Nonparametric Identification and Estimation in a Roy Model with Common Non-pecuniary Returns

Patrick Bayer
Duke University and NBER

Shakeeb Khan
Duke University

Christopher Timmins
Duke University and NBER

August 7, 2009

Abstract

This paper considers identification and estimation of a Roy model that includes, along with the usual sorting over wage draws, a common non-pecuniary component of utility associated with each choice alternative. The original Roy model (1951) was used as a tool to describe occupational sorting based on only wages, but it has broader applications to many polychotomous choice problems if a non-pecuniary component of utility is included. In this paper, we develop nonparametric estimators for such a model corresponding to two alternative assumptions under which we prove identification, derive asymptotic properties, and illustrate small sample properties with a series of Monte Carlo experiments. These estimators provide an alternative to arguments based on "identification at infinity". We apply one of those models to migration behavior and analyze the effect of Roy sorting on observed wage distributions in an application based on Dahl (2002). In particular, micro data from the 2000 Census are used to calculate the returns to a college education. If high-school and college graduates face different costs of migration, this would be reflected in different degrees of Roy-sorting-induced bias in their observed wage distributions. Correcting for this bias, the observed returns to a college degree are cut in half.

We would like to thank Richard Blundell, James Heckman, Hide Ichimura, Robert Moffitt, Tiemen Woutersen, and seminar participants at SUNY Albany, Duke University, Johns Hopkins University and the University of Chicago for their helpful comments. All remaining errors and omissions are our own.
Correspondence should be directed to all three authors at the Department of Economics, Duke University, Durham, NC 27708. (919) 660-1800. patrick.bayer@duke.edu, shakeebk@econ.duke.edu, christopher.timmins@duke.edu.

1. Introduction

In the original application of his model, Roy (1951) showed that the self-selection of individuals into occupations generally implies that observed wages (conditional on occupation choice) differ markedly from the underlying distribution of wages in the population. The Roy model has subsequently been applied to a wide class of problems in economics as its structure fits any setting in which individuals choose among a set of alternatives to maximize an outcome associated with that choice. Given its wide applicability, an important line of recent research has analyzed identification in the Roy model. Beginning with Heckman and Honore (1990), this literature has produced a series of results that clarify the conditions under which the underlying population distribution of wages can (or cannot) be identified in observational data.

In this paper, we study the identification and estimation of a Roy model that includes, in addition to the usual sorting with respect to wage draws, a common non-pecuniary component of utility associated with each alternative. An important limitation of the pure Roy model is that it assumes that individuals maximize only economic returns (e.g., wages). Yet non-pecuniary aspects of decisions are important in many economic applications. In the choice of occupation, for example, non-pecuniary components of utility would include the amenity value or injury risk associated with different jobs.[1] As with the pure Roy model, this generalized version is also applicable to settings in which the outcome of interest is not economic returns. In studying the choice of health behaviors or medical treatments, for example, the relevant outcome might be the survival rate, while the "non-pecuniary" component of utility might capture the enjoyment associated with a behavior (such as smoking) or disutility of side-effects associated with various treatments.[2] In this way, the generalized model developed here can be applied to a wide class of problems in economics.

In the analysis that follows, we provide two distinct sets of conditions under which the non-pecuniary values associated with each choice alternative are nonparametrically identified. As in Heckman and Honore (1990), the identification of the full population wage distribution requires additional identifying assumptions; the key insight of our paper is that the non-pecuniary value of alternatives in a generalized Roy model can be identified even in a single cross-section. The emphasis here is the identification of a particular parameter related to non-pecuniary returns.

[1]In modeling the choice of labor market/residence, the non-pecuniary component of utility would capture variation in amenities and cost-of-living across cities.

[2] Likewise, in the study of school choice, the relevant outcome might be achievement scores, while other factors affecting the choice of school (e.g., availability of special education programs) might be included as part of a separate component of utility.

Thus our identification comes at the expense of imposing more structure- i.e. a sector-specific constant characterizing non-pecuniary returns.

Throughout the paper, we consider the nonparametric identification of the model in a relatively demanding setting in which (i) the set of available choices is large and (ii) identification at infinity is difficult to undertake.[3] The objects to be identified are the population wage distributions and a common non-pecuniary utility associated with each choice. In developing a first set of conditions for identification, we impose only the relatively innocuous requirement that the distribution of pecuniary returns has a finite lower bound, whose value is independent of returns (both pecuniary and non-pecuniary) associated with other choices. Given this assumption, we demonstrate that the difference in the minimum order statistic for any two alternatives exactly identifies the difference in the non-pecuniary value of those choices. Intuitively, this follows directly from the observation that no individual will choose a less-preferred choice (on the basis of non-pecuniary considerations) unless the wage offered there exceeds this threshold. Thus, the minimum wage observed in the less-preferred sector should be exactly the minimum wage observed in the more-preferred sector plus the difference in non-pecuniary components.[4] Having identified the non-pecuniary component of utility, we show that it is then straightforward to (i) back-out underlying unconditional population wage distributions using transformed versions of the observed conditional wages distributions for each sector and the Kaplan-Meier (1958) procedure, if one assumes independence, or (ii) apply Petersen (1976) bounds to the transformed data to bound the unconditional population wage distribution.

While this estimator works very well in certain data environments, relying exclusively on differences in minimum order statistics to identify the non-pecuniary component of utility raises concerns about measurement error. As a result, we consider a second set of formal identifying assumptions. Our second identification proof is based on two key assumptions. First, we assume independence.[5] Second, we assume that information is available for (at least) two subsets of the

3 Identification of this sort of problem has been studied extensively in the literature, especially in the binomial choice problem. [See, for example, Heckman and Honore (1989) and Heckman (1990)] As there is a close link between Roy models and competing risk models, several of the papers in the survey in Powell(1994) are also related to the models we explore here. Some more recent related work includes Honore et al.(2002), Lee(2006), Honore and Lleras-Muney(2007), and Khan and Tamer(2007). While the identification at infinity strategy can be extended to the multinomial choice setting, the requisite demands on the data are enormous, requiring, for example, the availability of distinct combinations of covariates that compel individuals to select each choice with certainty.

4 Note that, within the pure Roy model, the minimum order statistics would be identical for all choices and the full empirical content of the data would in fact be absorbed by a specification including independent population wage distributions, as suggested by Heckman and Honore (1990).

5 Again, following the existing literature, the independence assumption can be relaxed in more generous data environments (e.g., when data is available for more than a single cross-section or when covariates of

population that differ in their non-pecuniary valuation of the set of choice alternatives. In the application that we present below, we consider the choice of regional labor market; in that context, moving costs (broadly defined) naturally imply that birth region affects the non-pecuniary value one ascribes to a particular destination. We then exploit the fact that wage offers are likely to be similar for individuals with similar characteristics from neighboring regions while the non-pecuniary value of residing in these regions will vary significantly with an individual's birthplace. We refer to this second assumption as "commonality", i.e., that a common wage distribution characterizes wage offers for all individuals regardless of birthplace. In one sense, this commonality assumption can be interpreted as an exclusion assumption – i.e., birth location can be regarded as an excluded variable that affects the non-pecuniary component of utility but does not affect wages. Moreover, this exclusion restriction is easy to employ in our empirical context as long as there are more than two locations. Given this assumption, we prove that both the non-pecuniary components of utility for each population subset and the overall population wage distributions are identified.

In this case, some intuition for why the model is identified by the commonality assumption can again be gained by referring back to Heckman and Honore (1990). Without non-pecuniary components of utility, the observed conditional wage distributions and choice probabilities map uniquely to a set of independent population wage distributions. With at least two subsets of the population that differ in their non-pecuniary valuations of alternatives, however, the resulting unconditional wage distributions that would reconcile the two subsets of the data would differ. What our identification proof ensures is that the identical unconditional wage distributions for each subset can only be reconciled at the true values of the non-pecuniary components of utility for each population subset.

Estimation of this model follows directly from the identification proof. As we show below, it is possible to write a system of equations based on the observed conditional wage distributions that must equal zero identically at the true values of the non-pecuniary parameters for each population subset. These equations serve as natural moments for a minimum distance estimator.

These results add to a sparse literature that has studied the nonparametric identification of a Roy model with many alternatives and non-pecuniary components of utility. Dahl (2002) proposes a multinomial version of the estimator developed in the binomial context by Ahn and

---

the sort used in the prior literature are available). See, for example, Khan and Tamer (2007) who achieve identification results under strong support conditions in a semi-parametric Roy model. Honore and Llera-Muney (2007) establish set identification when the independence assumption is relaxed.

Powell (1994). His extension relies on the key assumption that a non-parametric selection correction term can be based on the first-best choice probability. This assumption is not, however, based on a model of utility maximizing behavior. Other work has examined spatial sorting behavior based on wages and non-pecuniary benefits. Falaris (1987) and Davies, Greenwood, and Li (2001) study the determinants of migration decisions in Venezuela and the US, respectively. Falaris applies Lee's (1983) generalized polychotomous choice model to control for non-random selection bias in conditional wage distributions, while Davies, Greenwood, and Li essentially ignore it. The entire literature on wage-hedonics, beginning with Roback (1982), has similarly ignored this problem. In those papers, wage and housing price gradients across cities are used to back-out the value of urban amenities. Wage distributions conditional on non-random selection into cities are typically used to calculate the first of these gradients, leading to biased estimates.

We conclude this paper by applying our estimator to US Census data to study the effect of spatial sorting on returns to a college education, addressing the same question as Dahl (2002). College graduates are more likely to migrate than are high-school graduates, meaning that the bias in their conditional wage distributions induced by Roy sorting will be greater. Controlling for this bias for both high-school and college graduates, we find that the estimated returns to a college education at the median fall from 42% to only 18%.

The remainder of the paper is organized as follows. Section 2 introduces the Roy model, proves identification for the case in which wage distributions are assumed to have a finite lower support that is independent of all wage distributions, and develops a corresponding estimator. Section 3 proves identification under the alternative assumptions of independence and commonality, and develops a corresponding estimator. Section 4 outlines the asymptotic properties of our estimators, and section 5 shows how each estimator performs in finite samples and under less-than-ideal data circumstances. Section 6 uses the second estimator to recover an unbiased estimate of the returns to a college education. Section 7 concludes with a discussion of possible extensions to this research.

2. Identification and Estimation – Finite Lower Support

We begin our analysis by describing the Roy model with pecuniary and non-pecuniary returns and data environment that we study. We then prove identification under two separate sets of assumptions. The first case is characterized by (i) the assumption that the distribution of the endogenously determined payoffs (e.g., wages in a classic Roy model) has a finite lower support, the value of which is independent of (exogenous) wage distributions, and (ii) a weaker version of the typical independence assumption used in Roy models. We refer to this assumption as extreme quantile independence and describe it in more detail below.

Our second set of assumptions, which relies on independent wage draws, is applicable in situations where a finite lower support cannot be assumed, or where the minimum order statistic provides a noisy measure of the lower bound. That estimator is described in detail in Section 3. In both cases, we first prove identification with a simple model describing the sorting of individuals from a single origin location into one of two destinations (k = 1, 2). We indicate the wage earned by individual i, should he choose to settle in locations #1 and #2 as $\omega_{1,i}$ and $\omega_{2,i}$, respectively. In contrast to the classic Roy model, where sorting is simply across employment sectors and driven entirely by pecuniary compensation, we model sorting in a geographic context where the individual's location decision depends in part on his wage draw in each location, but also on non-wage determinants of utility specific to a particular location, which we label as "tastes".[6] Utility from choosing to settle in location k is given by the sum of wages ($\omega_{k,i}$) and tastes (τ_k):

$$(1) \qquad U_{k,i} = \omega_{k,i} + \tau_k$$

Without loss of generality, we normalize $\tau_1 = 0$.[7] The goal of our exercise is to recover estimates of τ_2, $f_1(\omega_1)$, and $f_2(\omega_2)$ (i.e., the taste parameter associated with location #2 and the unconditional wage distributions in each location). The difficulty arises from the fact that we only see (i) wage distributions conditional upon optimal sorting behavior, and (ii) an indicator of which location an individual chooses.

Note that our model is not as flexible as other generalizations of the Roy model found in the literature (e.g., Heckman and Vytlacil, (2006)) in that we do not allow τ_k to differ across individuals. This is done in order to support our identification strategy – a strategy that allows us to recover the values of taste parameters that themselves may be of interest in many empirical applications (see, for example, Deleire, Khan, and Timmins (2009)).

---

[6] Tastes would certainly include natural amenities and local public goods associated with the destination location. In addition, they may include "migration costs"; i.e., costs specific to someone moving from a particular origin to a particular destination. In a narrow sense, these costs would be comprised of re-location expenditures. In broader terms, these costs would likely involve the psychological costs of living far from one's birth location. 2000 Census data indicate that a majority of US household heads live in the narrowly defined region in which they were born. [Bayer, Keohane, and Timmins (2007)]

[7] As in all random-utility frameworks, utility is only identified up to an additive constant. This requires some sort of a normalization, which we use to eliminate one of the τ's from the two-destination example. In the more general N x N case, we estimate (N-1) τ's for each of the N origins.

2.1 Identification of Tastes Based on a Finite Lower Support

Our first approach uses only the conditional wage distributions and an indicator of location choice to recover τ_2, $f_1(\omega_1)$, and $f_2(\omega_2)$ according to the following argument based on minimum order statistics. For an individual i, we only observe $\omega_{2,i}$ if:

(2) $$\omega_{2,i} + \tau_2 \geq \omega_{1,i}$$

and we only observe $\omega_{1,i}$ if:

(3) $$\omega_{2,i} + \tau_2 < \omega_{1,i}$$

Note equations (2) and (3) can be regarded as an example of a generalized Roy model-(see Heckman and Vytlacil, *Handbook of Econometrics*, Chapter 70) for a thorough discussion on various generalizations) where here our mechanism for sector selection includes an additive taste parameter that is constant for individuals in a given sector/region. Denote the smallest wage (i.e., the minimum order statistic) that we observe from someone choosing to settle in location #1 by $\underline{w}_1$. $\underline{w}_2$ is similarly defined. We assume that $f_1(\omega_1)$ and $f_2(\omega_2)$ have finite lower points of supports (denoted by ω_1^* and ω_2^*, respectively) that are independent of the values of $\underline{w}_1$ and $\underline{w}_2$. We refer to this as an "extreme quantile independence" assumption, analogous to assumption sometimes invoked in the semiparametric literature, where one particular quantile of a random variable is independent of another random variable, but all the other quantiles are allowed to vary. In that sense we can see that our assumption is weaker than full independence assumption. We know that the smallest value of ω_1 that we could ever see given that individuals maximize utility:

(4) $$\begin{array}{lll} \underline{w}_1 = \omega_1^* & if & \omega_1^* > \omega_2^* + \tau_2 \\ \underline{w}_1 = \omega_2^* + \tau_2 & if & \omega_1^* \leq \omega_2^* + \tau_2 \end{array}$$

Similarly, the smallest value of ω_2 that we could ever see would be:

(5) $$\begin{array}{lll} \underline{w}_2 = \omega_2^* & if & \omega_1^* \leq \omega_2^* + \tau_2 \\ \underline{w}_2 = \omega_1^* - \tau_2 & if & \omega_1^* > \omega_2^* + \tau_2 \end{array}$$

In order to make sense of (4) and (5), define the following two cases:

$$\text{(6)} \qquad \begin{aligned} A: &\quad \omega_1^* > \omega_2^* + \tau_2 \\ B: &\quad \omega_1^* \leq \omega_2^* + \tau_2 \end{aligned}$$

We are not able to tell whether case A or B prevails in the data without recovering an estimate of τ_2. Conveniently, we are able to recover an estimate of τ_2 in either case. In particular:

$$\text{(7)} \qquad \tau_2 = \underline{w}_1 - \underline{w}_2$$

Equation (7) therefore describes our first estimator of τ_2 in the simplest 1 x 2 case. The same logic extends easily to any number of potential destinations (i.e., $\tau_k = \underline{w}_1 - \underline{w}_k$, $k = 1, 2, ..., K$). Begin by defining the following indicator variables:

$$\text{(8)} \qquad d_{1,i} \;=\; I\left[\omega_{1,i} > \max\left(\omega_{2,i} + \tau_2, \omega_{3,i} + \tau_3, ..., \omega_{K,i} + \tau_K\right)\right]$$

$$\text{(9)} \qquad d_{j,i} \;=\; I\left[\omega_{j,i} + \tau_j > \max\left(\omega_{1,i}, ..., \omega_{j-1,i} + \tau_{j-1}, \omega_{j+1,i} + \tau_{j+1}, ..., \omega_{K,i} + \tau_K\right)\right]$$

where we continue to normalize $\tau_1 = 0$. The observed wage of individual i is defined by:

$$\text{(10)} \qquad w_i = \sum_{k=1}^{K} \omega_{k,i} d_{k,i}$$

We next define the following indicator variables, which refer to the finite lower bounds in each location:

$$\text{(11)} \qquad \delta_1 \;=\; I\left[\omega_1^* > \max(\omega_2^* + \tau_2, \omega_3^* + \tau_3, ..., \; \omega_K^* + \tau_K)\right]$$

$$\text{(12)} \qquad \delta_k \;=\; I\left[\omega_k^* + \tau_k > \max(\omega_1^*, ..., \omega_{k-1}^* + \tau_{k-1}, \omega_{k+1}^* + \tau_{k+1}, ..., \; \omega_K^* + \tau_K)\right]$$

We then proceed by evaluating the minimum order statistic for an individual choosing to settle in location #1:

$$(13) \qquad \underline{w}_1 = \min(w_i \mid d_{1,i} = 1) = \omega_1^* \delta_1 + \sum_{k=2}^{K} (\omega_k^* + \tau_k)\, \delta_k$$

and in each location $j > 1$:

$$(14) \qquad \underline{w}_j = \min(w_i \mid d_{1,i} = j) = \omega_j^* \delta_j + \sum_{\substack{k=1 \\ k \neq j}}^{K} (\omega_k^* + \tau_k - \tau_j)\, \delta_k$$

By simple inspection, one can see that $\tau_k = \underline{w}_1 - \underline{w}_k, \ \forall\, k = 1, 2, ..., K$.

2.2 Identification of $f_1(\omega_1)$ and $f_2(\omega_2)$ with Kaplan-Meier

Having recovered an estimate of τ_2, it is a simple matter to recover $f_1(\omega_1)$ and $f_2(\omega_2)$ by employing a variation of the Kaplan-Meier (1958) procedure typically used in competing-risks models under independence assumptions.[8] The Kaplan-Meier procedure can be interpreted as a nonparametric maximum likelihood estimator of a censored distribution, and has been proven be asymptotically normally distributed- see, e.g. Gill(1980).

Our variation will be to apply the Kaplan Meier procedure to draws from $\omega_{1,i}+\tau_1$, where τ_1 can be estimated using the proposed procedure. In particular, we estimate $f_1(\omega_1)$ by first creating a new data vector which corresponds to only those values of utilities (i.e., $\omega_{1,i} + \tau_1$) that are "uncensored" for destination #1 (i.e., observed for individuals who optimally chose destination #1). Note that, because we were able to recover tastes with equation (7), we can treat utility (i.e., the sum of wages and tastes) as observed for the remainder of the exercise – our only goal is to recover its unconditional distribution, from which we can recover the unconditional distribution of wages. This vector of utilities will be of smaller dimension than the vector of all utilities, which includes draws for individuals who chose destination #1 or destination #2.

To implement the Kaplan Meier procedure, we can simply use standard software packages such as Stata. In the final step, we simply deduct our estimate of τ_2 from utility $U_{2,i}$ at each point in the support of its distribution. The resulting distribution is a non-parametric representation of $f_2(\omega_2)$. We then repeat this process in order to recover $f_1(\omega_1)$, recalling that τ_1 had been normalized to zero.

---

[8] As we mentioned previously, an alternative approach in this stage would be to relax independence and apply the Petersen (1976) bounds to the transformed data to bound the unconditional distributions.

Note that a portion of the unconditional distribution for one of these two locations will necessarily be censored. Suppose we are in case A, where ω_1^* is large relative to ω_2^* + τ_2. We are therefore able to observe the complete distribution $f_1(\omega_1)$, beginning with $\underline{w}_1 = \omega_1^*$. We are, however, unable to observe $f_2(\omega_2)$ to the left of $\underline{w}_2 = \omega_1^* - \tau_2 > \omega_2^*$. While we are unable to determine the shape of the distribution $f_2(\omega_2)$ between ω_2^* and $\underline{w}_2$ in the above case, we are able to bound from above the value of ω_2^* (i.e., the lower point of support for the censored distribution). In particular, knowing that $\omega_1^* = \underline{w}_1$, we know that $\omega_2^* < \underline{w}_1 - \tau_2$. We are unable to determine more about the shape of the distribution $f_2(\omega_2)$ between ω_2^* and $\underline{w}_2$ without resorting to parametric assumptions.

3. Identification and Estimation – Unbounded Support

While clean, transparent, and applicable in certain data environments, there are two practical problems with the technique outlined in Section 2. First, the payoff variable in question may not naturally have a finite lower support (e.g., theory might dictate using the natural log of wages in the utility function). Second, the minimum order statistic can be a very noisy statistic.[9] Unless one has confidence in the estimate of the minimum order statistic, that noise will be translated directly through to the estimates of the taste parameters and, subsequently, on to the Kaplan-Meier estimates.

As an alternative, we propose in this section an estimator that employs data from the full distribution of conditional wages. Importantly, this approach is valid for an unbounded support.[10] With that flexibility, however, comes the need for a new identification assumption. In particular, we begin by showing that, without such an assumption, τ_2, $f_1(\omega_1)$, and $f_2(\omega_2)$ are not identified. This negative proof, however, reveals just how easily identification can be achieved by exploiting the assumption of "commonality" described in Section 3.2.

3.1 Non-Identification in the 1 x 2 Case

We begin with a simple model of individuals sorting over two locations, indexed by 1 and 2. We assume for simplicity that the individuals are from location 1, and we therefore normalize

[9] For example, the bottom 2-3% of wage observations in the US Census data used for our empirical application in Section 7 are implausibly low (i.e., less than 50¢ per hour).
[10] In practice, this means that poorly measured data in the lower tail of the wage distribution will not have a significant impact on the estimation algorithm, whereas it can have severe effects on the minimum order statistic approach.

their taste for staying there to zero ($\tau_1 = 0$). Our interest is in recovering estimates of τ_2, $f_1(\omega_1)$, and $f_2(\omega_2)$.

We define a variable d_i, which functions as an indicator that individual i remained in his origin location:

$$(15) \qquad d_i = I[\omega_{1,i} > \omega_{2,i} + \tau_2]$$

Using this indicator, we can write down an expression for individual i's observed wage:

$$(16) \qquad w_i = d_i \omega_{1,i} + (1 - d_i)\omega_{2,i}$$

i.e., the individual receives his draw from location #1 if it was utility maximizing to stay there. Next, define the following joint probability distributions, both of which are easily observed in the data:

$$(17) \qquad \Psi_1(t) = P(d_i = 1, w_i \le t) \qquad \qquad \Psi_2(t) = P(d_i = 0, w_i \le t)$$

We will also work with the derivatives of these expressions, which we denote by:

$$(18) \qquad \psi_1(t) = \frac{\partial}{\partial t} P(d_i = 1, w_i \le t) \qquad \qquad \psi_2(t) = \frac{\partial}{\partial t} P(d_i = 0, w_i \le t)$$

For this estimator, we need to impose a stronger independence assumption. Rather than assuming only extreme quantile independence, we assume all wage draws are independent. We can then re-write $\Psi_1(t)$ as:

$$\Psi_1(t) = P(d_i = 1, w_i \le t)$$

$$(19) \qquad = P(\omega_{1,i} > \omega_{2,i} + \tau_2, \omega_{1,i} \le t) \quad = \quad P(\omega_{1,i} - \tau_2 > \omega_{2,i}, \omega_{1,i} \le t)$$

$$= \int_{-\infty}^{t} f_1(\omega_1) d\omega_1 \int_{-\infty}^{\omega_1 - \tau_2} f_2(\omega_2)\, d\omega_2 \quad = \quad \int_{-\infty}^{t} f_1(\omega_1)\, F_2(\omega_1 - \tau_2)\, d\omega_1$$

This means that we can define $\psi_1(t)$ as follows:

(20) $$\psi_1(t) = \frac{\partial}{\partial t}\int_{-\infty}^{t} f_1(\omega_1)\, F_2(\omega_1-\tau_2)\, d\omega_1 = f_1(t)\, F_2(t-\tau_2)$$

An analogous argument defines $\psi_2(t)$:

(21) $$\psi_2(t) = \frac{\partial}{\partial t}\int_{-\infty}^{t} f_2(\omega_2)\, F_1(\omega_2+\tau_2)\, d\omega_2 = f_2(t)\, F_1(t+\tau_2)$$

Going back to the final integral in equation (19) and carrying out integration-by-parts yields:

(22) $$\Psi_1(t) = \int_{-\infty}^{t} f_1(\omega_1)F_2(\omega_1-\tau_2)d\omega_1 = F_1(t)F_2(t-\tau_2) - \int_{-\infty}^{t} F_1(s)f_2(s-\tau_2)ds$$

Performing a change of variables $u = s-\tau_2$, equation (22) becomes:

(23) $$\Psi_1(t) = F_1(t)F_2(t-\tau_2) - \int_{-\infty}^{t-\tau_2} F_1(u+\tau_2)f_2(u)du$$

Next, we use the expressions for $\psi_1(t)$ and $\psi_2(t)$ defined in (20) and (21) to re-write equation (23) as follows:

(24) $$\Psi_1(t) = \frac{F_1(t)\psi_1(t)}{f_1(t)} - \int_{-\infty}^{t-\tau_2} \psi_2(u)du$$

Noting that the second integral in (24) is simply $\Psi_2(t-\tau_2)$, we can solve for the distribution of $\omega_{1,i}$ as a function of τ_2:

(25) $$\lambda_1(t) = \frac{f_1(t)}{F_1(t)} = \frac{\psi_1(t)}{\Psi_1(t)+\Psi_2(t-\tau_2)}$$

where $\lambda_1(t)$ is a function of the unconditional wage distribution in location #1. (25) is a single equation in two unknowns ($\lambda_1(t)$ and τ_2) for a particular value of *t*, and it is therefore not surprising that we cannot identify both of these values without making an additional assumption. One solution would involve making a parametric assumption about $F_1(t)$. For example, assuming $F_1(t) \sim N(\mu_1, \sigma_1^2)$ would reduce the equation to three parameters. The number of parameters would not increase, however, as one considered the expression evaluated at different values of *t*. By forcing the equation to hold for many values of *t*, we would have more equations than unknowns and could identify the model's parameters.

In the following section, we show how the assumption of commonality can be used to non-parametrically recover $\lambda_1(t)$ and τ_2 . This assumption will be analogous to the sort of exclusion restriction that is typically used to achieve identification – in particular, birth location is a variable that affects the location decision, but is assumed to be excluded from the determinants of wages.

3.2 Identification via Commonality in the 2 x 2 Case

Consider now the case of individuals born into one of two locations (again indexed by 1 and 2), who decide where to reside based on the maximization of utility. This introduces the need for additional notation – we use a superscript to indicate origin location and a subscript to indicate destination location.

The dummy variable indicating that an individual originating in location #1 chooses to stay in that location is given by:

$$(26) \qquad d_i^1 = I[\omega_{1,i}^1 > \omega_{2,i}^1 + \tau_2^1]$$

while the indicator that an individual originating in location #2 chooses not to migrate is given by:

$$(27) \qquad d_i^2 = I[\omega_{2,i}^2 > \omega_{1,i}^2 + \tau_1^2]$$

As before, we normalize the taste parameter for those choosing not to migrate to zero (i.e., $\tau_1^1 = \tau_2^2 = 0$). With these indicators, we can now write the expression for the observed wage of an individual *i* who originates in location #1:

(28) $$w_i^1 = d_i^1\omega_{1,i}^1 + (1-d_i^1)\omega_{2,i}^1$$

Based on these definitions for d and w, we define the following expressions analogously to the previous sub-section:

(29) $$\begin{array}{ll} \Psi_1^1(t) = P(d_i^1 = 1, w_i^1 \le t) & \Psi_2^1(t) = P(d_i^1 = 0, w_i^1 \le t) \\ \Psi_1^2(t) = P(d_i^2 = 0, w_i^2 \le t) & \Psi_2^2(t) = P(d_i^2 = 1, w_i^2 \le t) \end{array}$$

Continuing in a manner similar to the previous sub-section, we can use equation (29) to derive the following four expressions:

(30) $$\lambda_1^1(t) = \frac{f_1^1(t)}{F_1^1(t)} = \frac{\psi_1^1(t)}{\Psi_1^1(t) + \Psi_2^1(t-\tau_2^1)}$$

(31) $$\lambda_2^1(t) = \frac{f_2^1(t)}{F_2^1(t)} = \frac{\psi_2^1(t)}{\Psi_2^1(t) + \Psi_1^1(t+\tau_2^1)}$$

(32) $$\lambda_1^2(t) = \frac{f_1^2(t)}{F_1^2(t)} = \frac{\psi_1^2(t)}{\Psi_1^2(t) + \Psi_2^2(t+\tau_1^2)}$$

(33) $$\lambda_2^2(t) = \frac{f_2^2(t)}{F_2^2(t)} = \frac{\psi_2^2(t)}{\Psi_2^2(t) + \Psi_1^2(t-\tau_1^2)}$$

By itself, the expansion of the 1 x 2 case to the 2 x 2 case does nothing to help with identification. It does, however, allow us to introduce an additional assumption – commonality. Under the assumption of commonality, $\lambda_1^1(t) = \lambda_1^2(t)$ and $\lambda_2^1(t) = \lambda_2^2(t)\ \forall\ t$. Under this assumption, we can re-write equations (30)-(33) as the following two equations:

(34) $$\lambda_1^1(t) = \frac{\psi_1^1(t)}{\Psi_1^1(t) + \Psi_2^1(t-\tau_2^1)} = \frac{\psi_1^2(t)}{\Psi_1^2(t) + \Psi_2^2(t+\tau_1^2)} = \lambda_1^2(t)$$

(35) $$\lambda_2^1(t) = \frac{\psi_2^1(t)}{\Psi_2^1(t) + \Psi_1^1(t+\tau_2^1)} = \frac{\psi_2^2(t)}{\Psi_2^2(t) + \Psi_1^2(t-\tau_1^2)} = \lambda_2^2(t)$$

Estimation proceeds by forming minimum distance criterion functions based on equations (34) and (35):

(36) $\lambda_1^1(t;\tau_2^1) - \lambda_1^2(t;\tau_1^2) = 0$

(37) $\lambda_2^1(t;\tau_2^1) - \lambda_2^2(t;\tau_1^2) = 0$

and then relying on the properties of M-estimators to recover τ_2^1 and τ_1^2. [Davidson and MacKinnon (1993)] We then use these taste parameters along with a Kaplan-Meier procedure to recover estimates of $f_1(\omega_1)$ and $f_2(\omega_2)$ as described in Section 2.2.

We now provide sufficient conditions for identification and estimation of the taste parameters in the 2 x 2 setting with commonality. We begin by rearranging the expressions (34) and (35):

(38) $\Psi_2^1(t-\tau_2^1)\psi_1^2(t) - \psi_1^1(t)\Psi_2^2(t+\tau_1^2) = \Psi_1^2(t)\psi_1^1(t) - \Psi_1^1(t)\psi_1^2(t) = H(t)$

(39) $\Psi_1^1(t+\tau_2^1)\psi_2^2(t) - \psi_2^1(t)\Psi_1^2(t-\tau_1^2) = \Psi_2^2(t)\psi_2^1(t) - \Psi_2^1(t)\psi_2^2(t) = J(t)$

Note that the right-hand-side of each of these expressions is an observable function of the data for a particular value of *t*. Our identification result begins with the following lemma:

Lemma 1: *At the true parameter values (* τ_2^1*, τ_1^2* *), we have*

(40)
$$\left(\Psi_2^1(t-\tau_2^1*)\psi_1^2(t) - \psi_1^1(t)\Psi_2^2(t+\tau_1^2*) - H(t)\right)^2 + \left(\Psi_1^1(t+\tau_2^1*)\psi_2^2(t) - \psi_2^1(t)\Psi_1^2(t-\tau_1^2*) - J(t)\right)^2 = 0$$

$\forall\, t \in \Re$ *in the intersection of the supports of* $\psi_1^1(t)$, $\psi_2^1(t)$, $\psi_1^2(t)$, *and* $\psi_2^2(t)$.

This is simply a re-statement of our minimum distance criterion function described above. We will now show that, for each set of values of the taste parameters different from (τ_2^1*, τ_1^2*), denoted by ($\tilde{\tau}_2^1, \tilde{\tau}_1^2$), we must have:

$$\left(\Psi_2^1(t-\tilde{\tau}_2^1)\psi_1^2(t)-\psi_1^1(t)\Psi_2^2(t+\tilde{\tau}_1^2)-H(t)\right)^2 + \tag{41}$$

$$\left(\Psi_1^1(t+\tilde{\tau}_2^1)\psi_2^2(t)-\psi_2^1(t)\Psi_1^2(t-\tilde{\tau}_1^2)-J(t)\right)^2 > 0$$

for some $t \in \Re$ in the intersection of the supports of $\psi_1^1(t)$, $\psi_2^1(t)$, $\psi_1^2(t)$, and $\psi_2^2(t)$. To prove this result, first note that if $\tilde{\tau}_2^1 = \tau_2^1*$, then only $\tilde{\tau}_1^2 = \tau_1^2*$ will make equation (40) hold, by the monotonicity of the conditional c.d.f.'s that make-up that expression. By a similar argument, if $\tilde{\tau}_1^2 = \tau_1^2*$, then $\tilde{\tau}_2^1 = \tau_2^1*$ in order for equation (40) to hold. Therefore, we need only consider the case in which $\tilde{\tau}_2^1 \neq \tau_2^1*$ and $\tilde{\tau}_1^2 \neq \tau_1^2*$. I.e., is it possible that an imposter pair ($\tilde{\tau}_2^1$, $\tilde{\tau}_1^2$) could satisfy equation (40)?

Consider the following condition which we argue will be sufficient to rule out this possibility:

$$\psi_2^1(t-\tau_2^1*)\psi_1^2(t)\psi_2^1(t)\psi_1^2(t-\tau_1^2*) \neq \psi_1^1(t+\tau_2^1*)\psi_2^2(t)\psi_1^1(t)\psi_2^2(t+\tau_1^2*) \tag{42}$$

for some $t \in \Re$ in the intersection of the supports of $\psi_1^1(t)$, $\psi_2^1(t)$, $\psi_1^2(t)$, and $\psi_2^2(t)$. This condition has a simple interpretation – i.e., that the Jacobian matrix associated with equations (38) and (39) is non-singular. There are situations in which this condition will not hold; for example, when the two conditional wage distributions are identical and $\tau_2^1 = -\tau_1^2$.[11] We consider this to be a pathological case.

To establish the sufficiency of the above condition for identification, consider a local linearization of equations (38) and (39) around the true values of τ_2^1 and τ_1^2 and evaluated at *t*. For any pair of perturbations, Δ_2^1 and Δ_1^2, we require the net effect on the left-hand-side of each equation to be zero (since *H(t)* and *J(t)* are functions of only *t*).

$$\psi_2^1(t-\tau_2^1*)\psi_1^2(t)\Delta_2^1 + \psi_1^1(t)\psi_2^2(t+\tau_1^2*)\Delta_1^2 = 0 \tag{43}$$

[11] This would be the case if we took a single location and arbitrarily divided it into two locations with the exact same wage distributions and amenities. This condition therefore places a practical constraint on the level of geographic precision at which we can apply our estimator – i.e., at the level at which we can observe different spatial wage distributions.

(44) $\psi_1^1(t+\tau_2^1*)\psi_2^2(t)\Delta_2^1 - \psi_2^1(t)\psi_1^2(t-\tau_1^2*)\Delta_1^2 = 0$

If condition (42) holds, then the only solution to these expressions is given by $\Delta_2^1 = \Delta_1^2 = 0$, implying that no imposter values of ($\tilde{\tau}_2^1, \tilde{\tau}_1^2$) could satisfy the system. Note that every thing we have here is for a given t. Thus by varying values of t over values in the intersection of the supports of $\psi_1^1(t)$, $\psi_2^1(t)$, $\psi_1^2(t)$, and $\psi_2^2(t)$, equations (43) and (44) overidentify the parameters of interest under our assumptions of independence and commonality. Consequently a test for commonality could possibly be based on how far the value of our minimum distance objective function is from 0.

4. Asymptotics

Having described two identification strategies for both the taste parameters and unconditional wage distributions, we now outline the arguments that will be used in developing the asymptotic properties of our proposed estimators. We begin with a discussion of our minimum order statistic estimator. In practice, we simply replace population extreme quantiles in the identification argument with sample minimum order statistics. Asymptotic properties of minimum or maximum order statistics have been studied in recent work by Porter and Hirano (2003). Chernozhukov and Hong (2004) obtain similar results. As a preliminary step, we establish the rate of convergence of the estimator. The result is based on the following regularity conditions:

A1 The K+1 vectors of observed wage and choice indicators (w_i, $d_{k,i}$) are i.i.d. across individuals.

A2 The unconditional wage distributions for alternatives k = 1, 2, …, K are continuously distributed with positive density on $[\ell_k, \infty)$.

A3 $\min_{k=1,2,\ldots K} \ell_k > -\infty$

A4 $\min_{k=1,2,\ldots K} P(d_{k,i} = 1) > 0$

Theorem 0.1 *Under Assumptions A1-A4, we have*

(45) $\hat{\tau}_k - \tau_k = O_p(n^{-1})$

A proof that our estimator attains this rate of convergence under Assumption A2 follows from arguments similar to those used in van der Vaart (1998), Section 21.4.

Turning attention to the second stage estimator of the unconditional wage distributions, we proposed applying Kaplan-Meier to yield a consistent estimator of the distribution of $\omega_{k,i}+\tau_k$. We note the first stage estimator, which was shown to be "super-consistent", will have no effect on the limiting distribution of the second stage estimator. The next theorem establishes the limiting distribution of this estimator.

Theorem 0.2 *Under Assumptions A1-A4, our second stage estimator of the unconditional wage distribution has the following linear representation. Let $\pi(t) = P(\omega_{k,i} \le t)$ and define the set $\Omega = \{t : \pi(t) < 1\}$. Then for any $t \in \Omega$,*

$$(46) \qquad \sqrt{n}(\hat{F}_{\omega_{k,i}+\tau_k}(\cdot) - F_{\omega_{k,i}+\tau_k}(\cdot)) \Rightarrow F_{\omega_{k,i}+\tau_k}(\cdot)W(\varphi(\cdot))$$

where W is Brownian motion, and $\varphi(t) = \int_t^\infty \pi^{-1}(s)ds$.

A proof of the above theorem can be found by using the same arguments as in Fleming and Harrington (1991). We omit the details here.

We now turn our attention to the asymptotic properties of the unbounded support estimator. To illustrate the basic arguments involved, we will focus on the two-region setting. Our estimator of the taste parameter vector, $\hat{\tau} = (\hat{\tau}_2^1, \hat{\tau}_1^2)$, is obtained by minimizing the minimum distance objective function:

$$(47) \qquad \hat{\tau} = \arg\min_{\tau} \frac{1}{n}\sum_{i=1}^{n} Q(\tau, t_i)$$

The asymptotic properties of our unbounded support estimator are based on the following assumptions:

B1 The K+1 vectors of observed wage and choice indicators (w_i, $d_{k,i}^j$) are i.i.d. across individuals.

B2 The true vector τ_0 lies in the interior of a compact parameter space.

B3 The functions $\psi_m^l(\cdot)$, l, m = 1, 2 are assumed to be uniformly bounded and twice continuously differentiable, with uniformly bounded first and second derivatives.

B4 The kernel function $K(\cdot)$ used to approximate $\psi_m^l(\cdot)$ has bounded support, integrates to one, and has mean zero.

B5 The bandwidth h associated with kernel function $K(\cdot)$ satisfies $\sqrt{n}h^2 \to 0$ and $nh \to \infty$.

Theorem 0.3 *Under Assumptions B1-B4,*

(48) $$\hat{\tau} \xrightarrow{p} \tau_0$$

The proof of the above result can be shown by establishing the four sufficient conditions in Theorem 2.1 in Newey and McFadden (1994), which can be characterized as compactness, identification, uniform convergence, and continuity.

Furthermore, by Newey and McFadden (1994) Theorem 8.11, we can establish the parametric rate of convergence as well as the asymptotic normality of our estimator. The parametric rate is attainable despite the nonparametric rate of convergence achieved by some components because the parameter of interest (τ_0) is a smooth functional of the nonparametric components. Our next theorem is based on the following assumptions:

B6 The functions $\psi_m^l \quad l,m = 1,2$ are assumed to be uniformly bounded and p times continuously differentiable, with uniformly bounded p^{th} order derivatives.

B7 The kernel function K integrates to one, has mean zero, and is of p^{th} order.

B8 The bandwidth h associated with the kernel function satisfies $\sqrt{n}h^p \to 0$ and $nh \to \infty$.

The following theorem establishes the root-*n* consistency and asymptotic normality of our estimator. Its proof is omitted as it follows from the same arguments used in proving Theorem 8.11 in Newey and McFadden (1994).

Theorem 0.4 *Under Assumptions B1, B2, B4-B8,*

(49) $$\sqrt{n}(\tau - \tau_0) \Rightarrow N(0, \Omega_x)$$

Where $\Omega_x > 0$.

Remark 1 *The exact form of* Ω_x *is complicated, as it involves higher order derivatives of the functions* ψ_m^l. *The details of its form in the 2x2 case, as well as a sketch of the proof of the theorem are left to the appendix. Due to its complicated form, we employ sampling methods for inference on* τ_0 *in our application in order to avoid introducing additional nonparametric methods.*

Remark 2 *Note that, in this case, the first stage estimator converges at the parametric rate, and consequently will affect the limiting distribution of the second stage estimator. While the* *precise effect on the limiting distribution can be derived using arguments similar to those used by Newey and McFadden (1994) Section 8, we omit the details here.*

5. Monte Carlo Results

In this section, we use Monte Carlo experiments to describe the properties of both estimators in small samples and with less-than-ideal data. We consider a simple setting with just three locations that serve as both origins and destinations, and we model the sorting decisions of individuals who care about both pecuniary returns (i.e., wages) and non-pecuniary factors (i.e., migration costs and amenities) in deciding where to live. In each experiment, we consider some number of identical individuals (N) originating in each location, and we use their simulated behavior to recover the matrix of taste parameters:

$$(50)\qquad \begin{bmatrix} \tau_1^1 & \tau_2^1 & \tau_3^1 \\ \tau_1^2 & \tau_2^2 & \tau_3^2 \\ \tau_1^3 & \tau_2^3 & \tau_3^3 \end{bmatrix} = \begin{bmatrix} 0 & -0.5 & -0.2 \\ -0.4 & 0 & -0.6 \\ -0.3 & -0.1 & 0 \end{bmatrix}$$

For the sake of simplicity in exposition, we focus our attention on the performance of the estimators in recovering these taste parameters. Unconditional wage distributions in each location could be recovered by applying the Kaplan-Meier technique described in section 2.2 for each set of Monte Carlo estimates.

We begin by looking at the minimum order statistic estimator. The results of nine Monte Carlo experiments are described in Table 1. The first three experiments use the baseline framework in which wages are random variables determined by the following (*j* denotes origin location):

$$(51)\qquad \begin{aligned} \omega_1^j &= \sqrt{x_1^2 + 2.25} \\ \omega_2^j &= \sqrt{x_2^2 + 1.75} \qquad j = 1, 2, 3 \qquad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}\right] \\ \omega_3^j &= \sqrt{x_3^2 + 2.75} \end{aligned}$$

Columns describe the various taste parameter estimates, while rows summarize the mean, standard deviation, and mean squared error of 500 Monte Carlo simulations for each experiment.

With an increasing number of individuals in each origin location, the minimum order statistic becomes a better measure of the true lower bound on wages in a particular location, and our estimates of the taste parameters improve accordingly. This is evident in the declining MSE as N increases from 1,000 to 10,000 to 50,000 for each parameter. Even with as few as 1,000 observations, however, taste parameter estimates based on the minimum order statistic are quite precise.

The fourth and fifth experiments in Table 1 relax the assumption of a finite lower bound for the unconditional wage distribution. In particular,

$$\begin{aligned} & f(\omega_1^j) \sim N(2.25, 0.5) \\ (52) \qquad & f(\omega_2^j) \sim N(1.75, 0.5) \qquad\qquad j = 1, 2, 3 \\ & f(\omega_3^j) \sim N(2.75, 0.5) \end{aligned}$$

The impact of this model mis-specification is evident in an increase in the MSE by a factor of 100 to 10,000, depending upon the parameter. Conditional upon this mis-specification, however, MSE's still fall as N increases from 10,000 to 50,000.

The sixth and seventh experiments address an important concern with our minimum order statistic estimator – measurement error. Because the estimator relies on a single value of wages for each origin and destination combination, it could become severely biased if that value were mis-measured. In these experiments, we return to the same wage distributions used in the first three experiments (i.e., assuming a finite lower bound), but we add to each wage an i.i.d. normally distributed random variable with zero mean and variance equal to 0.25. This has a significant impact on the precision of our estimates, raising the MSE's associated with our taste parameters by nearly as much as the absence of a finite lower bound. In contrast to that model mis-specification, however, this is primarily the result of an increase in the bias of our estimator, as opposed to its standard deviation.

In the eighth and ninth experiments, we demonstrate a desirable feature of the minimum order statistic estimator – the fact that it is robust to arbitrary forms of correlation in wage draws. Using the same wage distributions as in our baseline specifications, we assume a correlation of 0.25 between wage draws in all locations. As is evident from the table, taste parameter estimates are virtually identical to the baseline case.

Table 2 describes the results of nine Monte Carlo experiments that similarly illustrate the properties of our unbounded support estimator. Using the same matrix of taste parameters, we assume in our baseline experiment that wages are drawn from the same distributions as described in equation (52). The first three experiments demonstrate the effect of increasing the number of individuals originating from each location (N) from 1,000 to 50,000. MSE's of all taste parameter estimates fall with an increase in the sample size. In general, however, results are not as precise as under the (properly specified) minimum order statistic estimator (conditional upon N).

In the next two experiments, we show the implications of violating our key identifying assumption – commonality. In particular, we allow non-migrants to receive a higher wage on average than individuals migrating into their birth location (i.e., a "home advantage" in the labor market).

$$
\begin{array}{lllll}
f(\omega_1^j) \sim N(2.25, 0.5) & if & j = 2, 3 & otherwise & f(\omega_1^1) \sim N(2.5, 0.5) \\
f(\omega_3^j) \sim N(1.75, 0.5) & if & j = 1, 3 & otherwise & f(\omega_2^2) \sim N(2, 0.5) \\
f(\omega_3^j) \sim N(2.75, 0.5) & if & j = 1, 2 & otherwise & f(\omega_3^3) \sim N(3, 0.5)
\end{array} \tag{53}
$$

We assume, moreover, that the researcher properly identifies this home advantage and uses only moments formed between pairs of migrant groups (e.g., migrants from locations #2 and #3 living in location #1) in forming our minimum distance objective function. Not surprisingly, with this limited set of moments the model does not perform as well as in the baseline specification. It does, however, do a reasonable job of estimating all parameters (even with only 10,000 observations per origin location). When N is set equal to 50,000, the estimates become quite precise, indicating that our estimation strategy is indeed valid under situations of "limited commonality".

The sixth and seventh experiments describe what happens when another key assumption used in the derivation of the unbounded support estimator – independence – is violated. Recall that, in the derivation of equation (19), we assumed individuals received draws from independent wage distributions. Here, we assume that wage draws exhibit a positive correlation (0.25) across locations. MSE's for all taste parameters rise dramatically, highlighting this as an important shortcoming of our estimation strategy. In current research, we are exploring how correlation might be better handled using panel data. With only cross-sectional data, these results highlight the importance of controlling for as many forms of observable heterogeneity as possible (i.e., wages may be systematically higher for certain groups – the estimation algorithm should be run separately for them). Our final set of experiments describe the effect of measurement error on our unbounded support estimator. As was the case for the minimum support estimator, we simply add to each wage an i.i.d. normal measurement error with mean zero and variance 0.25. In contrast to the minimum order statistic estimator, however, the results of the unbounded support estimator are affected very little.

In summary, Monte Carlo simulations suggest that our minimum order statistic estimator performs extremely well when properly specified. It is, moreover, robust to arbitrary forms of correlation in an individual's wage draws, but it performs very poorly when wages are observed with error or when they are drawn from a distribution without a finite lower bound. These failures motivate our derivation of the unbounded support estimator. When properly specified, experiments show that it also performs well. Moreover, its performance is not adversely affected by measurement error in wages or by limited commonality (if the researcher properly recognizes this in forming the minimum distance objective function). In contrast to the minimum order statistic estimator, however, it performs poorly when wage draws are correlated across locations (motivating our current work with panel data).

6. Empirical Application: Measuring the Returns to College Education

In order to demonstrate the performance of our estimator in an empirical setting, we examine a question similar to that posed by Dahl (2002) – i.e., what are the returns to a college education (relative to graduating from high school) before and after controlling for the non-random spatial sorting of workers across the United States? The results of the basic Roy model (1951) suggest that sorting shifts the means of the (observed) conditional wage distributions up from their (unobserved) unconditional values. Whether spatial sorting increases or reduces the estimated returns to a college education will depend upon whether this shift is proportionally bigger for high school or college educated individuals. If, for example, college educated individuals were more mobile and, hence, more able to migrate in response to favorable idiosyncratic wage draws, we would expect spatial sorting to create an upward bias in the estimated returns to a college education. Whether or not this is the case (and how big is the resulting bias) is an empirical question.

In order to answer that question, we use data extracted from the 2000 US Census 5% microsample, available from the IPUMS (www.ipums.org). Specifically, we consider a sample of 470,918 high school graduates taken from each of nine divisions of the United States used by the Census Bureau, along with a corresponding sample of 429,584 college graduates.[12] We use only data describing male household heads.[13] For each individual, we observe annual income from wages and salary, the individual's region of residence, and the individual's region of birth.[14] Tables 3 and 4 summarize the long-run migration probabilities observed in the data for high school and college graduates, respectively, for each of four summary birth and destination regions. In particular, each row indicates the birth region while each column indicates the region in which the individual is observed in the 2000 Census. Each entry describes the fraction of individuals originating in the row birth region who are found to be living in the column destination region. 80.6% of high-school graduates born in New England are found to be living in New England. The fraction of high school graduate "stayers" is similarly high for other regions.[15] For college graduates, a noticeably lower percentage remains tied to their respective birth regions.

Because Census wage data, which are derived from self-reported income and hours information, are quite noisy in the lower tail (see footnote 5), and because we see individuals from multiple birth places, we opt for our unbounded support estimator. This estimator makes an independence assumption and assumes that individuals from different birth regions will receive wage draws from a common destination wage distribution. Note that, with different data, the extreme quantile estimator (which only assumes extreme quantile independence and that wage distributions have a finite lower bound) might be used instead. Deleire, Khan and Timmins (2009) use this estimator, along with CPS wage data, to recover an estimate of the value of a statistical life (VSL) controlling for Roy sorting across occupations.

Tables 5 and 6 report the estimates of the taste parameters for high school and college graduates, respectively. Results are measured in terms of the natural log of hourly wages,

---

[12] Regional Definitions: (1) *New England* (CT, ME, MA, NH, RI, VT), (2) *Middle Atlantic* (NJ, NY, PA), (3) *East North Central* (IL, IN, MI, OH, WI), (4) *West North Central* (IA, KS, MN, MO,NE, SD, ND), (5) *South Atlantic* (DE, DC, FL, GA, MD, NC, SC, VA, WV), (6) *East South Central* (AL, KY, MS, TN), (7) *West South Central* (AR, LA, OK, TX), (8) *Mountain* (AZ, CO, ID, MT, NV, NM, UT), and (9) *Pacific* (AK, CA, HI, OR, WA).

[13] We use only household heads because we assume they are more likely to have made their own geographic location decision, and we use only individuals less than 35 years of age as they are more likely to have recently migrated. Older individuals may have migrated further in the past in response to different wage or amenity distributions.

[14] We drop any individuals reporting zero annual income, self-employed individuals, individuals not born in the United States, and individuals who worked fewer than 45 weeks in the previous year. The US Census describes both the individual's birth state as well as the PUMA in which he/she was living five years prior. We use birth state to define birth region, which becomes our measure of "origin location", but a similar analysis could be performed using location five years prior as the "origin", leading to a short-run measure of mobility cost.

[15] Note that the fraction of "stayers" would be smaller if we had used a finer geographic division (e.g., states), but would still constitute a clear plurality.

standard errors are derived from the results of 750 bootstrap simulations, and point estimates are bias-corrected. A college graduate from the mid-Atlantic, for example, faces a statistically significant cost of -0.622 per year in moving to the Pacific region. Considering the mean wage amongst all college graduates ($26.22), this amounts to a compensating variation (defined implicitly according to $\ln(26.22) = \ln(26.22 + CV) - 0.622$) of $22.62 per hour. All off-diagonal taste parameters are negative and significant, revealing the tendency for individuals of all levels of education to remain in their birth regions.

Next, we use these estimates to recover the unconditional income distributions for each region and education group with the Kaplan-Meier procedure described in Section 2.2. Results are reported in Figures 1 and 2. In every case, the unconditional wage distribution lies below the observed distribution. Importantly, the correction for Roy sorting is generally larger for college graduates, who are more prone to migrate from their birth regions. We record the medians and 75th percentiles of each of these distributions in Table 7. The median log-wage for a high-school graduate from the South Atlantic, for example, falls from 2.63 to 2.55. Defining the returns to a college education at the median to be the difference between the median of the college and high-school graduate log-wage distributions, we report those returns in Table 8. Returns are analogously defined at the 75th percentile. In every region, the returns to a college education fall once we control for Roy sorting. As in Dahl (2002), we find significant evidence of an upward bias in the returns to a college education because of Roy sorting, although we generally find the size of that bias to be bigger. On average, our measure of returns falls from 42% to 18% at the median, and from 45% to 34% at the 75th percentile. These results suggest that observed wage distributions, which are distorted by Roy sorting, seriously overstate the true returns to a college education, particularly for those in the heart of the wage distribution.

7. Conclusion

This paper considers identification and estimation of a multi-sector Roy model which includes a common non-pecuniary component of utility associated with each alternative. Two identification results are established – one under an extreme quantile independence condition and the other under a commonality/independence assumption, where commonality provides us with a convenient exclusion restriction that facilitates identification. Estimation procedures based on both identification results are proposed, and their asymptotic properties are derived. These models are able to identify the common taste parameter explicitly, and provide identification in a data environment where the approach used in previous work (i.e., identification at infinity) may prove difficult.

Our second estimator is used to recover an estimate of the returns to a college education, controlling for different migration rates of high-school and college graduates. We report the values of the taste parameters that, along with wage draws, determine migration decisions, and we recover non-parametric estimates of the unconditional wage distributions from which individuals receive draws. The results suggest that an estimate based on conditional distributions may overstate returns by more than a factor of two at the median. An application of our first estimator to sorting across occupations, where individuals care about pecuniary returns and other job attributes (including fatality risk) yields similarly stark results. In that analysis, the wage-hedonic estimate of the value of a statistical life rises by a factor of four and becomes statistically significant. [Deleire, Khan and Timmins (2009)]

Finally, it is important to note that the models we develop here could be applied to estimate a "generalized" version of the competing risks model. Specifically, the "risk" indicator is no longer simply determined by the minimum of the two variables whose distributions one aims to identify. So it would be useful to further explore how the results attained here can be useful in some of the variations introduced in that literature- i.e. allowing for data complications such as left truncation and length biased sampling.

Appendix: Asymptotic Properties of Proposed Minimum Distance Estimator for τ.

Theorem A1. Under Assumptions B1-B8, letting $\hat{\tau}$ denote our proposed minimum distance estimator of the parameter τ_0

$$\sqrt{n}(\hat{\tau}-\tau_0) \Rightarrow \mathrm{N}\left(0, V^{-1}\Delta V^{-1}\right) \qquad (A.1)$$

where V, a Hessian term, is defined below, and

$$\Delta = \mathrm{E}\left[(\delta_{1i}+\delta_{2i}+\delta_{3i})(\delta_{1i}+\delta_{2i}+\delta_{3i})'\right]$$

δ_{1i}, δ_{2i}, δ_{3i} are separate score terms, also described below. δ_{2i}, δ_{3i} correspond to "correction" terms, that account for ψ and Ψ being estimated and plugged into the objective function, respectively. As we will see, both V and Δ involve unknown functions and their derivatives, thus making them difficult to estimate. This motivates the use of the bootstrap in our application to conduct inference.

We begin by describing the estimator in more detail. From a random sample of n observations of wages w_i, and a given t, let $\hat{\psi}(t)$, $\hat{\Psi}(t)$ denote sample analog estimators (based on a random sample of $i = 1,2,..n$ observations) $\psi(t)$, $\Psi(t)$ respectively. Specifically,

$$\hat{\Psi}_1^1(t) = \frac{1}{n}\sum_{i=1}^{n} d_i^1 I\left[w_i^1 \leq t\right] \qquad (A.2)$$

and an estimator of its derivative w.r.t. t, is

$$\hat{\psi}_1^1(t) = \frac{1}{nh}\sum_{i=1}^{n} d_i^1 K_h(w_i - t) \qquad (A.3)$$

where $K()$ is a kernel function, h is a bandwidth sequence, and $K_h(\cdot) = K(\cdot / h)$. Consequently, we can construct estimators $\hat{\lambda}_1^1(t, \tau_2^1)$, $\hat{\lambda}_1^2(t, \tau_1^2)$ based on (34), and similarly for (35); these are estimators for a given t as a function of τ_2^1, τ_1^2. For a grid of values $j = 1,2,..n$ of t our minimum distance estimator minimizes the following objective function:

$$Q_n(\tau) = \frac{1}{n}\sum_{j=1}^{n}\left(\hat{\lambda}_1^1(t_j, \tau_2^1) - \hat{\lambda}_1^2(t_j, \tau_1^2)\right)^2 + \left(\hat{\lambda}_2^1(t_j, \tau_2^1) - \hat{\lambda}_2^2(t_j, \tau_1^2)\right)^2 \qquad (A.4)$$

where τ denotes τ_2^1, τ_1^2. Note that despite the quadratic distance function, $Q_n(\tau)$ is not differentiable in τ. Note also, the objective function involves a double summation as we are summing across observations and grid values, so the objective function can be viewed as a U-process in τ, analogous to objective functions used in, e.g. Sherman(1993). The key behind an asymptotic normality result in both that case and our example is that the expectation of the objective function, now taken with respect to t_i , w_i is differentiable in τ. The proof will work with the frameworks detailed in Sherman (1993,1994).

Sufficient conditions for rates of convergence and asymptotic normality can be found in two theorems in Sherman(1994), which we state here. Keeping our notation deliberately close to Sherman(1994b), here we denote our sample objective function by $Q_n(\tau)$ and denote our limiting objective function by $Q(\tau)$. From Theorem 1 in Sherman(1994b), sufficient conditions for rates of convergence are that

1. $\hat{\tau} - \tau_0 = O_p(\delta_n)$
2. There exists a neighborhood of β_0 and a constant $\kappa > 0$ such that $Q(\tau) - Q(\tau_0) \geq \kappa\|\tau - \tau_0\|^2$ for all τ in this neighborhood.
3. Uniformly over $O_p(\delta_n)$ neighborhoods of τ_0

$$Q_n(\tau) = Q(\tau) + O_p\left(\|\tau - \tau_0\| / \sqrt{n}\right) + o_p\left(\|\tau - \tau_0\|^2\right) + O_p(\varepsilon_n) \qquad (A.5)$$

in which case $\hat{\tau} - \tau_0 = O_p\left(\max(\varepsilon_n^{1/2}, n^{-1/2})\right)$. We can use the above result (sometimes repeatedly), to establish root-n consistency. Arguments for specific examples can be found in
Sherman(1994), Khan (2001), Khan and Tamer(2009). Since identical arguments can be used for the problem at hand, we omit showing this here.

After root-n consistency has been established we can apply Theorem 2 in Sherman(1994b) to attain asymptotic normality. A sufficient condition is that uniformly over $O_p\left(1/\sqrt{n}\right)$ neighborhoods of τ_0,

$$Q_n(\tau) - Q_n(\tau_0) = \frac{1}{2}(\tau - \tau_0)' V(\tau - \tau_0) + \frac{1}{\sqrt{n}}(\tau - \tau_0)' W_n + o_p\left(\frac{1}{n}\right) \qquad (A.6)$$

where W_n converges in distribution to a $N(0, \Delta)$ random vector, and V is positive definite. In this case the asymptotic variance of $\hat{\tau} - \tau_0$ is $V^{-1}\Delta V^{-1}$. We will now write the objective functions in terms of $\psi()$, $\Psi(\ )$ as in (34),(35); Specifically,

$$Q_n(\tau) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{\psi}_1^1(t_j)}{\hat{\Psi}_2^1(t_j) - \hat{\Psi}_2^2(t_j - \tau_2)} - \frac{\hat{\psi}_1^2(t_j)}{\hat{\Psi}_1^2(t_j) - \hat{\Psi}_2^2(t_j + \tau_1)}\right)^2 \qquad (A.7)$$
$$+ \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{\psi}_2^1(t_j)}{\hat{\Psi}_2^1(t_j) - \hat{\Psi}_2^2(t_j + \tau_2)} - \frac{\hat{\psi}_2^2(t_j)}{\hat{\Psi}_2^1(t_j) - \hat{\Psi}_1^2(t_j - \tau_1)}\right)^2$$

We will aim to represent $Q_n(\tau)$, a second order $U-$ process as in Theorem 1 In Sherman(1994). To illustrate our arguments, we will concentrate on the first "half" of the objective function:

$$Q_{1n}(\tau) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{\psi}_1^1(t_j)}{\hat{\Psi}_2^1(t_j) - \hat{\Psi}_2^2(t_j - \tau_2)} - \frac{\hat{\psi}_1^2(t_j)}{\hat{\Psi}_1^2(t_j) - \hat{\Psi}_2^2(t_j + \tau_1)}\right)^2 \qquad (A.8)$$

noting the second "half" can be dealt with similarly. It will prove convenient to subtract

$$\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{\psi}_1^1(t_j)}{\hat{\Psi}_2^1(t_j) - \hat{\Psi}_2^2(t_j + \tau_0)} - \frac{\hat{\psi}_1^2(t_j)}{\hat{\Psi}_1^2(t_j) - \hat{\Psi}_2^2(t_j - \tau_0)}\right)^2$$

which is simply the sample objective function evaluated at the true parameter. Note this has no effect on the optimization problem. Turning attention to the proof for the problem at hand, we will adopt the following notation for convenience:

Let $\psi_{1i}, \Psi_{21i}(\tau), \Psi_{11i}, \psi_{2i}, \Psi_{22i}(\tau), \Psi_{12i}$ denote $\psi_1^1(t_i)$, $\Psi_2^1(t_i - \tau_2^1)$, $\Psi_1^1(t_i)$, $\psi_1^2(t_i)$, $\Psi_2^2(t_i + \tau_1^2)$, $\Psi_1^2(t_i)$ respectively, and as before let $\hat{}$ denote estimated values. Also, let $H_{1i}(\tau) = (\Psi_{11i} - \Psi_{21i}(\tau))^{-1}$, $H_{2i}(\tau) = (\Psi_{12i} - \Psi_{22i}(\tau))^{-1}$.
Then we can define an objective function based on (34), as

$$Q_{1n}(\tau)\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\psi}_{1i}\hat{H}_{1i}(\tau) - \hat{\psi}_{2i}\hat{H}_{2i}(\tau)\right)^2 \qquad (A.9)$$

And we can distribution theory for $\hat{\tau}$. Results or an estimator based on optimizing sample analogs for the sum of (34) and (35) would follow from similar arguments, so here we only focus on the objective function in $(A.9)$. Our first step is work with the infeasible objective function:

$$\tilde{Q}_{1n}(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left(\psi_{1i}H_{1i}(\tau) - \psi_{2i}H_{2i}(\tau)\right)^2 \qquad (A.10)$$

In what follows, to further simplify notation, we will let $\kappa(t_i, \tau)$ denote the term inside the above summation; note this a standard optimization problem, where the objective function is smooth in τ. So we have, uniformly in $o_p(1)$ neighborhoods of τ_0,

$$\tilde{Q}_{1n}(\tau) - Q_n(\tau_0) = \frac{1}{2}(\tau - \tau_0)' V(\tau - \tau_0) + \frac{1}{\sqrt{n}}(\tau - \tau_0)' W_n + o_p\left(\frac{1}{n}\right) \qquad (A.11)$$

where here $V = E[\nabla_{\tau\tau}\kappa(t_i, \tau_0)]$, where $\nabla_{\tau\tau}$ denotes the second derivative operator, and

$$W_n = \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\nabla_\tau \kappa(t_i, \tau_0)$$

where ∇_τ denotes the first derivative operator; note the term in the summation has expected value 0 by our identification result. Therefore $\nabla_\tau \kappa(t_i, \tau_0)$ can be regarded as one component of the influence function in the linear representation of our estimator, which we will denote here by δ_{1i}.

We next deal with representations for remainder terms that arise by replacing the feasible objective function with the infeasible. The first remainder term is of the form:

$$-2\frac{1}{n}\sum_{i=1}^{n}\left(\psi_{1i}H_{1i}(\tau_0) - \psi_{2i}H_{2i}(\tau_0)\right)\psi_{1i}H_{1i}^2(\tau_0)\left(\hat{H}_{1i}^{-1}(\tau) - H_{1i}^{-1}(\tau)\right) \qquad (A.12)$$

At this stage we can expand the form of $\hat{H}_{1i}^{-1}(\tau)$ inside the above summation. This will result in a second order U-statistic. Importantly, this U statistic is mean 0 since $\hat{H}_{1i}^{-1}(\tau)$ is an unbiased estimator of $H_{1i}^{-1}(\tau)$. Let the kernel of the resulting U-statistic be denoted $u_1(t_i, w_j, \tau)$. That is $(A.14)$ can be expressed as:

$$\frac{1}{n(n-1)}\sum_{i \neq j} u_1(t_i, w_j, \tau)$$

where recall w_j denotes the wages in our sample, and t_i denotes the draws of cutpoints to evaluate the estimator of ψ, Ψ at. As the conditional expectation of the kernel given t_i is also 0 we can represent the U-statistic in $(A.14)$ as

$$\frac{1}{n}\sum_{i=1}^{n} \nabla_\tau E[u_1(t_i, w_j, \tau_0) \mid w_j](\tau - \tau_0) + r_n \qquad (A.13)$$

where r_n is $o_p(1/n)$ uniformly in τ in $O(1/\sqrt{n})$ neighborhoods of τ_0, using identical arguments to that in Sherman(1993). Thus $\nabla_\tau E[u_1(t_i, w_j, \tau_0) \mid w_j]$ can be considered an additional component in the influence function of the estimator.

We note identical arguments can be used to deal with a remainder term involving $\hat{H}_{2i}(\tau) - H_{2i}(\tau)$, resulting in a term analogous to $(A.13)$. Denote the sum of these two terms by $\frac{1}{n}\sum_{i=1}^{n} \delta_{2i}'(\tau - \tau_0)$, so δ_{2i} represents the "correction" that arises from estimating $H_{1i}(\tau)$, $H_{12}(\tau)$

.

Next, we note there will be a remainder term analogous to $(A.12)$, but now involving the terms $\hat{\psi}_{1i} - \psi_{1i}$, $\hat{\psi}_{2i} - \psi_{2i}$. These can be handled the same way – i.e., expanding the forms of the estimators, and attaining a U-statistic representation. For example, the linear term involving $\hat{\psi}_{1i} - \psi_{1i}$ is of the form

$$2\frac{1}{n}\sum_{i=1}^{n} (\psi_{1i} H_{1i}(\tau) - \psi_{2i} H_{2i}(\tau)) H_{1i}^{1}(\tau_0)(\hat{\psi}_{1i} - \psi_{1i}) \qquad \text{(A.14)}$$

Carrying the same arguments as above, the representation for (A.14) will be

$$2\frac{1}{n}\sum_{i=1}^{n} \left(\nabla_\tau(\psi_{1i} H_{1i}(\tau_0) - \psi_{2i} H_{2i}(\tau_0)) H_{1i}^{1}(\tau_0) f_T(w_i) - E[\nabla_\tau(\psi_{1i} H_{1i}(\tau_0) - \psi_{2i} H_{2i}(\tau_0)) H_{1i}^{1}(\tau_0) f_T(w_i)]\right)$$

where $f_T()$ denotes the density function of t_i. We can attain an analogous representation for the linear term involving $\hat{\psi}_{2i} - \psi_{2i}$. Collecting both pieces we will denote the resulting influence function for this component by δ_{3i}.

Finally, we note that the additional remainder terms are of the order $(\hat{\psi}_{1i} - \psi_{1i})^2$, $(\hat{\psi}_{2i} - \psi_{2i})^2$, $\left(\hat{H}_{2i}(\tau) - H_{2i}(\tau)\right)^2$, $\left(\hat{H}_{1i}(\tau) - H_{1i}(\tau)\right)^2$. These are all negligible in the sense that they result in terms that are $o_p(1/n)$ for τ uniformly in $O\left(1/\sqrt{n}\right)$ neighborhoods of τ_0.
Thus our result follows from $(A.6)$.

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.

Bayer, P., N. Keohane, and C. Timmins (2006). "Migration and Hedonic Valuation: The Case of Air Quality." NBER Working Paper 12106.

Borjas, George (1987). "Self-Selection and the Earnings of Immigrants." *American Economic Review*. 77:531-553.

Chernozhukov, V. and H. Hong (2004). "Likelihood Estimation and Inference in a Class of Nonregular Econometric Models." *Econometrica*. 72(5):1445-1480.

Dahl, Gordon (2002). "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets." *Econometrica.* 70(6):2367-2420.

Davies, Greenwood and Li (2001). "A Conditional Logit Approach to U.S. State-to-State Migration." *Journal of Regional Science*. 41(2):337-360.

Deleire, T, S. Khan, and C. Timmins (2009). "Roy Model Sorting and Non-Random Selection in the Valuation of a Statistical Life." Unpublished Manuscript.

Falaris, Evangelos (1987). "A Nested Logit Migration Model with Selectivity." *International Economic Review*. 28:429-43.

Fleming, T.R. and D.P. Harrington (1991). *Counting Processes and Survival Analysis*. Wiley-Interscience.

Heckman, James (1990). "Varieties of Selection Bias." *American Economic Review*. 80(2):313-18.

Heckman, James and Bo Honore (1989). “The Identifiability of the Competing Risk Model.” *Biometrika*. 76:325-30.

Heckman, James and Bo Honore (1990). “The Empirical Content of the Roy Model.” *Econometrica*. 58:1121-49.

Heckman, James and Edward Vytlacil (2008), “Econometric Evaluation of Social Programs, Part I”, *Handbook of Econometrics, Vol. 8,* 4779-5144.

Honore, Bo and Andrea Lleras-Muney (2007). “Bounds in Competing Risks Models and the War on Cancer” *Econometrica*. Forthcoming.

Kaplan, E.L. and P. Meier (1958). “Nonparametric Estimation from Incomplete Data.” *Journal of the American Statistical Association*. 53:457-481.

Kennan, John and James Walker (2005). “The Effect of Expected Income on Individual Migration Decisions.” NBER Working Paper 9585.

Khan, Shakeeb and Elie Tamer (2007). “Inference on Randomly Censored Regression Models Using Conditional Moment Inequalities”. *forthcoming, Journal of Econometrics.*

Lee, Simon (2006). “Identification of Competing Risks Model with Unknown Transformation of Latent Failure Times”, *Biometrika,* 93:996-1002.

Newey, W.K. and D. McFadden (1994). “Estimation and Hypothesis Testing in Large Samples.” In Engle, R.F. and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4. Amsterdam: North-Holland.

Petersen, A.V. (1976). “Bounds for a Joint Distribution with Fixed Sub-Distribution Functions.” *Proceedings of the National Academy of Science*. 73:11-13.

Porter, J. and K. Hirano (2003). “Asymptotic Efficiency in Parametric Structural Models with Parameter Dependent Support.” *Econometrica*. 71(5):1307-1338.

Roback, Jennifer (1982). “Wages, Rents, and the Quality of Life.” *Journal of Political Economy*. 90:1257-78.

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

van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge, U.K.: Cambridge University Presss.

Table 1: Monte Carlo Simulations
Minimum Order Statistic Estimator

		τ_2^1	τ_3^1	τ_1^2	τ_3^2	τ_1^3	τ_2^3
		-0.5	-0.2	-0.4	-0.6	-0.3	-0.1
	Baseline (N = 1,000)						
1	Mean	-0.542	-0.203	-0.408	-0.610	-0.316	-0.113
	Std Dev	0.031	0.002	0.006	0.007	0.011	0.009
	MSE	0.003	1.28×10^{-5}	1.03×10^{-4}	1.53×10^{-4}	3.63×10^{-4}	2.50×10^{-4}
	Baseline (N = 10,000)						
2	Mean	-0.510	-0.201	-0.402	-0.602	-0.303	-0.103
	Std Dev	0.007	4.06×10^{-4}	0.001	0.001	0.002	0.002
	MSE	1.40×10^{-4}	5.36×10^{-7}	4.68×10^{-6}	7.26×10^{-6}	1.71×10^{-5}	1.23×10^{-5}
	Baseline (N = 50,000)						
3	Mean	-0.503	-0.200	-0.400	-0.600	-0.301	-0.101
	Std Dev	0.002	1.45×10^{-4}	4.19×10^{-4}	5.00×10^{-4}	7.70×10^{-4}	6.51×10^{-4}
	MSE	1.79×10^{-5}	6.63×10^{-8}	5.54×10^{-7}	8.12×10^{-7}	2.02×10^{-6}	1.33×10^{-6}
	Unbounded Support (N = 10,000)						
4	Mean	-0.595	-0.173	-0.389	-0.564	-0.350	-0.184
	Std Dev	0.202	0.212	0.204	0.198	0.215	0.208
	MSE	0.050	0.046	0.042	0.040	0.048	0.050
	Unbounded Support (N = 50,000)						
5	Mean	-0.569	-0.177	-0.383	-0.572	-0.354	-0.161
	Std Dev	0.193	0.189	0.202	0.187	0.183	0.199
	MSE	0.042	0.036	0.041	0.036	0.036	0.043
	Measurement Error (N = 10,000)						
6	Mean	-0.668	-0.248	-0.502	-0.711	-0.439	-0.228
	Std Dev	0.062	0.048	0.050	0.055	0.057	0.056
	MSE	0.032	0.005	0.013	0.015	0.023	0.020
	Measurement Error (N = 50,000)						
7	Mean	-0.660	-0.248	-0.491	-0.697	-0.430	-0.223
	Std Dev	0.050	0.043	0.044	0.044	0.046	0.047
	MSE	0.028	0.004	0.010	0.011	0.019	0.017
	Correlated Wage Draws (N = 10,000)						
8	Mean	-0.516	-0.201	-0.402	-0.603	-0.305	-0.104
	Std Dev	0.011	3.89×10^{-4}	0.001	0.002	0.003	0.002
	MSE	3.78×10^{-4}	4.61×10^{-7}	5.38×10^{-6}	9.76×10^{-6}	2.96×10^{-5}	1.97×10^{-5}
	Correlated Wage Draws (N = 50,000)						
9	Mean	-0.506	-0.200	-0.401	-0.601	-0.302	-0.101
	Std Dev	0.004	1.40×10^{-4}	4.68×10^{-4}	5.76×10^{-4}	0.001	7.84×10^{-4}
	MSE	4.83×10^{-5}	5.54×10^{-8}	6.87×10^{-7}	1.02×10^{-6}	3.41×10^{-6}	2.12×10^{-6}

Table 2: Monte Carlo Simulations
Unbounded Support Estimator

		τ_2^1	τ_3^1	τ_1^2	τ_3^2	τ_1^3	τ_2^3
		-0.5	-0.2	-0.4	-0.6	-0.3	-0.1
	Baseline (N = 1,000)						
1	Mean	-0.731	-0.230	-0.441	-0.652	-0.404	-0.173
	Std Dev	0.444	0.332	0.422	0.544	0.323	0.156
	MSE	0.250	0.111	0.180	0.298	0.115	0.029
	Baseline (N = 10,000)						
2	Mean	-0.620	-0.199	-0.390	-0.584	-0.373	-0.171
	Std Dev	0.079	0.066	0.081	0.063	0.053	0.057
	MSE	0.021	0.004	0.007	0.004	0.008	0.008
	Baseline (N = 50,000)						
3	Mean	-0.614	-0.197	-0.381	-0.573	-0.375	-0.181
	Std Dev	0.047	0.033	0.043	0.032	0.029	0.042
	MSE	0.015	0.001	0.002	0.002	0.006	0.008
	Home Advantage (N = 10,000)						
4	Mean	-0.576	0.175	-0.103	-0.218	-0.290	-0.303
	Std Dev	0.163	0.415	0.347	0.460	0.133	0.269
	MSE	0.032	0.313	0.208	0.358	0.018	0.114
	Home Advantage (N = 50,000)						
5	Mean	-0.510	-0.101	-0.309	-0.426	-0.354	-0.193
	Std Dev	0.073	0.204	0.156	0.157	0.112	0.128
	MSE	0.005	0.078	0.032	0.055	0.015	0.025
	Correlated Wage Draws (N = 10,000)						
6	Mean	-1.362	-0.441	-0.959	-1.367	-0.814	-0.356
	Std Dev	0.487	0.322	0.625	1.032	0.657	0.176
	MSE	0.980	0.162	0.702	1.652	0.696	0.096
	Correlated Wage Draws (N = 50,000)						
7	Mean	-1.368	-0.419	-0.950	-1.268	-0.706	-0.421
	Std Dev	0.290	0.057	0.155	0.107	0.171	0.135
	MSE	0.838	0.051	0.327	0.458	0.194	0.121
	Measurement Error (N = 10,000)						
8	Mean	-0.643	-0.206	-0.401	-0.601	-0.382	-0.176
	Std Dev	0.077	0.068	0.083	0.066	0.055	0.060
	MSE	0.026	0.005	0.007	0.004	0.010	0.009
	Measurement Error (N = 50,000)						
9	Mean	-0.632	-0.202	-0.395	-0.592	-0.385	-0.187
	Std Dev	0.052	0.035	0.046	0.034	0.029	0.042
	MSE	0.020	0.001	0.002	0.001	0.008	0.009

Table 3: Mobility Matrix, High School Graduates
2000 US Census, 5% IPUMS Random Sample

		Destination Region								
		New England	Mid-Atlantic	E North Central	W North Central	South Atlantic	E South Central	W South Central	Mountain	Pacific
Birth Region	New England	0.806	0.035	0.011	0.005	0.083	0.008	0.012	0.017	0.024
	Mid-Atlantic	0.016	0.809	0.011	0.005	0.100	0.008	0.012	0.019	0.020
	E North Central	0.004	0.013	0.766	0.024	0.072	0.034	0.026	0.031	0.031
	W North Central	0.002	0.006	0.053	0.770	0.028	0.011	0.034	0.053	0.043
	South Atlantic	0.008	0.036	0.016	0.007	0.863	0.029	0.017	0.010	0.015
	E South Central	0.003	0.009	0.065	0.009	0.082	0.776	0.032	0.008	0.015
	W South Central	0.002	0.007	0.022	0.025	0.035	0.022	0.814	0.030	0.043
	Mountain	0.004	0.009	0.017	0.032	0.027	0.010	0.049	0.747	0.105
	Pacific	0.006	0.011	0.021	0.027	0.035	0.012	0.040	0.097	0.750

Table 4: Mobility Matrix, College Graduates
2000 US Census, 5% IPUMS Random Sample

		Destination Region								
		New England	Mid-Atlantic	E North Central	W North Central	South Atlantic	E South Central	W South Central	Mountain	Pacific
Birth Region	New England	0.619	0.073	0.033	0.012	0.129	0.012	0.024	0.029	0.070
	Mid-Atlantic	0.060	0.546	0.053	0.012	0.182	0.014	0.030	0.035	0.070
	E North Central	0.017	0.033	0.600	0.040	0.112	0.028	0.043	0.051	0.077
	W North Central	0.010	0.020	0.091	0.537	0.074	0.018	0.075	0.085	0.090
	South Atlantic	0.020	0.049	0.046	0.014	0.709	0.043	0.043	0.028	0.049
	E South Central	0.009	0.019	0.068	0.017	0.189	0.560	0.076	0.024	0.038
	W South Central	0.008	0.016	0.032	0.028	0.078	0.032	0.696	0.047	0.064
	Mountain	0.012	0.021	0.037	0.040	0.062	0.014	0.074	0.562	0.179
	Pacific	0.014	0.021	0.032	0.022	0.061	0.012	0.042	0.091	0.706

Table 5
Taste Parameter Estimates
High School Graduates

Destination Region

Birth Region		New England	Mid-Atlantic	E North Central	W North Central	South Atlantic	E South Central	W South Central	Mountain	Pacific
	New England	0	-0.863	-0.562	-0.899	-0.967	-0.972	-0.865	-0.686	-0.948
			(0.05)	(0.05)	(0.15)	(0.06)	(0.10)	(0.08)	(0.05)	(0.08)
	Mid-Atlantic	-0.803	0	-0.877	-1.092	-0.902	-0.988	-0.921	-0.324	-0.976
		(0.06)		(0.08)	(0.13)	(0.05)	(0.09)	(0.08)	(0.03)	(0.09)
	E North Central	-1.036	-0.895	0	-0.960	-1.203	-1.055	-1.011	-0.685	-1.211
		(0.17)	(0.18)		(0.12)	(0.08)	(0.08)	(0.07)	(0.06)	(0.10)
	W North Central	-0.390	-0.823	-0.631	0	-1.119	-0.830	-0.815	-0.788	-0.899
		(0.05)	(0.07)	(0.06)		(0.11)	(0.13)	(0.10)	(0.05)	(0.09)
	South Atlantic	-0.847	-0.742	-0.727	-1.069	0	-0.704	-0.808	-0.801	-0.528
		(0.08)	(0.08)	(0.05)	(0.09)		(0.06)	(0.06)	(0.07)	(0.08)
	E South Central	-0.405	-0.585	-0.445	-0.433	-0.874	0	-0.816	-0.995	-0.932
		(0.04)	(0.04)	(0.06)	(0.03)	(0.07)		(0.06)	(0.09)	(0.17)
	W South Central	-0.763	-1.231	-0.682	-1.009	-1.108	-0.969	0	-0.699	-1.075
		(0.08)	(0.15)	(0.09)	(0.09)	(0.09)	(0.08)		(0.07)	(0.08)
	Mountain	-0.878	-0.555	-0.644	-0.774	-0.952	-0.774	-0.859	0	-0.784
		(0.15)	(0.05)	(0.05)	(0.10)	(0.07)	(0.11)	(0.06)		(0.05)
	Pacific	-0.710	-0.804	-0.550	-0.806	-0.775	-0.994	-0.648	-0.305	0
		(0.18)	(0.07)	(0.05)	(0.06)	(0.06)	(0.10)	(0.05)	(0.03)	

Table 6
Taste Parameter Estimates
College Graduates

Destination Region

Birth Region	New England	Mid-Atlantic	E North Central	W North Central	South Atlantic	E South Central	W South Central	Mountain	Pacific
New England	0	-0.745	-0.578	-1.361	-0.540	-1.068	-0.492	-0.435	-0.448
		(0.02)	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)
Mid-Atlantic	-0.614	0	-0.602	-0.351	-0.347	-1.004	-0.700	-0.294	-0.622
	(0.01)		(0.01)	(0.01)	(0.01)	(0.04)	(0.02)	(0.01)	(0.01)
E North Central	-0.562	-1.087	0	-0.798	-0.515	-0.932	-0.647	-0.679	-0.341
	(0.02)	(0.03)		(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
W North Central	-0.475	-0.672	-0.420	0	-0.908	-1.238	-0.861	-0.573	-0.647
	(0.01)	(0.02)	(0.01)		(0.02)	(0.05)	(0.02)	(0.01)	(0.02)
South Atlantic	-1.097	-0.975	-0.821	-1.303	0	-0.958	-0.890	-0.858	-0.835
	(0.02)	(0.03)	(0.02)	(0.03)		(0.02)	(0.02)	(0.02)	(0.02)
E South Central	-1.129	-1.063	-0.631	-0.450	-0.513	0	-0.646	-0.924	-0.803
	(0.04)	(0.03)	(0.01)	(0.01)	(0.01)		(0.02)	(0.02)	(0.02)
W South Central	-0.781	-1.330	-0.972	-0.632	-0.630	-0.136	0	-0.823	-0.693
	(0.02)	(0.04)	(0.03)	(0.02)	(0.01)	(0.00)		(0.02)	(0.02)
Mountain	-1.106	-1.186	-0.663	-0.840	-0.542	-0.898	-0.666	0	-0.442
	(0.03)	(0.04)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)		(0.01)
Pacific	-1.253	-1.125	-0.861	-0.809	-0.654	-0.499	-0.761	-0.443	0
	(0.05)	(0.08)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	

Table 7
Log Wages by Education and Region
2000 US Census, 5% IPUMS Random Sample
Raw Data and Corrected for Spatial Selection

	High School				College			
	Median		75th Percentile		Median		75th Percentile	
	Raw Data	Selection Corrected	Raw Data	Selection Corrected	Raw Data	Selection Corrected	Raw Data	Selection Corrected
New England	2.77	2.67	3.03	2.97	3.20	2.87	3.51	3.32
Mid-Atlantic	2.76	2.65	3.04	2.97	3.24	2.76	3.56	3.32
E. North Central	2.74	2.61	3.03	2.96	3.14	2.76	3.44	3.25
W. North Central	2.63	2.50	2.92	2.83	2.98	2.53	3.31	3.05
South Atlantic	2.63	2.55	2.93	2.87	3.09	2.83	3.43	3.3
E. South Central	2.60	2.48	2.91	2.82	3.01	2.59	3.36	3.11
W. South Central	2.60	2.48	2.92	2.84	3.04	2.78	3.39	3.24
Mountain	2.67	2.49	2.96	2.86	3.04	2.63	3.36	3.14
Pacific	2.79	2.63	3.07	2.97	3.22	2.97	3.52	3.4
Average	*2.69*	*2.56*	*2.98*	*2.90*	*3.11*	*2.75*	*3.43*	*3.24*

Table 8
Percentage Returns to College Education

	Median		75th Percentile	
	Raw Data	Selection Corrected	Raw Data	Selection Corrected
New England	0.43	0.20	0.48	0.35
Mid-Atlantic	0.48	0.11	0.52	0.35
E. North Central	0.40	0.15	0.41	0.29
W. North Central	0.35	0.03	0.39	0.22
South Atlantic	0.46	0.28	0.50	0.43
E. South Central	0.41	0.11	0.45	0.29
W. South Central	0.44	0.30	0.47	0.40
Mountain	0.37	0.14	0.40	0.28
Pacific	0.43	0.34	0.45	0.43
Average	*0.42*	*0.18*	*0.45*	*0.34*

Figure 1: High School Graduates
Conditional and Unconditional Log Wage Distributions by Destination Region

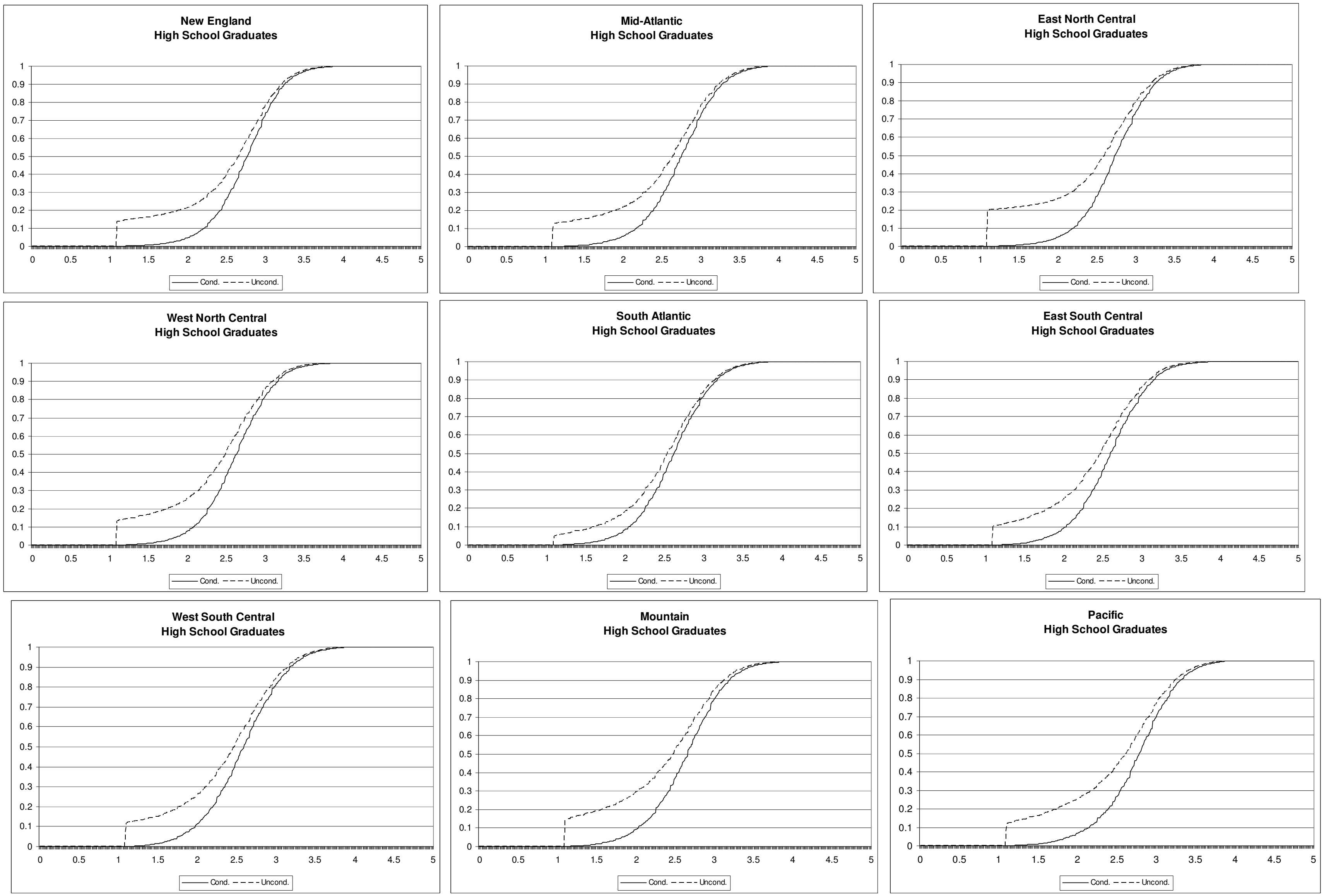

Figure 2: College Graduates
Conditional and Unconditional Log Wage Distributions by Destination Region