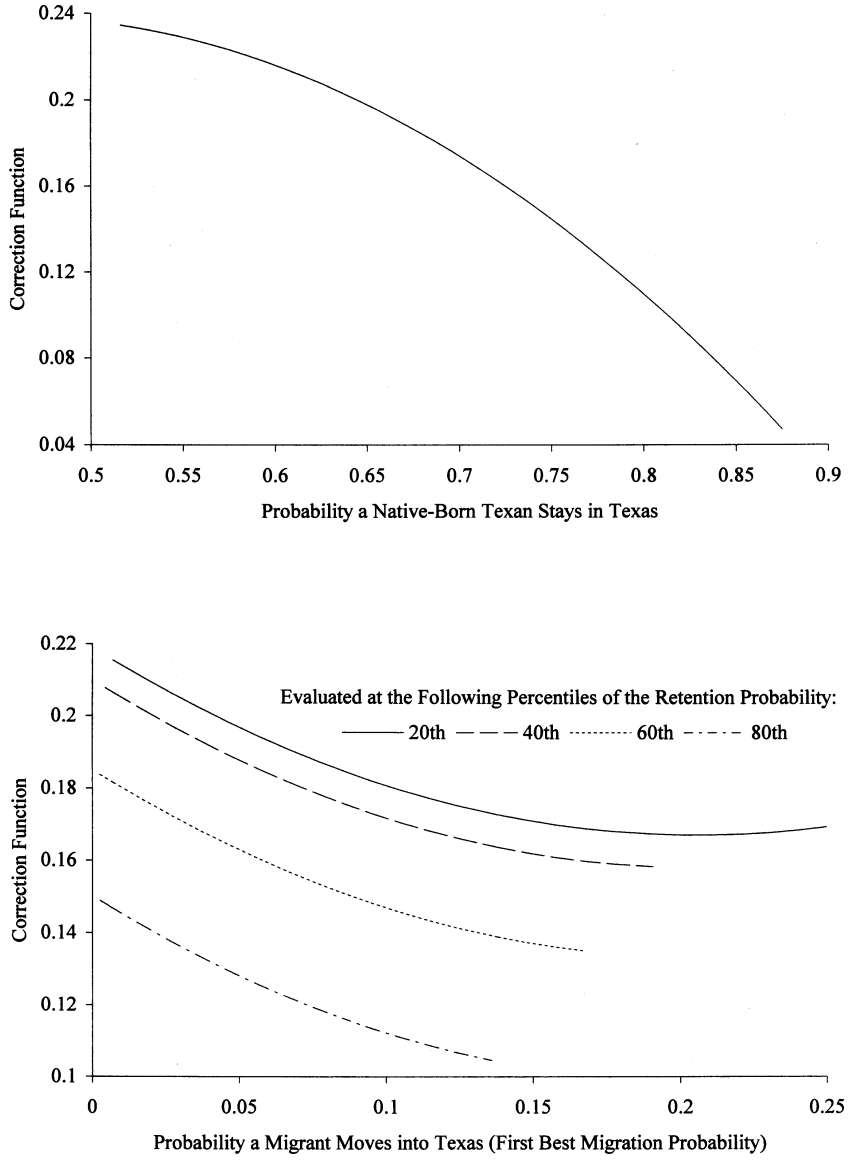

FIGURE 2.—Correction as a function of the migration probability for stayers and movers.

5.3.2. *Summary Results for the Entire U.S.*

To graphically illustrate the effect of selection bias on the estimated returns for all 51 states, Figure 3 plots the corrected versus uncorrected return to college for each state. All but seven of the estimated returns decrease when correcting for self-selected migration. Using a Wilcoxon signed-rank test, the corrected returns

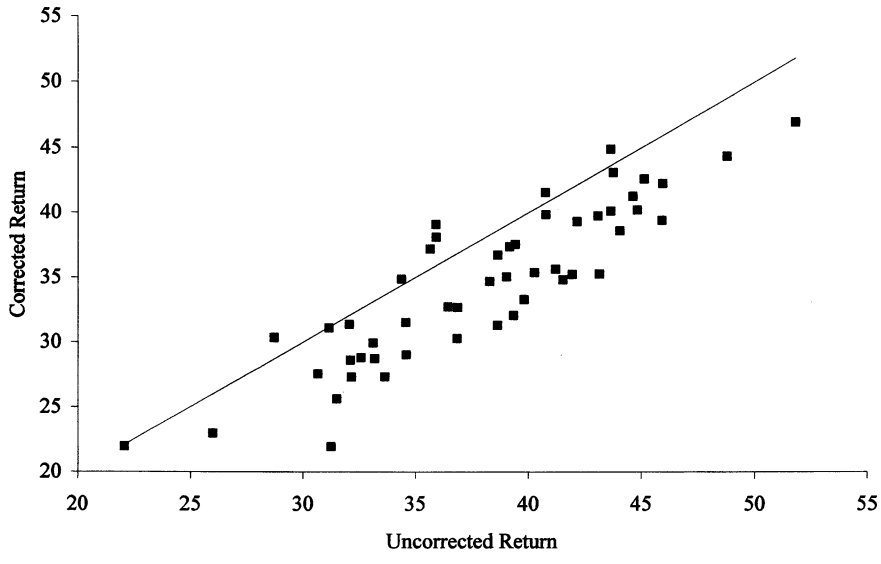

FIGURE 3.—Corrected versus uncorrected returns to a college education by state.

as a group are statistically different from the uncorrected returns at conventional significance levels. While the OLS estimates differ significantly from the corrected returns as a group, the corrected and uncorrected returns are strongly correlated ($\rho = 0.89$, p-value < 0.001).

For each state, Table IV presents the coefficients on the college education dummies from the corrected and uncorrected wage equations. For convenience, the states are listed alphabetically within the four regions of the U.S. The table confirms that with few exceptions, the naive uncorrected return to college is biased upward. While the average bias for all states is around nine percent, there is considerable heterogeneity between and within regions on the extent of the bias. The average bias appears to be larger in the Midwest and lower in the West. In 27 states the bias in the college education coefficient is significant at the one percent confidence level using a Hausman-type test (see footnote 25). Four additional states register a significant bias at the five percent level. The table also reports on the significance of the correction functions. For thirty-four states the correction terms enter the wage equation significantly at the five percent level using the Wald test described in footnote 26. Together with the significant changes in the return to college, these tests indicate that self-selected mobility plays an important role in earnings determination for many states.

One possible explanation for the upward bias in the OLS estimates is that college-educated individuals are more likely to sort into states that provide a better match for their particular skills and talents compared to those with a high school education. College migration choices might be more responsive to unobserved earnings because highly educated individuals are more likely to move for a fixed moving cost or because variation in unobserved earnings across states

TABLE IV

CORRECTED AND UNCORRECTED ESTIMATES OF THE RETURN TO COLLEGE BY STATE

State	Uncorrected College Return	Corrected College Return	Hausman Test for Difference	Wald Test for Correction Terms	State	Uncorrected College Return	Corrected College Return	Hausman Test for Difference	Wald Test for Correction Terms
		ALL REGIONS					MIDWEST		
Mean	38.1	34.7			Illinois	36.4	32.7***	6.04	109.96
Std. Dev.	5.96	6.05				(1.0)	(1.2)	[0.000]	[0.000]
					Indiana	36.9	30.3***	4.87	26.88
		NORTHEAST				(1.4)	(1.9)	[0.000]	[0.000]
					Iowa	31.5	25.6***	2.76	9.99
Connecticut	33.7	27.3***	5.67	70.47		(2.1)	(3.0)	[0.006]	[0.189]
	(1.6)	(1.9)	[0.000]	[0.000]	Kansas	32.1	28.6**	2.50	8.34
Maine	35.9	38.1	0.88	11.94		(2.1)	(2.5)	[0.012]	[0.304]
	(3.0)	(3.9)	[0.378]	[0.102]	Michigan	39.0	35.1***	4.20	68.74
Massachusetts	30.7	27.6***	5.79	76.48		(1.1)	(1.5)	[0.000]	[0.000]
	(1.1)	(1.2)	[0.000]	[0.000]	Minnesota	32.2	27.3***	4.11	43.49
New Hampshire	38.7	36.7	1.16	20.53		(1.5)	(1.9)	[0.000]	[0.000]
	(2.4)	(3.0)	[0.248]	[0.005]	Missouri	34.6	29.0***	4.00	37.82
New Jersey	39.8	33.3***	9.87	252.03		(1.6)	(2.1)	[0.000]	[0.000]
	(1.1)	(1.3)	[0.000]	[0.000]	Nebraska	32.6	28.8	1.53	6.73
New York	43.1	39.8***	5.08	132.59		(2.8)	(3.8)	[0.127]	[0.458]
	(0.9)	(1.1)	[0.000]	[0.000]	North Dakota	40.8	39.9	0.33	6.27
Pennsylvania	38.6	31.3***	8.82	102.60		(5.0)	(5.7)	[0.744]	[0.509]
	(0.9)	(1.2)	[0.000]	[0.000]	Ohio	39.4	32.1***	8.14	71.47
Rhode Island	36.9	32.7***	2.65	44.59		(1.0)	(1.3)	[0.000]	[0.000]
	(3.1)	(3.4)	[0.008]	[0.000]	South Dakota	40.8	41.6	0.36	6.26
Vermont	35.9	39.1	0.72	6.41		(4.7)	(5.2)	[0.672]	[0.510]
	(4.6)	(6.4)	[0.474]	[0.492]	Wisconsin	31.3	21.9***	6.69	104.54
Mean	37.0	34.0				(1.5)	(2.0)	[0.000]	[0.000]
Std. Dev.	3.6	4.7			Mean	35.6	31.1		
					Std. Dev.	3.7	5.6		

TABLE IV—*Continued*

State	Uncorrected College Return	Corrected College Return	Hausman Test for Difference	Wald Test for Correction Terms	State	Uncorrected College Return	Corrected College Return	Hausman Test for Difference	Wald Test for Correction Terms
		SOUTH					WEST		
Alabama	46.0	42.3***	3.02	16.16	Alaska	33.2	28.7	1.55	14.68
	(1.9)	(2.3)	[0.003]	[0.024]		(6.2)	(6.8)	[0.121]	[0.040]
Arkansas	38.3	34.7**	2.22	10.07	Arizona	48.8	44.4***	4.50	35.80
	(2.7)	(3.2)	[0.026]	[0.185]		(2.0)	(2.2)	[0.000]	[0.000]
Delaware	34.6	31.5	0.99	5.57	California	43.8	43.1**	2.43	463.56
	(4.1)	(5.1)	[0.324]	[0.591]		(0.7)	(0.8)	[0.015]	[0.000]
D.C.	33.1	30.0	0.72	6.68	Colorado	43.7	40.1***	3.16	12.85
	(12.1)	(12.9)	[0.472]	[0.463]		(1.8)	(2.1)	[0.002]	[0.076]
Florida	44.8	40.2***	6.60	117.57	Hawaii	43.7	44.9	0.45	3.82
	(1.1)	(1.3)	[0.000]	[0.000]		(6.1)	(6.7)	[0.654]	[0.801]
Georgia	41.2	35.7***	4.99	32.84	Idaho	28.7	30.3	0.86	6.27
	(1.4)	(1.8)	[0.000]	[0.000]		(4.2)	(4.6)	[0.391]	[0.508]
Kentucky	41.6	34.8***	4.77	22.44	Montana	26.0	23.0	0.90	9.54
	(2.1)	(2.5)	[0.000]	[0.002]		(4.8)	(5.9)	[0.368]	[0.217]
Louisiana	42.2	39.3**	1.99	5.42	Nevada	35.7	37.2	1.07	37.07
	(2.0)	(2.5)	[0.047]	[0.609]		(3.8)	(4.1)	[0.284]	[0.000]
Maryland	43.2	35.3***	8.50	93.66	New Mexico	45.1	42.6	1.17	16.79
	(1.4)	(1.7)	[0.000]	[0.000]		(3.5)	(4.1)	[0.243]	[0.019]
Mississippi	40.3	35.4**	4.28	25.26	Oregon	31.2	31.1	0.07	34.85
	(2.7)	(2.9)	[0.000]	[0.001]		(2.3)	(2.6)	[0.946]	[0.000]

North Carolina	44.1	38.6***	6.14	62.45	Utah	34.4	34.9	0.49	15.53
	(1.3)	(1.6)	[0.000]	[0.000]		(2.7)	(2.9)	[0.623]	[0.030]
Oklahoma	39.4	37.6	1.52	8.58	Washington	32.1	31.4	0.94	113.15
	(2.2)	(2.5)	[0.128]	[0.284]		(1.5)	(1.7)	[0.347]	[0.000]
South Carolina	39.2	37.4	1.51	22.92	Wyoming	22.1	22.0	0.03	5.29
	(2.0)	(2.3)	[0.132]	[0.002]		(6.5)	(8.0)	[0.977]	[0.625]
Tennessee	45.9	39.4***	5.47	33.69					
	(1.7)	(2.1)	[0.000]	[0.000]	Mean	36.0	34.9		
Texas	51.8	47.0***	11.27	110.39	Std. Dev.	8.3	7.9		
	(0.9)	(1.0)	[0.000]	[0.000]					
Virginia	42.0	35.2***	7.00	84.09					
	(1.4)	(1.7)	[0.000]	[0.000]					
West Virginia	44.6	41.3*	1.78	19.81					
	(3.3)	(3.8)	[0.076]	[0.006]					
Mean	41.9	37.4							
Std. Dev.	4.4	4.1							

Note: Standard errors in parentheses, *p*-values in brackets; both adjusted for the sampling variability of the estimated migration probabilities appearing in the correction functions (see footnote 24).

***Significantly different from the uncorrected estimate under the null hypothesis of no selection bias at the 1% level; **significantly different at the 5% level; *significantly different at the 10% level.

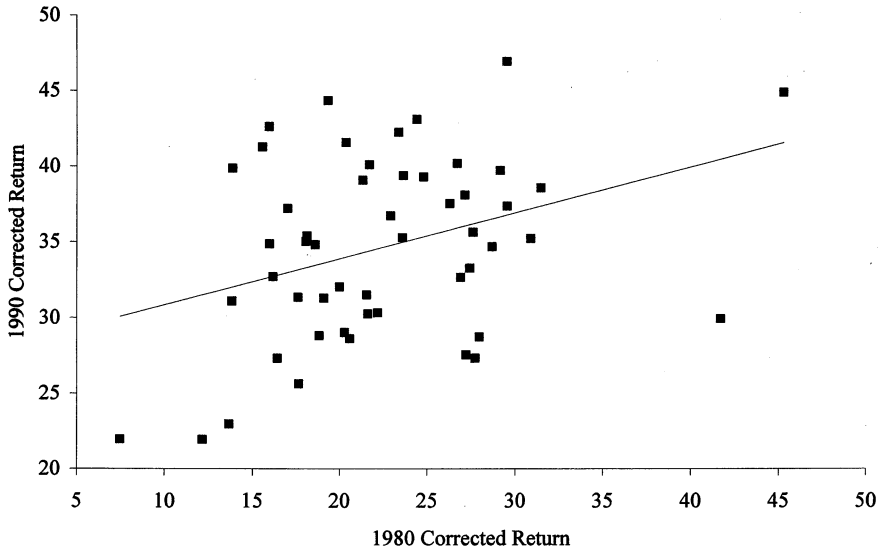

FIGURE 4.—Corrected returns to a college education by state in 1990 versus 1980.

is greater for individuals with a college degree. This could generate a positive correlation between schooling level and the error term in the wage regressions for the self-selected samples and hence an upward bias in the uncorrected estimates of the return to education.[30]

While the estimated returns to a college education are significantly biased, Figure 3 and Table IV document that correcting for the bias does not narrow the range of returns across states. This suggests that the productivity of a college-educated worker relative to a high school-educated worker varies from state to state, and that migration is ineffective in equalizing the wages of comparably educated labor across space. One explanation for such variation is that differences in the returns to a college education are equalizing differentials for amenity differences (Roback (1988)). If amenities remain largely unchanged over time, permanent differentials in the return to education could exist from state to state. Figure 4 examines the stability in the corrected returns by comparing the estimates in this paper to estimates using 1980 data. The graph plots the corrected returns in 1990 versus 1980, with the solid line in the graph representing a linear fit. The graph reveals that state-specific returns to college education are significantly correlated between 1980 and 1990 ($\rho = 0.35$, p-value $= 0.012$) and that the return to education has risen over time. One possible interpretation is that

[30] Of course, observing more college migration than high school migration does not guarantee a positive bias in the OLS estimates. For example, the high school individuals who move could have relatively larger error terms in the self-selected samples so that the correlation between schooling and the error term was not positive. More generally, in a Roy model with multiple sectors and selection based on earnings and tastes there are a variety of reasons the OLS estimates could be upward biased.

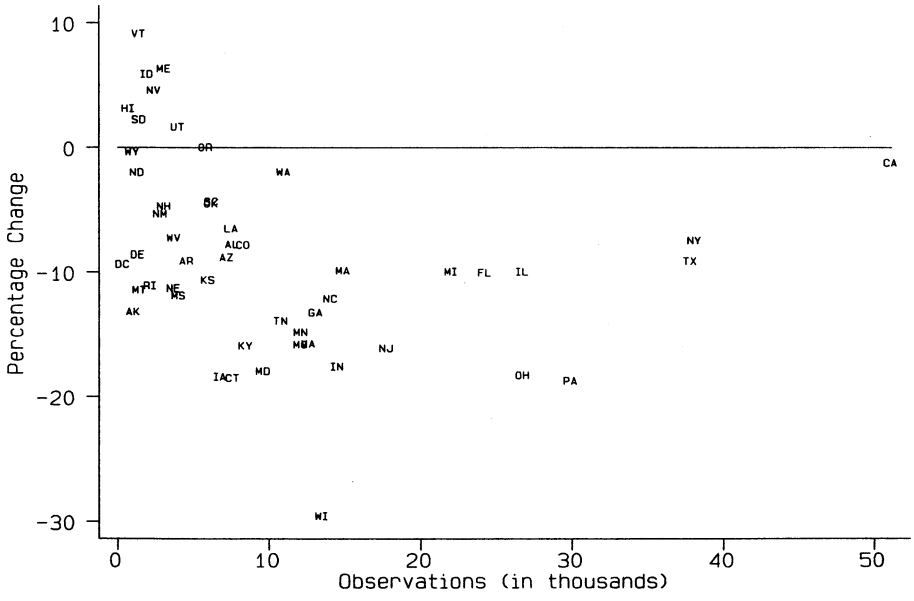

FIGURE 5.— Percentage change in the estimated return to a college education when correcting for selection bias versus the number of observations used in estimation by state.

the rise in the return to education was a nationwide increase that affected state-specific labor markets in a similar fashion (see Bound and Johnson (1992), Katz and Murphy (1992)).

How do the corrected estimates in a state depend on the number of observations? Consistency of the corrected estimates requires sufficient variability in the cell migration probabilities (see Section 4.3), and the number of distinct migration probabilities is strongly correlated with the size of a state. For each state, Figure 5 plots the percentage change in the return to a college education when correcting for selection bias versus the number of observations used in estimation. The plotting symbol is a state's two letter postal abbreviation. For states with few observations the estimated bias in the uncorrected returns is highly variable, ranging from roughly -13 percent to $+9$ percent. As expected, the results for these smaller states are estimated with less precision, with little evidence of selection bias using the tests reported in Table IV. However, for the larger states, the bias is overwhelmingly negative and significant. For example, in states with at least ten thousand observations, the average bias in the college education coefficient is -13 percent. These findings suggest the estimation approach developed in this paper is more suitable for reasonably large datasets. A similar conclusion emerges from the Monte Carlo simulations presented in Appendix C.

Another interesting question is how the semiparametric approach developed in this paper compares to the parametric approach developed by Lee. As a reminder, the results presented in Tables III and IV use series expansions to

estimate two unknown correction functions that depend on the first-best probability (p_{ikk}) for stayers and the first-best probability (p_{ijk}) plus the retention probability (p_{ijj}) for movers. In contrast, Lee's method involves two parametric correction functions that depend on the first-best probability for stayers and the first-best probability for movers. Since Lee's approach does not include the retention probability in the movers' correction function, it is not too surprising that the Lee estimates differ. The average return to education across all 51 states when using Lee's correction as outlined in Appendix A is 36.3 percent (std. dev. = 6.00). In comparison, the average return is 34.7 percent (std. dev. = 6.05) for the corrected estimates appearing in Table IV. Perhaps a more parallel comparison to the Lee estimates would be to use a semiparametric approach where the unknown correction function for movers is a function of only the first-best migration probability. These estimates are very similar to the Lee estimates, with the mean difference between the two estimates being less than 0.01 (std. dev. = 0.60) for the 51 states. Further evidence on when Lee's approach is likely to yield similar estimates to the semiparametric approach developed in this paper is discussed in the Monte Carlo appendix (Appendix C).[31]

5.4. *Step 3: Testing the Roy Model*

In this section, I test the appropriateness of the Roy model and the estimation approach taken in this paper. Using aggregate information for high school and college individuals, I estimate the responsiveness of migration flows to differences in corrected returns and amenities.

5.4.1. *A Model of Migration Flows and Amenity Differences*

If the true return to education differs across local labor markets and individuals behave according to comparative advantage, migration flows by education level should respond to differences in returns across states. In the spirit of the Roy model developed in this paper, an equation describing migration flows from state

[31] One could think about extending Lee's approach so the retention probability could be part of the selection correction function just as Heckman's method (1979) has been extended to higher dimensions (see Maddala (1983, pp. 278–283)). For example, the researcher could specify the inverse cumulative distribution functions corresponding to the first-best migration probability and the retention probability to both be inverse standard normals. The next requirement is to estimate the covariance term between the error terms associated with these two inverse CDFs. One way to do this is to construct residuals assuming linear probability models, i.e., by taking the difference between predicted probabilities and actual migration choices. The researcher could then specify that the error terms associated with these two inverse CDFs and the error term in the wage equation have a trivariate standard normal distribution. The approach involves numerical integration, and correction of the standard errors involves further numerical integration. This extension to Lee's approach yields estimates that are close, although slightly larger on average, compared to the semiparametric estimates appearing in Table IV. While a detailed discussion of this extension is beyond the scope of the paper, the estimates are available from the author on request.

j to state k for college-educated movers in terms of earnings and amenities is

$$\ln(p_{jk}^{CD}) = \theta_0^{CD} + \theta_1(y_k^{CD} - y_j^{CD}) + \theta_2^{CD}(A_k - A_j) + \theta_3^{CD} D_{jk} + \nu_{jk}^{CD} \tag{26}$$

where $\ln(\bullet)$ denotes the natural log, y_k^{CD} represents the average earnings of individuals with a college degree in state k, A_k is a vector of amenity variables associated with state k, D_{jk} is a vector of cost variables for moving from j to k, and ν_{jk}^{CD} is an error term. Note that for estimation, one would need to substitute in estimates of y_k^{CD} and y_j^{CD} since their true values are unavailable. Define a similar equation for individuals with a high school education, superscripting the appropriate variables and coefficients with an "HS" instead of a "CD". Equation (26) formalizes the assumption that migration flows are determined by earnings and amenity differences across states. Notice that schooling level does not change the package of amenities offered by a state. However, the value individuals in different education classes place on those amenities is expected to differ, which accounts for the education-specific coefficients on these variables. In contrast, while state-specific earnings depend on education level, the coefficient θ_1 is not superscripted by schooling level in the migration flow equations. The implication is that the log of college and high school migration flows respond identically to a given difference in earnings.

Differencing the log migration flows of college- and high school-educated individuals yields

$$\begin{aligned} \ln(p_{jk}^{CD}) - \ln(p_{jk}^{HS}) = {} & (\theta_0^{CD} - \theta_0^{HS}) + \theta_1(y_k^{CD} - y_k^{HS}) \\ & - \theta_1(y_j^{CD} - y_j^{HS}) + (\theta_2^{CD} - \theta_2^{HS})(A_k - A_j) \\ & + (\theta_3^{CD} - \theta_3^{HS}) D_{jk} + (\nu_{jk}^{CD} - \nu_{jk}^{HS}). \end{aligned} \tag{27}$$

Assuming the only component of earnings that differs by schooling level across states is the return to education, the expression $y_k^{CD} - y_k^{HS}$ represents the return to a college education relative to high school in state k. This relative return is simply the coefficient on the college dummy in the earnings equation, which I denote as β_k^{CD} for state k. Making this substitution and simplifying the notation of equation (27),

$$\ln(p_{jk}^{CD}) - \ln(p_{jk}^{HS}) = \theta_0 + \theta_1 \Delta\beta^{CD} + \theta_2 \Delta A + \theta_3 D_{jk} + \nu_{jk} \tag{28}$$

where $\theta_0 = \theta_0^{CD} - \theta_0^{HS}$, $\theta_2 = \theta_2^{CD} - \theta_2^{HS}$, $\theta_3 = \theta_3^{CD} - \theta_3^{HS}$, $\Delta A = A_k - A_j$, $\nu_{jk} = \nu_{jk}^{CD} - \nu_{jk}^{HS}$, and $\Delta\beta^{CD} = \beta_k^{CD} - \beta_j^{CD}$.

Since the true value of $\Delta\beta^{CD}$ is not available, I substitute an estimate into equation (28) using results from Step 2. The coefficient estimates from a simple OLS regression will be consistent, but the standard errors will be biased due to the extra sampling variability. Accounting for this sampling variability increases the standard error for the estimate of θ_1 between 21 and 78 percent, depending on the specification. The standard errors for the other estimated coefficients

TABLE V

RESPONSIVENESS OF COLLEGE RELATIVE TO HIGH SCHOOL MIGRATION FLOWS TO DIFFERENCES IN THE RETURN TO COLLEGE AND AMENITIES

Dependent Variable: $\ln(p_{jk}^{CD}) - \ln(p_{jk}^{HS})$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	0.51***	0.51***	0.54***	0.52***	0.52***	0.53***	0.52***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)
Δ Corrected Return to College	2.29***	—	2.38***	1.89***	3.45***	2.39***	2.89***
	(0.4)		(0.35)	(0.39)	(0.62)	(0.40)	(0.77)
Δ Uncorrected Return to College	—	2.24***	—	—	—	—	—
		(0.32)					
Distance in Miles	—	—	−0.23	—	—	—	0.10
			(0.22)				(0.26)
Δ Unemployment Rate	—	—	−3.18**	—	—	—	−1.03
			(1.45)				(2.66)
Included Amenity Variables							
Quality of Life[a]				×			×
Climate[b]					×		×
State Spending and Taxing[c]						×	×
F test for Amenity Variables	—	—	—	10.83	3.92	7.70	5.44
				[.0000]	[.0000]	[.0000]	[.0000]
R-squared	0.0452	0.0479	0.0507	0.1325	0.0907	0.0855	0.1721
Observations	1,871	1,871	1,871	1,871	1,871	1,706	1,706

Notes: Huber-White standard errors in parentheses, p-values in brackets; the standard errors and F tests are adjusted for the sampling variability of the estimated state returns to a college education (see footnote 32). The symbol Δ represents the difference operator for the value of a variable between state k and state j. All explanatory variables are averages of 1980 and 1990 values except for the climate variables, which are already long-term averages. See Appendix D for variable sources and definitions.

[a]Quality of Life variables are Δ Population Density, Δ Doctors per Capita, Δ Dentists per Capita, Δ Hospital Costs, Δ Teacher's Salaries, Δ School Expenditures per Capita, Δ School Expenditures per Pupil, Δ Crime Rate, Δ Violent Crime Rate, and Δ Incarceration Rate.

[b]Climate variables are Δ Average Temperature, Δ Maximum Temperature, Δ Minimum Temperature, Δ Afternoon Humidity, Δ Annual Precipitation, Δ Number of Rainy Days, Δ Number of Sunny Days, and Δ Average Wind Speed.

[c]State Spending and Taxing variables are Δ State Spending on Education, Δ State Spending on Health and Human Services, Δ State Spending on Highways, Δ State Spending on Public Welfare, Δ Miscellaneous State Spending, Δ State Sales Tax, and Δ Average State Income Tax.

***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

increase anywhere from 9 to 106 percent, depending on the coefficient and specification. The standard errors are also corrected to account for possible heteroskedasticity.[32]

[32] A feasible estimator for the asymptotically correct covariance matrix is

$$(\widehat{X}'\widehat{X})^{-1}\widehat{X}'\hat{\theta}^2\widehat{V}(\Delta\hat{\beta}^{CD})\widehat{X}(\widehat{X}'\widehat{X})^{-1} + (\widehat{X}'\widehat{X})^{-1}\sum_i \hat{x}_i\hat{x}_i' r_i^2 (\widehat{X}'\widehat{X})^{-1}$$

where $\widehat{X}$ is the matrix of the explanatory variables appearing in equation (28), with $\Delta\hat{\beta}^{CD}$ substituting for $\Delta\beta^{CD}$, and r_i is the estimated residual for observation i. The expression is analogous to that in footnote 24, with a few important differences. First, the second term in the expression is the Huber-White covariance matrix to account for possible heteroskedasticity (White (1980)). Second, the derivative with respect to $\Delta\hat{\beta}^{CD}$ is a constant term. Third, the estimated covariance matrix of $\Delta\hat{\beta}^{CD}$, while easy to calculate, is not a block-diagonal matrix. To understand the elements of this matrix, first note the returns from which $\Delta\hat{\beta}^{CD}$ is formed are estimated on different samples, with the exception that the migration probabilities used in the correction functions may share common denominators. With the assumption that the denominator in the migration probability (i.e., the total number of observations in a cell) does not create a correlation in the returns to education, these

5.4.2. *Estimation Results*

Table V lists the estimation results for equation (28) using a variety of return to college and amenity variable combinations. All specifications indicate that migration is at least partly driven by comparative advantage, with state differences in college returns significantly influencing migration flows. The intercept is positive in every specification, confirming that the college group is more mobile than the high school group on average. Throughout Table V, the coefficients on the variables should be interpreted as how college migration responds to a given difference in earnings or amenities across states relative to high school migration.

Columns (1) and (2) examine how migration responds to differences in corrected and uncorrected returns to college. Both the corrected and uncorrected state-to-state differences in returns significantly explain migration flows. To understand the impact these return differences have on migration patterns, consider the coefficient estimate from column (1). For a one-standard deviation increase (0.0724) in the difference in the corrected return to college, the percentage change in the high school migration probability subtracted from the percentage change in the college migration probability increases by 16.6 percentage points. Notice the corrected returns do not induce spurious correlation as they would if they were mechanically related to the migration probabilities. That is, the corrected returns are not equal to the uncorrected returns plus some function of the migration probabilities. In fact, the estimates of the corrected returns partial out the effects of the correction functions (which in turn depend on the migration probabilities).

Is the cell assignment assumption based on education level plausible? The remaining specifications in Table V test if college- and high school-educated individuals are differentially motivated by amenity differences. Column (3) adds in a distance variable as well as state differences in the five-year average unemployment rate. Column (4) includes measures of population density differences, crime rate differences, and other quality of life differences from state to state. Column (5) adds in climate difference variables and column (6) adds in differences in state spending and taxing measures. Column (7) includes all of the variables of columns (3) through (6). Details of the amenity variables and how they are constructed are found in Appendix D. Many of the variables appear to significantly affect migration flow differences, although the coefficients are often imprecisely estimated and not robust to the inclusion or exclusion of other variables.[33] Table V indicates that migration is driven by amenity differences in general, however, with these variables being jointly significant in every specification.

---

returns are independent. This assumption merely says that the total cell population in a state does not affect the return to education in other states. This implies the covariance matrix for $\Delta\hat{\beta}^{CD}$ is composed of the negatively- and positively-signed variances for the return coefficients $\hat{\beta}_1^{CD}, \ldots, \hat{\beta}_N^{CD}$. For example, the covariance between $\hat{\beta}_3^{CD} - \hat{\beta}_7^{CD}$ and $\hat{\beta}_6^{CD} - \hat{\beta}_3^{CD}$ equals minus the variance of $\hat{\beta}_3^{CD}$, a variance that was consistently estimated in Step 2.

[33] The individual coefficients for these amenity variables are not reported in Table V due to space considerations, but are available from the author on request.

In the specifications that include the amenity variables, the coefficient on the difference in the return to college, θ_1, continues to be estimated as a positive and significant value. The magnitude suggests a quantitatively important effect of college return differentials on migration flows. To see this, note that the number of college-educated individuals in the sample who move from their birth state to another state due to differences in the return to college is

$$\sum_{\forall k}\sum_{\forall j}\left[p_{jk}POP_j - \frac{p_{jk}POP_j}{1+I(\beta_k^{CD}>\beta_j^{CD})\theta_1(\beta_k^{CD}-\beta_j^{CD})}\right] \tag{29}$$

where $I(\bullet)$ is an indicator function and POP_j is the number of potential migrants in state j. The first term in brackets is the total number of college-educated individuals who migrate from state j to k, while the second term is the number of college-educated individuals who migrate for reasons other than the difference in returns. I estimate this quantity using the corrected college return coefficients from Table IV, the predicted value for θ_1 in column (7) of Table V, and the number of college-educated individuals born in each state in the sample. The calculation implies that 3.9 percent of the college-educated men in the sample moved in response to a higher return in another state. Looked at another way, the results indicate that state differences in the return to education account for 9.6 percent of the migration of college-educated individuals. While much of the variation in mobility choices remains unexplained, the strong and robust effect of the return differentials in explaining migration flows supports a Roy model of comparative advantage.

6. CONCLUSION

This paper provides a simple way to model and correct for selection bias when there are many choices. Two questions motivate the application: how does self-selected migration affect the observed returns to education in state-specific labor markets, and does self-selection play a role in explaining the wide range of returns across states? To answer these questions, I outline a Roy model of mobility and earnings with multiple sectors. To correct for sample selection bias in a model with so many alternatives, I develop and apply a new semiparametric technique. My main methodological insight is that a combination of Lee's maximum order statistic approach with the use of selection probabilities in an index framework results in a flexible semiparametric correction for selectivity bias in polychotomous choice models. In its simplest form, the selection correction takes the form of an unknown function of the first-best choice probability. An extension allows a subset of the other probabilities to also enter the correction function.

My main substantive finding is that self-selection significantly biases the observed returns to schooling in state-specific labor markets. Tests of the Roy model support the role of comparative advantage by education level in mobility decisions. I find that self-selection of more highly educated people to states

with higher returns to education generally leads to *upward* biases in the return to a college education, in many cases by 10 to 20 percent. However, the variation between states in returns does not narrow, suggesting that state-specific amenities and other unmeasurable nonwage variables play important roles in the migration decisions of individuals with different levels of education. Consistent with a range of corrected returns, I find that relative (college to high school) migration flows respond significantly to differences in the corrected return to education and differences in amenities across states.

It should be emphasized that the new semiparametric approach I develop in this paper is not specific to my model of mobility and earnings. My methodology can also be applied to more conventional choice models, where the selection probabilities are not estimated semiparametrically by the fraction of individuals who choose different alternatives. For example, my basic approach could be applied to a Roy model of occupational choice, where individuals choose from many alternative careers based on comparative advantage. The researcher could estimate the choice probabilities using a conditional logit model and then use the first-best probability as the single index appearing in the correction function. While such an approach parameterizes the selection equations, it is flexible regarding the joint distribution of the error terms in the outcome and multiple selection equations and hence the form of the selection correction function. Future research could adapt the semiparametric correction method of this paper to a variety of other polychotomous choice models.

Dept. of Economics, University of Rochester, 229 Harkness Hall, Rochester, NY 14627, U.S.A.; dahl@troi.cc.rochester.edu.

Manuscript received November, 1998; final revision received January, 2002.

APPENDIX A: LEE'S METHODOLOGY

Using the same notation as Section 3, this appendix describes Lee's parametric approach for correcting selection bias. While the random variables $\max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk})$ indexed over i are not identically distributed, Lee points out that new random variables that are identically distributed can be constructed using the transformation $\epsilon_{ijk} = G_{jk}^{\epsilon-1}\{H_{jk}(0|V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk})\}$, where G_{jk}^{ϵ} is any continuous univariate cumulative distribution function. The selection rule can be expressed in the following equivalent ways:

$$
\begin{aligned}
y_{ik} \text{ observed} \quad \text{if and only if} \quad & e_{ijm} - e_{ijk} \le V_{ijk} - V_{ijm} \quad \forall\, m \\
& \text{or } \max_m(V_{jm} - V_{jk} + e_{ijm} - e_{ijk}) \le 0 \\
& \text{or } \epsilon_{ijk} \le G_{jk}^{\epsilon-1}\{F_{jk}^{e}(V_{j1} - V_{jk}, \ldots, V_{jN} - V_{jk})\}.
\end{aligned}
$$

The final equivalence reduces the N error terms from the selection criteria into the single error term, ϵ_{ijk}. By construction, G_{jk}^{ϵ} is a well-defined marginal distribution for ϵ_{ijk}.

Lee next makes the critical assumption that the random vectors (u_{ik}, ϵ_{ijk}) indexed over i are independent and identically distributed with joint distribution function G_{jk}, thus specifying the form of F_{jk}, the joint distribution for $(u_{ik}, e_{ij1} - e_{ijk}, \ldots, e_{ijN} - e_{ijk})$. Given the equivalent formulations of

the selection rule above, the cumulative distribution function F_{jk} can be expressed in the following ways:

$$\begin{aligned}F_{jk}(r, V_{j1}-V_{jk}, \ldots, V_{jN}-V_{jk}) &= Pr(u_{ik} < r, e_{ij1}-e_{ijk} < V_{j1}-V_{jk}, \ldots, e_{ijN}-e_{ijk} \\ &\qquad < V_{jN}-V_{jk}) \\ &= Pr\big(u_{ik} < r, \epsilon_{ijk} < G_{jk}^{\epsilon-1}[F_{jk}^{e}(V_{j1}-V_{jk}, \ldots, V_{jN}-V_{jk})]\big) \\ &= G_{jk}\big(r, G_{jk}^{\epsilon-1}[F_{jk}^{e}(V_{j1}-V_{jk}, \ldots, V_{jN}-V_{jk})]\big).\end{aligned}$$

Lee's assumption allows $e_{ij1}, \ldots, e_{ijN}$ to affect the distribution function F_{jk} only through the N-to-1 distribution function F_{jk}^{e}. The distributional transformation of the maximum and an assumed distribution for G_{jk} completely specifies the form of the selection correction.

The standard parametric assumptions for Lee's approach are: (i) specify G_{jk}^{ϵ} to be a univariate standard normal CDF, and (ii) specify the joint distribution of u_{ik} and ϵ_{ijk} to be bivariate standard normal. The expression for the expectation of u_{ik} conditional on sample selection is then

$$\begin{aligned}&E[u_{ik}|s_i, M_{ijk}=1, V_{j1}-V_{jk}, \ldots, V_{jN}-V_{jk}] \\ &\quad = \rho_{jk}\pi_{jk}[F_{jk}^{e}(V_{j1}-V_{jk}, \ldots, V_{jN}-V_{jk})] \qquad (k=1, \ldots, N),\end{aligned}$$

where ρ_{jk} is the correlation between u_{ik} and ϵ_{ijk} and $\pi_{jk}(F_{jk}^{e}) = -\phi[\Phi^{-1}(F_{jk}^{e})]/F_{jk}^{e}$, with ϕ and Φ representing the normal PDF and CDF respectively. Hence the conditional expectation of y_{ik} is

$$\begin{aligned}E[y_{ik}|s_i, M_{ijk}=1, V_{j1}-V_{jk}, \ldots, V_{jN}-V_{jk}] &= \alpha_k + x_i'\delta_k + s_i\beta_k + \rho_{jk}\pi_{jk}(F_{jk}^{e}) \\ &= \alpha_k + x_i'\delta_k + s_i\beta_k + \rho_{jk}\pi_{jk}(p_{ijk}).\end{aligned}$$

APPENDIX B: PROOF THAT $\psi_{jk}(V_{j1}-V_{jk}, \ldots, V_{jN}-V_{jk}) = \lambda_{jk}(p_{ijk})$

This appendix proves that under the index sufficiency assumption (A-2), the multiple index correction function of equation (13) can be reduced to the single index correction function of equation (16). For notational brevity let $\vec{V}$ denote $V_{j1}-V_{jk}, \ldots, V_{jN}-V_{jk}$. Since $\psi_{jk}(\vec{V}) = E[u_{ik}|M_{ijk}=1]$ is a function of the conditional density $f_{jk}(u_{ik}, e_{ij1}-e_{ijk}, \ldots, e_{ijN}-e_{ijk}|M_{ijk}, \vec{V})$ and since

$$f_{jk}(u_{ik}, e_{ij1}-e_{ijk}, \ldots, e_{ijN}-e_{ijk}|M_{ijk}, \vec{V}) = g_{jk}(u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})|M_{ijk}, \vec{V}),$$

it suffices to show that

$$\begin{aligned}&g_{jk}(u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})|M_{ijk}, \vec{V}) \\ &\quad = g_{jk}(u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})|M_{ijk}, p_{ijk}).\end{aligned}$$

By Bayes Theorem,

$$\begin{aligned}&g_{jk}\Big(u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})|M_{ijk}, \vec{V}\Big) \\ &\quad = \frac{P(M_{ijk}=1|u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk}), \vec{V}) \times g_{jk}(u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})|\vec{V})}{P(M_{ijk}=1|\vec{V})}.\end{aligned}$$

The denominator of this expression, $P(M_{ijk}=1|\vec{V})$, equals p_{ijk}. By the index sufficiency assumption (A-2),

$$g_{jk}\Big(u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})|\vec{V}\Big) = g_{jk}\Big(u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})|p_{ijk}\Big).$$

By definition, the event $\{M_{ijk}=1\}$ is equivalent to $\{\max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk}) \le 0\}$, an expression that depends only on $\max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk})$. Therefore, $P(M_{ijk}=1|u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk}), \vec{V}) = P(M_{ijk}=1|u_{ik}, \max_m(V_{jm}-V_{jk}+e_{ijm}-e_{ijk}), p_{ijk})$ and the result is proved.

APPENDIX C: MONTE CARLO SIMULATIONS

Baseline Monte Carlo Design

Consider data generated from a Roy model of self-selection similar to the model of mobility and earnings described in Section 2. A simplified representation capturing the essence of this data-generating process is:

$$y_{ik} = s_i\beta_k + u_{ik} \qquad (k = 1, \ldots, N),$$

$$u_{ik} = \tau_k a_i + b_{ik} \qquad (k = 1, \ldots, N),$$

$$t_{ijk} = z_i\gamma_{jk} + w_{ijk} \qquad (k = 1, \ldots, N),$$

$$V_{ijk} = y_{ik} + t_{ijk} \qquad (k = 1, \ldots, N),$$

$$y_{ik} \text{ observed} \quad \text{if and only if} \quad V_{ijk} \geq V_{ijm} \quad \forall\, m$$

where the same notation is used as before. The exogenous variable s_i takes on integer values between 1 and 5 with equal probability, while z_i takes on integer values between 1 and 10 with equal probability. Without loss of generality, the simulations that follow focus on estimation of the outcome equation for $k = 1$.

The form and severity of selectivity bias depends on the coefficient vectors and the assumptions made regarding the distributions of the error terms. Varying the error distributions reveals how well the proposed method performs when the index sufficiency assumption (A-2) holds as well as when it does not. The baseline model for discussion is $a_i \sim N(0,1)$, $b_{ik} \sim N(0,1)$, $w_{ijk} \sim N(0,1)$, and $\tau_k = 1$ for $k = 1, \ldots, N$. In the baseline model a_i is independent over i, b_{ik} is independent over i and over k, w_{ijk} is independent over i, over j, and over k, and a_i, b_{ik}, and w_{ijk} are all independent of each other. This baseline specification satisfies assumption (A-2) as noted in Section 3.3. In addition, in the baseline model the correction functions are identical for all originating sectors (i.e., $\lambda_{jk} = \lambda_k, \forall\, j$).

The purpose of the simulations is to assess how well the method proposed in this paper can estimate β_1 in the presence of self-selection. While there are many parameters that could be varied in the simulations, this paper focuses on three aspects of polychotomous choice models: (i) the number of alternatives, (ii) the distribution of the error terms u_{ik} and w_{ijk}, (iii) and the sample size. To assess the large and small sample properties of the estimation method, I consider self-selected samples generated by models with 10,000 and 1,000 observations originating in each sector. I examine models that satisfy the index sufficiency assumption as well as models that violate it.

The direction and size of the selection bias also depends on the full set of coefficients in the model. However, to maintain focus on the primary objectives of this exercise, these coefficient vectors are not treated as experimental parameters. Instead, arbitrary values were chosen for the true values of the coefficients, with β_2 to β_N ranging from 0 to 2, and each γ_{jk} ranging from $-.25$ to .25. It is impractical to report all of these coefficients since there are $N + N^2$ coefficients that vary in an N-sector baseline model. These coefficients are available from the author on request. In all of the simulations the parameter of interest, β_1, equals 1.

Migration probabilities are estimated by grouping observations into cells based on the discrete variables s_i and z_i, and calculating cell fractions as described in Section 4.2. Corrected estimates are presented for the method of this paper using polynomial expansions of the first-best migration probability (labeled "1st Best Probability" in the table), the Lee approach as developed in Appendix A assuming joint normality, and second order polynomial expansions that include the products and cross-products of other migration probabilities as well. Except where noted, a single correction function is used instead of separate correction functions for each originating sector j. This results in an efficiency gain, since by construction these correction functions will be identical in all but one of the simulations. For further details on these implementation choices, see Section 4.

Choice Models of Various Dimensions

To see how well the estimation method performs as the dimensionality of the model grows, the first panel of the Appendix Table considers baseline models with $2, 3, 5, 10, 25$, and 50 sectors. Caution

should be exercised in comparing the amount of selectivity bias across specifications with differing numbers of sectors. Since higher dimensional models have additional parameters, the severity of the bias is not directly comparable. What can be seen from the table, however, is that the method developed in this paper is able to take care of selection bias in low-dimensional as well as high-dimensional models.

The Appendix Table begins by reporting simulation results for a two sector baseline model. Because there is only one migration probability (for the single choice) for each observation, no index sufficiency assumption is required. The results for the two sector model can be viewed as a comparison benchmark, since for a single choice model the estimation method of this paper is just an application of Ahn and Powell (1993) with a slightly different semiparametric estimation method. As expected, the OLS estimate is significantly biased away from the true value of $\beta_1 = 1$. Both the Lee approach and the semiparametric approach of this paper eliminate the bias in large samples and a majority of the bias in small samples.

Specifications (2) through (6) examine baseline models of various dimensions. Once again, the OLS estimates are significantly biased. The estimator relying on the first-best migration probability eliminates most of the bias, even as the number of choices under consideration increases. For the 25 and 50 sector models, the selection corrected estimates appearing in the table only eliminate around 90% of the bias. Although not reported in the tables, as the sample size increases, the remaining bias in these estimates disappears. For example, if there are 20,000 observations per sector, the first best probability estimate is 0.999 in the 25 sector model and 1.012 in the 50 sector model. In the smaller samples (1,000 observations per sector), the estimator eliminates only a fraction of the bias and the reported standard deviations are large. For all of the baseline models, the Lee estimator yields very similar results.

Deviations from Baseline in a Two Choice Model

The second panel in the Appendix Table examines the effect of a variety of distributional assumptions for a three sector model. To maintain comparability, all 3 sector models use the same set of coefficients for β_k and γ_{jk}. The first specification draws the error terms in the outcome equations from lognormal rather than normal distributions and sets $\tau_k = 0$ for all k. The semiparametric approach developed in this paper does a good job of eliminating selection bias, at least in the large sample. In contrast, the Lee approach relying on joint normality does not completely eliminate the bias. Specification (8) increases the variance of the error distribution so that the bias is approximately twice as much as in (7). Once again, the semiparametric estimator relying on the first-best probability outperforms the Lee estimator relying on joint normality. The Lee approach could be improved by using other distributions for the transformations discussed in Appendix A, but the researcher would somehow need to choose the appropriate distributions (see Lee (1982)).

The point of specifications (7) and (8) is that selectivity corrections based on a parametric approach can be significantly biased when the distributions are misspecified (Arabmazar and Schmidt (1981, 1982), Goldberger (1983)). Another example when the Lee approach performs poorly occurs when u_{ik} is drawn from a mixture of normal distributions. In a simulation not reported in the table, the mean OLS estimate was 1.318, the Lee estimate was 1.045, and the first best probability estimate was 1.008.

Specification (9) models the variance in the fixed effect component of unobserved earnings as a function of s_i. This provides one way to model the observation made by labor economists that the variance in earnings is larger for more educated individuals. Both the Lee approach and the semiparametric approach developed in this paper work well in the presence of this conditional heteroskedasticity.

Specification (10) draws the error term w_{ijk} from a multivariate normal distribution, with correlation in this variable across the j originating sectors. This implies the correction functions are sector-specific, and suggests that using a single correction function may be inappropriate. If a single correction function is used, even though sector-specific correction functions are called for, the estimation method of this paper still eliminates a large fraction of the bias. However, only by adding separate control functions for each sector does the bias go to zero.

APPENDIX TABLE

MONTE CARLO RESULTS
(True Parameter Value Equals 1)

	10,000 Observations per Sector				1,000 Observations per Sector			
	Mean	Std. Dev.	RMSE	Ave. Sample Size	Mean	Std. Dev.	RMSE	Ave. Sample Size
BASELINE MODELS OF VARIOUS DIMENSIONS								
(1) *2 Sectors*								
OLS	1.074	0.014	0.075	5,579	1.074	0.044	0.086	558
Lee	1.003	0.016	0.016		1.022	0.049	0.054	
1st Best Probability	1.005	0.016	0.017		1.022	0.049	0.054	
(2) *3 Sectors*								
OLS	1.139	0.019	0.140	4,314	1.138	0.064	0.152	432
Lee	0.987	0.025	0.028		1.037	0.079	0.087	
1st Best Probability	0.995	0.027	0.027		1.044	0.082	0.093	
+Remaining Probability	1.026	0.037	0.032		1.068	0.103	0.123	
(3) *5 Sectors*								
OLS	1.084	0.022	0.087	3,271	1.080	0.068	0.106	326
Lee	1.003	0.030	0.030		1.060	0.077	0.097	
1st Best Probability	1.002	0.030	0.030		1.059	0.078	0.098	
(4) *10 Sectors*								
OLS	1.098	0.016	0.099	5,604	1.104	0.050	0.115	562
Lee	0.992	0.019	0.021		1.051	0.054	0.074	
1st Best Probability	0.999	0.020	0.020		1.053	0.055	0.076	
(5) *25 Sectors*								
OLS	1.127	0.026	0.130	3,259	1.124	0.072	0.143	326
Lee	1.023	0.031	0.038		1.096	0.075	0.122	
1st Best Probability	1.026	0.031	0.041		1.096	0.076	0.122	
(6) *50 Sectors*								
OLS	1.139	0.030	0.142	2,612	1.135	0.100	0.168	263
Lee	1.020	0.032	0.037		1.097	0.102	0.141	
1st Best Probability	1.028	0.032	0.042		1.098	0.104	0.143	
DEVIATIONS FROM BASELINE IN A THREE SECTOR MODEL								
(7) *Lognormal distribution for* u_{ik} $[\log(u_{ik}) \sim N(0, 2.7)]$								
OLS	1.130	0.016	0.131	3,492	1.131	0.052	0.141	350
Lee	1.023	0.017	0.029		1.053	0.057	0.078	
1st Best Probability (4th order)	1.007	0.017	0.019		1.047	0.059	0.075	
(8) *Lognormal distribution for* u_{ik}*, with approximately twice as much bias as* (7) $[\log(u_{ik}) \sim N(0, 4.4)]$								
OLS	1.250	0.022	0.251	3,693	1.249	0.071	0.259	369
Lee	1.058	0.022	0.062		1.119	0.077	0.141	
1st Best Probability (4th order)	1.016	0.024	0.029		1.104	0.079	0.107	
(9) *Conditional heteroskedasticity for* a_i $[a_i \sim N(0, s_i^2)]$								
OLS	1.140	0.050	0.149	4,215	1.140	0.173	0.222	422
Lee	1.000	0.057	0.057		1.047	0.196	0.202	
1st Best Probability	1.005	0.063	0.064		1.054	0.205	0.212	

APPENDIX TABLE—*Continued*

	10,000 Observations per Sector				1,000 Observations per Sector			
	Mean	Std. Dev.	RMSE	Ave. Sample Size	Mean	Std. Dev.	RMSE	Ave. Sample Size
(10) *Correlation across j for* w_{ijk}								
OLS	1.131	0.019	0.133	6,030	1.127	0.060	0.140	603
Lee	1.041	0.020	0.045		1.059	0.065	0.088	
1st Best Probability	1.040	0.021	0.045		1.061	0.066	0.090	
Sector-Specific Control Functions	1.000	0.031	0.031		1.105	0.077	0.130	
(11) *Fixed effect loading factor* τ_k *equals* β_k								
OLS	1.121	0.020	0.123	6,211	1.118	0.066	0.135	622
Lee	0.954	0.029	0.054		1.036	0.078	0.085	
1st Best Probability	0.960	0.031	0.051		1.041	0.079	0.089	
1st Best + Additional Probability	1.002	0.047	0.047		1.045	0.099	0.109	
DEVIATION FROM BASELINE IN A FIFTY SECTOR MODEL								
(12) *Random Correlation across k for* b_{ik} [*i.e.,* u_{ik} *correlated across states*]								
OLS	1.204	0.024	0.205	2,523	1.207	0.078	0.221	250
Lee	1.107	0.026	0.110		1.177	0.078	0.193	
1st Best Probability	1.108	0.026	0.111		1.182	0.081	0.199	
+Maximum Predicted Probability	1.154	0.028	0.157		1.211	0.087	0.228	
+One Other Probability	1.109	0.027	0.113		1.183	0.081	0.200	
+Two Other Probabilities[a]	1.111	0.029	0.115		—	—	—	
+Three Other Probabilities[a]	1.124	0.030	0.128		—	—	—	

Notes: 500 replications for all specifications. "Lee" stands for the method relying on the normality assumptions as developed in Appendix A and "1st Best Probability" stands for the semiparametric approach developed in the current paper using assumption (A-2). RMSE stands for root mean squared error. See Appendix C for further details on the Monte Carlo designs and a discussion of the results.

[a]Not estimable for the small sample, since the covariance matrix of independent variables was often singular.

How does the estimator perform when the index sufficiency assumption does not hold? In specification (11), the loading factor on the fixed effect is allowed to vary by state, so that $\tau_k = \beta_k$. In the context of a model of earnings, this specification can be interpreted as setting the return to unobserved ability equal to the return to education in a state. As discussed in Section 3.3, this formulation violates the index sufficiency assumption. Under these circumstances, the estimation method of this paper does not eliminate the bias, although the estimate has a smaller bias compared to the OLS estimate. If the remaining probability is included in the control function, then no index sufficiency assumption is needed. When this additional probability is included in a second-order polynomial expansion of the correction function, the bias is eliminated at the cost of a small increase in the variance of the estimate.

The final specification examines the assumption of index sufficiency in a high-dimensional model. Fifty error terms for the fifty outcome equations are drawn from a multivariate normal distribution, with random covariances across sectors. Although a slightly different formulation compared to specification (11), this error term structure also violates the index sufficiency assumption. The Lee estimator and the estimator that assumes single index sufficiency (A-2) each eliminate about half of the bias. The curse of dimensionality makes it impossible to include all 50 probabilities in the correction function, so a few other probabilities were added instead. Including other probabilities, whether it be the maximum predicted probability or other probabilities chosen at random, does not reduce

the bias. These results suggest that it may be difficult to detect violations of the index sufficiency assumption in high-dimensional models (see Section 3.3).

APPENDIX D: DATA SOURCES FOR TABLE V

Distance and Unemployment Variables

(1) *Distance in Miles:* Author's calculations of distance between the state capitals in miles, using the "Great Circle" formula:

$$Distance = \arccos\{[\sin(Lat_j)\sin(Lat_k)] + [\cos(Lat_j)\cos(Lat_k)\cos(Lon_j - Lon_k)]\} \times M$$

where Lat_j = latitude of capital j, Lat_k = latitude of capital k, Lon_j = longitude of capital j, Lon_k = longitude of capital k, and $M = 69.16$ miles, the average value of a degree. Source for latitude and longitude of state capitals: Munro, D., editor (1988): *Cambridge World Gazetteer: A Geographical Dictionary*. Cambridge, U.K.: Cambridge University Press.

(2) *Unemployment Rate:* Five year average unemployment rate for individuals age 16 and older. Source for unemployment rate: U.S. Bureau of the Census (various years): *March Current Population Survey*.

Quality of Life Variables

(3) *Population Density:* Persons per square mile.
(4) *Doctors per Capita:* Doctors per person.
(5) *Dentists per Capita:* Dentists per person.
(6) *Hospital Costs:* Average hospital cost per day (to hospital), in dollars.
(7) *Teachers' Salaries:* Average salary for a teacher in the public schools, in dollars.
(8) *School Expenditures per Capita:* Annual public school expenditures, per capita, in dollars.
(9) *School Expenditures per Pupil:* Annual public school expenditures, per pupil, in dollars.
(10) *Crime Rate:* Crime rate per 100,000 persons.
(11) *Violent Crime Rate:* Violent crime rate per 100,000 persons.
(12) *Incarceration Rate:* Number of prisoners sentenced to more than one year, per 100,000 residents.

Source for variables (3)–(12): Horner, E., editor (various years): *Almanac of the 50 States: Basic Data Profiles with Comparative Tables*. Palo Alto: Information Publications.

Climate Variables

(13) *Average Temperature:* Normal daily mean temperature in Fahrenheit degrees.
(14) *Maximum Temperature:* Normal daily maximum temperature in Fahrenheit degrees.
(15) *Minimum Temperature:* Normal daily minimum temperature in Fahrenheit degrees.
(16) *Annual Precipitation:* Normal annual precipitation in inches.

Source for variables (13)–(16): U.S. National Oceanic and Atmospheric Administration, *Climatography of the United States*, No. 81. Airport data based on a standard thirty-year period, 1961–1990. Data for selected cities in all 50 states and the District of Columbia.

(17) *Number of Rainy Days:* Average days per year with precipitation of 0.01 inch or more.
(18) *Number of Sunny Days:* Average days per year that are either clear or partly cloudy.
(19) *Afternoon Humidity:* Annual average relative afternoon humidity (percent).
(20) *Average Wind Speed:* Annual average wind speed in miles per hour.

Source for variables (17)–(20): U.S. National Oceanic and Atmospheric Administration, *Comparative Climatic Data*, Annual. Airport data for period of record through 1993. Data for selected cities in all 50 states and the District of Columbia.

State Spending and Taxing Variables

(21) *State Spending on Education:* Annual per capita state expenditures on public education.

(22) *State Spending on Health/Human Services:* Annual per capita state expenditures on health and human services.

(23) *State Spending on Highways:* Annual per capita state expenditures on highways.

(24) *State Spending on Public Welfare:* Annual per capita state expenditures on public welfare.

(25) *Miscellaneous State Spending:* Total annual per capita state expenditures, minus expenditures on education, health/human services, highways, and public welfare.

(26) *State Sales Tax:* Total annual sales tax revenue per capita.

(27) *Average State Income Tax:* Average annual income tax revenue per capita.

Source for variables (21)–(27): U.S. Bureau of the Census, Government Finance Series, *General Revenue Tables*. No data for Hawaii, Alaska, or the District of Columbia. All variables measured in dollars.

REFERENCES

AHN, H., AND J. L. POWELL (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3–29.

ANDREWS, D. W. K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59, 307–345.

ANDREWS, D. W. K., AND Y. WHANG (1990): "Additive and Interactive Regression Models: Circumvention of the Curse of Dimensionality," *Econometric Theory*, 6, 466–479.

ARABMAZAR, A., AND P. SCHMIDT (1981): "Further Evidence on the Robustness of the Tobit Estimator to Heteroskedasticity," *Journal of Econometrics*, 17, 253–258.

——— (1982): "An Investigation of the Robustness of the Tobit Estimator to Non-Normality," *Econometrica*, 50, 1055–1063.

ASHENFELTER, O., AND D. CARD (1985): "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *Review of Economics and Statistics*, 67, 648–660.

BORJAS, G. J. (1987): "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77, 531–553.

BORJAS, G. J., S. G. BRONARS, AND S. J. TREJO (1992): "Self-selection and Internal Migration in the United States," *Journal of Urban Economics*, 32, 159–185.

BOUND, J., AND G. JOHNSON (1992): "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations," *American Economic Review*, 82, 371–392.

CARD, D., AND A. B. KRUEGER (1992): "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, 100, 1–40.

CARD, D., AND A. PAYNE (1998): "School Finance Reform, The Distribution of School Spending, and the Distribution of SAT Scores," National Bureau of Economic Research Working Paper 6766.

CARD, D., AND D. SULLIVAN (1988): "Measuring the Effect of Subsidized Training Programs on Movements in and out of Employment," *Econometrica*, 56, 497–530.

CHISWICK, B. R. (1974): *Income Inequality: Regional Analyses within a Human Capital Framework*. New York: National Bureau of Economic Research.

CHOI, K. (1992): "Identification and Estimation of Nonparametric and Semiparametric Sample Selection Models," Ph.D. Dissertation, University of Chicago.

COSSLETT, S. R. (1991): "Semiparametric Estimation of a Regression Model with Sample Selectivity," in *Semiparametric Methods in Economics and Statistics*, ed. by W. A. Barnett, J. Powell, and G. E. Tauchen. Cambridge, U.K.: Cambridge University Press.

DOLTON, P. J., G. H. MAKEPEACE, AND W. VAN DER KLAAUW (1989): "Occupational Choice and Earnings Determination: The Role of Sample Selection and Non-Pecuniary Factors," *Oxford Economic Papers*, 41, 573–594.

DUBIN, J., AND D. MCFADDEN (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, 345–362.

DYNARSKI, M. (1987): "The Scholastic Aptitude Test: Participation and Performance," *Economics of Education Review*, 6, 263–273.

FALARIS, E. M. (1987): "A Nested Logit Migration Model with Selectivity," *International Economic Review*, 28, 429–443.

GOLDBERGER, A. (1983): "Abnormal Selection Bias," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, ed. by S. Karlin, T. Amemiya, and L. Goodman. New York: Academic Press.

GRONAU, R. (1974): "Wage Comparisons, a Selectivity Bias," *Journal of Political Economy*, 82, 1119–1144.

HAM, J. C., AND C. HSIAO (1984): "Two-Stage Estimation of Structural Labor Supply Parameters Using Interval Data from the 1971 Canadian Census," *Journal of Econometrics*, 24, 133–158.

HAM, J. C., AND R. J. LALONDE (1996): "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica*, 64, 175–205.

HAM, J. C., X. LI, AND P. B. REAGAN (2001): "Selection and Matching Estimates of the Effect of Migration on the Wages of Young Men," Unpublished Manuscript, Ohio State University.

HAUSMAN, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 47, 1251–1272.

HAY, J. (1980): "Occupational Choice and Occupational Earnings: Selectivity Bias in a Simultaneous Logit-OLS Model," Ph.D. Dissertation, Yale University.

HECKMAN, J. J. (1974): "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42, 679–693.

——— (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

——— (1981): "Addendum to Sample Selection Bias as a Specification Error," *Evaluation Studies Review Annual*, 5, 69–74.

HECKMAN, J. J., AND B. E. HONORÉ (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121—1149.

HECKMAN, J. J., AND R. ROBB (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. Heckman and B. Singer. Cambridge, U.K.: Cambridge University Press.

——— (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in *Drawing Inferences from Self-Selected Samples*, ed. by H. Wainer. New York: Springer-Verlag.

HECKMAN, J. J., AND G. SEDLACEK (1990): "Self-Selection and the Distribution of Hourly Wages," *Journal of Labor Economics*, 8, S329–S363.

ICHIMURA, H. (1987): "Consistent Estimation of Index Model Coefficients," Ph.D. Dissertation, Massachusetts Institute of Technology.

ICHIMURA, H., AND L. LEE (1991): "Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation," in *Semiparametric Methods in Economics and Statistics*, ed. by W. A. Barnett, J. Powell, and G. E. Tauchen. Cambridge, U.K.: Cambridge University Press.

KATZ, L. F., AND K. M. MURPHY (1992): "Changes in Relative Wages, 1963–1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107, 35–78.

KLEIN, R. W., AND R. H. SPADY (1993): "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 61, 387–421.

LEE, L. (1978): "Unionism and Relative Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," *International Economic Review*, 19, 415–433.

——— (1982): "Some Approaches to the Correction of Selectivity Bias," *Review of Economic Studies*, 49, 355–372.

——— (1983): "Generalized Econometric Models with Selectivity," *Econometrica*, 51, 507–512.

——— (1995): "Semiparametric Maximum Likelihood Estimation of Polychotomous and Sequential Choice Models," *Journal of Econometrics*, 65, 381–428.

MADDALA, G. S. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, U.K.: Cambridge University Press.

MANSKI, C. F. (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 205–228.

MATZKIN, R. L. (1993): "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 58, 137–168.

McFadden, D. L. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press.

——— (1984): "Econometric Analysis of Qualitative Response Models," in *Handbook of Econometrics, Vol. II*, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North Holland.

Mroz, T. A. (1987): "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, 55, 765–799.

Murphy, K. M., and R. H. Topel (1985): "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics*, 3, 370–379.

Nakosteen, R. A., and M. Zimmer (1980): "Migration and Income: The Question of Self-Selection," *Southern Economic Journal*, 46, 840–851.

Newey, W. K. (1988): "Two Step Series Estimation of Sample Selection Models," Unpublished Manuscript, Princeton University.

——— (1994a): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

——— (1994b): "Series Estimation of Regression Functionals," *Econometric Theory*, 10, 1–28.

——— (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168.

Newey, W. K., J. L. Powell, and J. R. Walker (1990): "Semiparametric Estimation of Selection Models: Some Empirical Results," *American Economic Review*, 80, 324–328.

Powell, J. L. (1987): "Semiparametric Estimation of Bivariate Latent Variable Models," Social Systems Research Institute Working Paper 8704, University of Wisconsin.

——— (1998): "Estimation of Semiparametric Models," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle and D. McFadden. Amsterdam: North Holland.

Roback, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90, 1257–1278.

——— (1988): "Wages, Rents, and Amenities: Differences among Workers and Regions," *Economic Inquiry*, 26, 23–41.

Robinson, C., and N. Tomes (1982): "Self-selection and Interprovincial Migration in Canada," *Canadian Journal of Economics*, 15, 474–502.

Rosen, S. (1983): "A Note on Aggregation of Skills and Labor Quality," *Journal of Human Resources*, 18, 425–431.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

Ruggles, S., and M. Sobek (1997): Integrated Public Use Microdata Series: Version 2.0, http://www.ipums.umn.edu. Minneapolis: Historical Census Projects, University of Minnesota.

Schmertmann, C. P. (1994): "Selectivity Bias Correction Methods in Polychotomous Sample Selection Models," *Journal of Econometrics*, 60, 101–132.

Sjaastad, L. A. (1962): "The Costs and Returns of Human Migration," *Journal of Political Economy*, 70, 80–93.

Trost, R. P., and L. Lee (1984): "Technical Training and Earnings: A Polychotomous Choice Model with Selectivity," *Review of Economics and Statistics*, 66, 151–156.

Vella, F. (1998): "Estimating Models with Sample Selection Bias: A Survey," *Journal of Human Resources*, 33, 127–172.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

Willis, R. J., and S. Rosen (1979): "Education and Self-Selection," *Journal of Political Economy*, 87, S7–S36.