

Propensity Score Matching, a Distance-Based Measure of Migration, and the Wage Growth of Young Men^{*}

John C. Ham^a
Xianghong Li^b
Patricia B. Reagan^c

This Draft: December 2004

Abstract

Our analysis of migration differs from previous research in three important aspects. First, we exploit the confidential geocoding in the NLSY79 to obtain a distance-based measure. Second, we let the effect of migration on wage growth differ by schooling level. Third, we use propensity score matching to measure the effect of migration on the wages of those who move. We develop an economic model and use it to (i) assess the appropriateness of matching as an econometric method for studying migration and (ii) choose the conditioning variables used in the matching procedure. Our data set provides a rich array of variables on which to match. We find a significant effect of migration on the wage growth of college graduates of 10 percent, and a marginally significant effect for high school dropouts of -12 percent. If we use either a measure of migration based on moving across county lines or state lines, the significant effects of migration for college graduates and dropouts disappear.

JEL Classification Code: J6, J3, C4.

Keywords: Propensity score matching, distance-based migration, wage growth.

^{*}Corresponding author: John C. Ham; email: ham.25@osu.edu. An earlier version of this paper was presented under the title "Matching and Selection Estimates of the Effect of Migration on Wages for Young Men." We would like to thank Dwayne Benjamin, Stephen Cosslett, Songnian Chen, Xiaohong Chen, William Greene, Cheng Hsiao, Guido Imbens, John Kennan, Lung-fei Lee, Audrey Light, Geert Ridder, Aloysius Siow, Jeffrey Smith, Petra Todd, Insan Tunali, Bruce Weinberg, Tiemen Woutersen, James Walker and Jeff Yankow for very helpful comments and discussions. We are also grateful to seminar participants at Arizona, CEMFI, the Federal Reserve Bank of New York, McGill/University of Montreal, McMaster, Minnesota (Industrial Relations), Ohio State, Pompeu Fabra, Rutgers, Toronto, UC Berkeley, UC Davis, UC San Diego, USC, Vanderbilt, Western Ontario, Wisconsin and the Upjohn Institute, as well as at the Econometric Society and Society of Labor Economics meetings. Eileen Kopchik and Yong Yu provided outstanding programming help. This research was partially supported by the NSF and NICHD. We are responsible for all errors. This paper also reflects the views of the authors and not necessarily the views of the Federal Reserve Bank of New York or the Federal Reserve System. The paper in no way represents the views of the NSF or NICHD.

^a Department of Economics and CHRR, Ohio State University, the Federal Reserve Bank of New York, and IZA.

^b Department of Economics, York University.

^c Department of Economics and CHRR, Ohio State University.

I. Introduction

Internal migration is an important economic phenomenon in the United States. Between 2002 and 2003, about 40.1 million Americans moved, and about 60 percent of these movers were 20 to 29 years old.¹ Labor economists typically model migration as an investment in human capital, and a natural question to ask is what the return to this investment is in terms of higher wages. While this issue has received some attention, previous empirical research has focused more on the causes of moving rather than on the consequences. Most migration studies find that factors such as age, education, job tenure, wage on the current job, skills, family composition, length of residence in the current location, local amenities, and the local cost of living affect the migration decision. However, evidence on whether moving increases wages is mixed. By using data on young men from the 1979-1996 waves of the National Longitudinal Surveys of Youth 1979 (NLSY79), we attempt in this paper to identify the average individual *contemporaneous* wage gain from U.S. internal migration for those who move.

We contribute to migration research in several aspects. First, we allow migration effects to differ across education groups and find that this distinction is important.² Previous studies pool different education groups to estimate average returns for all migrants. If returns to migration are positive for some education group(s), such as college graduates, and negative or zero for other groups, then the overall sample average may be statistically indistinguishable from zero. We find a significant positive migration effect for college graduates of around 10 percent. We also find a marginally significant negative effect for high school dropouts of about -12 percent. For the overall sample and the other educational groups, we do not find a significant migration effect.

Second, we use a distance-based measure of migration instead of a measure based on moving across a state or county line. By exploiting the confidential geocoding of the data by the Center for Human Resource Research at the Ohio State University, we can obtain the exact latitude and longitude of the respondent's residence at the time of each interview. This allows us to calculate a distance-based measure of migration. Compared to a measure based on moving across a state or county line, the measures commonly used in the literature, the distance-based measure of migration corresponds more closely to the theoretical notion of changing local labor markets described by Hanushek (1973). We find that measuring migration by changing state underestimates migration by about 36%, and measuring migration by changing county overestimates migration by about 43%. Further, we are not able to find a significant migration

¹ See <http://www.census.gov/prod/2004pubs/p20-549.pdf>

² To the best of our knowledge, Yankow (1999) is the only other author who allows migration effects to differ by education.

effect for college graduates or high school dropouts by using these alternative measures of moving across state or county.

Third, selection bias can be a severe problem in migration studies. Further, we argue below that investigating the effect of migration on those who move is at least as interesting as examining the unconditional migration effect. For both these reasons we use propensity score matching method (Rosenbaum and Rubin 1983) to address this problem. We use a theoretical model to ascertain the appropriateness of matching for our empirical problem. From an empirical perspective, matching requires a strong ignorable treatment assignment assumption, which may not seem completely reasonable in many applications. From a theoretical perspective, if an individual knows more about his wage on his new job than the researcher, the condition necessary for matching to be valid will not hold. However, we use our theoretical model to show that as a practical matter, if matching and differencing do a good job of eliminating the permanent component of wages, the violation of ignorable treatment assignment is likely to be relatively minor. Thus, many economists are likely to consider matching as a reasonable empirical strategy. This theoretical model also helps us determine the appropriate variables on which to base the matching, and the NLSY79 data provide a rich set of such variables. Our results are not sensitive to the propensity score model specification, bandwidth choice and trimming level, and they pass balancing tests and a specification test.

Potentially of use to other applied researchers are the following findings. First, there is no advantage in going to higher order polynomials than the local linear regression in the matching procedure, and there may well be a cost in terms of overparamaterizing the model. Second, a variable bandwidth works well for us, and the results are not sensitive to the size of the bandwidth. Third, the Andrews-Buchinsky (2000, 2001) procedure for choosing the number of bootstrap repetitions is quite helpful, and in one case suggests a higher number of repetitions than the number often used by applied researchers. In the case of this estimator, using too small a number of bootstrap repetitions gives misleading results.

Our paper is organized as follows. We review the migration literature in Section II. In Section III we review the matching literature and present our econometric model. In Section IV we use our theoretical model to examine the assumption of ignorable treatment assignment that underlies all matching studies. The next two sections describe our data and our empirical results respectively. Section VII concludes the paper.

II. Literature Review

The most common theoretical model of migration treats the decision to migrate as an investment in human capital: individuals migrate if the present value of real income in a destination minus the cost of moving exceeds what could be earned at the place of origin (Sjaastad 1962).³ The empirical studies based on this model can be classified into two broad areas, for our purposes: those examining the determinants of migration and those focused on the consequences of migration for wages and earnings.⁴ While the determinants of migration are not the focus of our paper, they play a crucial role in our estimation of the propensity score. Polachek and Horvath (1977) and Plane (1993) find that migration propensities vary over a person's life cycle. Geographic mobility peaks during the early to mid-twenties and declines with age thereafter because the time horizon over which gains from migration can be realized grows shorter. These studies also find that the propensity to migrate increases with education. Highly educated workers operate in labor markets that compete across broad geographic areas, whereas workers with low levels of education operate in more geographically isolated labor markets. Workers with more education also may be better informed about opportunities outside their local labor market and better able to evaluate that information.

In addition, the migration decision is affected by migration cost and the non-wage benefits of different locations. Goss and Schoening (1984) provide some indirect evidence that households with fewer assets are less mobile, since they find that the probability of migration declines with the duration of unemployment. Lansing and Mueller (1967) report that many moves are attributable to family related issues, such as proximity to family members or health considerations. Presumably being close to one's family is a non-wage advantage of a given location.

Of course, in the human capital model of migration, expected wage gains, local demand shocks, and inter-regional differences in returns to skill play an important role in the migration decision. Shaw (1991), Borjas, Bronars, and Trejo (1992b), Dahl (2002), and Kennan and Walker (2003) use a Roy model of comparative advantage to explain migration. Although the human capital model of migration clearly predicts a higher present value of lifetime earnings for those who migrate, the literature on the consequences of migration reaches no consensus on the contemporaneous returns to migration. Estimates of the average contemporaneous returns can be

³ See also McCall and McCall (1987). They develop a "multi-armed bandit" approach to the migration decision. Workers rank locations by their pecuniary and nonpecuniary attributes, and then sample locations sequentially until a suitable match is found. Search costs limit the number of markets sampled.

⁴ Greenwood (1997) provides an excellent review of the literature.

negative, zero, or positive. Positive contemporaneous returns are found by Bartel (1979) for younger workers, Hunt and Kau (1985) for repeat migrants, and Gabriel and Schmitz (1995) and Yankow (2003) for less-educated workers. Negative contemporaneous returns are found by Polachek and Horvath (1977), Borjas, Bronars, and Trejo (1992a), and Tunali (2000).⁵ Studies that find statistically insignificant contemporaneous returns include Bartel (1979) for older workers, Hunt and Kau (1985) for one-time migrants, and Yankow (2003) for workers with more than a high school degree.

The sign and significance of the migration effect depend on the sample chosen and on how researchers address three critical questions. First, what definition of migration is used? Although all authors have in mind a migration as a change of labor market, most define migration as occurring if a geographic boundary is traversed. The majority of authors, including most of those cited above, focus on interstate migration. A few, such as Hunt and Kau (1985) and Gabriel and Schmitz (1995), define migration as a change of Metropolitan Statistical Area (MSA). Falaris (1987) defines it as a change of Census region. Finally, some authors, such as Linneman and Graves (1983), study inter-county migration. By comparing alternate definitions of migration, we show later that migration counts are highly sensitive to the definition used and that the estimated returns to migration are also sensitive to the definition of migration.⁶

The second question affecting the estimated effect of migration concerns the choice of comparison group. Most authors use all workers who do not migrate as the comparison group. But it is well known that there is wage growth associated with voluntary job turnover (Topel and Ward 1992). Since most migrants change jobs, the “return to migration” may confound returns to job changing with a return to geographic mobility. Bartel (1979) was the first to focus on the relationship between the types of job separation and migration. Others, such as Yankow (1999), condition on job changing but do not differentiate between types of job turnover. Finally, Raphael and Riker (1999) consider only workers who were laid off.

Third, what is the treatment of sample selection? Because migration is a choice variable and not randomly assigned, there is no reason to presume that migrants constitute a random sample of all workers. Nakosteen and Zimmer (1980, 1982) were among the first to provide evidence of positive self-selection into migration. Robinson and Tomes (1982) and Gabriel and

⁵ A negative return is not necessarily inconsistent with utility maximization, since a high growth rate can overcome a negative contemporaneous effect. Alternatively, Tunali (2000) views migration as a lottery and finds that while a substantial portion of migrants experience wage reductions after moving, a minority realize very high returns. Individuals are willing to invest in an activity that has a high probability of yielding negative returns because of the potential for a very large payoff.

⁶ The distance-based measure that we use is also used by Baumann and Reagan (2002) to study mobility in Appalachia.

Schmitz (1995) also find favorable self-selection. However, Hunt and Kau (1985) and Borjas, Bronars, and Trejo (1992a) find no evidence of self-selection.

Note that the sample selection studies in migration attempt to estimate an unconditional effect of moving or an average treatment effect. We see two problems with this. First, it may not be interesting to ask the effect on wages of moving to a new location for a randomly chosen individual. Many individuals will already be in a location that gives them relatively high wages, so we could easily expect this treatment effect to be negative. Second, the choice of a new location is ambiguous – is it the individual’s best alternative or a randomly chosen location? In this paper we use matching to look at a less ambitious, but arguably better-specified question: the effect on wage growth for those who move. Note that there is no ambiguity here as to what the respective locations are in interpreting this effect.⁷

III. Econometric Model

3.1 Estimating the Effect of Moving

Our goal is to use propensity score matching to estimate the effect of internal migration on between-job wage growth for those who quit their first job and move.⁸ Following the notation in the evaluation literature, let $D = 1$ if an individual moves and $D = 0$ otherwise. We then define the outcome for movers ($D = 1$) as Y_1 and the outcome for stayers ($D = 0$) as Y_0 . As will be discussed in Section IV, we use difference-in-difference matching, so the outcome is the logarithm of the starting wage on the second job minus the logarithm of the ending wage on the first job for each individual. Our goal is to identify the average treatment effect on the treated (i.e., the effect of migration on those who migrate).

$$\Delta = E(Y_1 - Y_0 \parallel D = 1) = E(Y_1 \parallel D = 1) - E(Y_0 \parallel D = 1). \quad (3.1)$$

⁷ See Heckman, LaLonde and Smith (1999) for a discussion of when estimating the effect of “treatment on the treated” may be more useful than estimating an average treatment effect.

⁸ The seminal paper in the matching literature is Rosenbaum and Rubin (1983). For recent contributions in the economics literature see Abadie and Imbens (2002), Hahn (1998), Heckman, Ichimura, and Todd (1997, 1998), Heckman, Ichimura, Smith and Todd (1998), and Hirano, Imbens and Ridder (2003). See Imbens (2004) and Heckman, LaLonde and Smith (1999) for a discussion of previous work in the statistics literature and empirical applications of matching such as Lechner (2000). See Frölich (2004) and Zhao (2004) for Monte Carlo evaluation of various matching approaches.

We observe the first term on the right-hand side of equation 3.1. However, we do not observe the second term on the right-hand side (i.e., the wage gain movers would have experienced had they not moved). We will use matching to estimate $E(Y_0 \| D = 1)$. However, for matching to be valid, certain assumptions must hold. The fundamental assumption underlying matching estimators is *ignorable treatment assignment* (ITA) (Rosenbaum and Rubin 1983) or *selection on observables* (Heckman and Robb 1985). This assumption is represented by

$$(Y_0, Y_1) \perp D \| X^*, \quad (3.2A)$$

where X^* is a vector of variables that are unaffected by the treatment. This assumption states that, conditional on a set of observables X^* , the respective treatment outcome is independent of actual treatment status. In empirical work X^* usually contains pretreatment variables and time-invariant individual characteristics.

Since we are estimating the average treatment effect on the treated, condition (3.2A) can be weakened to the following mean independence assumption involving only Y_0

$$E(Y_0 \| X^*, D) = E(Y_0 \| X^*). \quad (3.2B)$$

In the next section we use a theoretical model and the variables available to us to argue that our rich dataset makes this assumption plausible and thus matching is a suitable estimation approach for our problem.

To identify the treatment effect on the treated, matching also requires that

$$\Pr(D = 1 \| X^*) < 1. \quad (3.3)$$

This common support condition requires that at each level of X^* , the probability of observing nonparticipants is positive. (This condition can be enforced by adding a common support constraint, as discussed in Section 3.6.)⁹

⁹ Another implicit assumption required by the matching estimator is the stable unit-treatment value assumption (SUTVA) (Rosenbaum and Rubin 1983). It says that the outcome of unit i given treatment is independent of the outcome of unit j given treatment. To satisfy this assumption, we have to ignore

Matching on all variables in X^* becomes impractical as the number of variables increases. To overcome this curse of dimensionality, Rosenbaum and Rubin (1983) propose propensity score matching, which reduces a multidimensional matching problem to a one-dimensional problem. Specifically, instead of matching on a vector X^* , we match on an index function $P(X^*)$. $P(X^*)$ is the propensity score (i.e., the probability of moving conditional on X^*), where

$$P(X^*) = \Pr(D = 1 \parallel X^*). \quad (3.4)$$

Rosenbaum and Rubin show that if the conditions in equations 3.2A and 3.3 are satisfied, then

$$(Y_0, Y_1) \perp D \parallel P(X^*) \quad (3.5A)$$

and

$$\Pr(D = 1 \parallel P(X^*)) < 1. \quad (3.6)$$

It is straightforward to show that given 3.2B and 3.3, we have

$$E(Y_0 \parallel P(X^*), D) = E(Y_0 \parallel P(X^*)) \quad (3.5B)$$

and

$$\Pr(D = 1 \parallel P(X^*)) < 1. \quad (3.6)$$

If ITA or the mean independence assumption holds given X^* , it also holds conditional on $P(X^*)$. Of course, we must choose what variables to include in X^* . We defer this issue to Section IV, where we use an economic model to guide our choice. Fortunately, the NLSY79 is a rich data set offering many possible variables.

We should note that it is also possible to calculate an unconditional effect of moving. We do not do so for two reasons. First, as we mentioned above, we believe the conditional effect for those who move is more interesting. Second, calculating an unconditional effect involves using matching to estimate the wage gain stayers would have experienced had they moved. This, in our

general equilibrium effects. Since our 378 migrants are from a random nationwide sample, SUTVA is reasonable in our problem.

case, requires matching a small number of movers to a large number of stayers. Not surprisingly, the corresponding standard errors are so large as to render this calculation uninformative.

3.2 Choice of Matching Method

We now discuss the issue of which propensity score matching estimator to use. Let N_1 be the number of movers and N_0 be the number of stayers. The outcomes for the two groups can be written as $Y_1 = \{Y_{1i}\}_{i=1}^{N_1}$ and $Y_0 = \{Y_{0j}\}_{j=1}^{N_0}$ respectively. Consider member i of the mover group. The simplest method of matching is to use nearest neighbor matching (with replacement). Here we approximate $E(Y_{0i} \parallel D=1)$ using Y_{0j} , the outcome for the member j of the stayer group whose propensity score $\hat{P}(X_j^*)$ is closest to $\hat{P}(X_i^*)$.

Nearest neighbor matching, although intuitively appealing, is inefficient: it uses only one observation in the comparison group to estimate the potential outcome for a treated observation. For this reason we do not use nearest neighbor matching, and instead consider more efficient matching methods. Heckman, Ichimura, and Todd (1997, 1998), and Heckman, Ichimura, Smith, and Todd (1998) incorporate local regression into matching. For each observation i ($i = 1, \dots, N_1$) in the treatment group, local regression matching opens a window around $\hat{P}(X_i^*)$ and uses all observations in the comparison group with propensity scores in that window to construct a weighted mean $\hat{m}(\hat{P}(X_i^*))$ to approximate $E(Y_{0i} \mid D=1)$. Within the window, the closer $\hat{P}(X_j^*)$ is to $\hat{P}(X_i^*)$, the greater the weight the observation j gets in estimating $\hat{m}(\hat{P}(X_i^*))$.

To formally define local regression, suppose we observe two paired vectors (w_j, z_j) , where $j = 1$ to n . At each point of interest, W_0 , local regression estimates $m(W_0)$ by solving the following minimization problem:

$$\min_{\alpha_0, \beta_0^1, \dots, \beta_0^M} \sum_{j=1}^n \left\{ z_j - \alpha_0 - \sum_{l=1}^M \beta_0^l (w_j - W_0)^l \right\}^2 K \left(\frac{w_j - W_0}{h(W_0)} \right). \quad (3.7)$$

where $K(\cdot)$ is a kernel function and $h(W_0)$ is the bandwidth. In our case the bandwidth varies with W_0 , as will be discussed later. This minimization problem yields $\hat{m}(W_0) = \hat{\alpha}_0$.

Applying local regression to our study, we let $(w_j, z_j) = (\hat{P}(X_j^*), Y_{0j})$. For each mover i ($i = 1, N_1$), we run a local regression at the estimated propensity score $\hat{P}(X_i^*)$ and estimate $\hat{m}(\hat{P}(X_i^*))$. Of course, to implement this procedure we must choose M , the highest order of the polynomial. Generally, the larger M is, the smaller will be the asymptotic bias but the larger will be the asymptotic variance. Fan and Gijbels (1996) prove that asymptotically a choice of $M = q$, where q is an odd number, dominates a choice of $M = q - 1$. The intuition is that moving from $q - 1$ to q introduces an extra parameter, reducing the asymptotic bias (especially in boundary regions and highly clustered regions). There is no corresponding increase, however, in the asymptotic variance. (Their result implies that kernel regression is asymptotically dominated by local linear regression.) Fan and Gijbels (1996) also point out that in practice the typical optimal choice is usually $M = 1$ and occasionally $M = 3$. Thus, their work suggests that we should use in our problem a local linear regression or possibly a local cubic regression.

However, the above discussion does not consider the finite sample behavior of the estimators. Frölich (2004) investigates finite-sample performance of matching estimators including kernel regression ($M = 0$) and local linear regression ($M = 1$). He concludes that kernel regression is more robust to misspecification in the bandwidth than local linear regression. Two aspects of Frölich's results are worth noting. First, his results are based on the use of a global bandwidth, and local linear estimators have a well-known problem over regions of sparse data with such a bandwidth. One solution is to use a variable or locally adaptive bandwidth (Fan and Gijbels 1996). We use this approach as discussed immediately below. Second, in Frölich's results, the quality of local linear regression depends on the sample size of the treatment group compared to the sample size of the comparison group. Frölich's results suggest that local linear matching performs reasonably well when the comparison group is large relative to the treatment group (a ratio of the comparison group to the treatment group on the order of 5 to 1). Our data include 1700 stayers and 378 movers, and thus we expect that our local linear regression matching estimator should perform reasonably well.

Finally, in calculating the average migration effect for all movers we only match individual j to individual i if individual j is in individual educational group. Rosenbaum and

Rubin (1983) define such a procedure as finer balancing. Following Rosenbaum and Rubin (1985), we first estimate the propensity score using the entire sample and then match movers with stayers in the same educational group based on the estimated propensity score.

3.3 Choice of the Bandwidth Parameter

The choice of a bandwidth or smoothing parameter is often the most important decision a researcher makes in nonparametric regression. There is a trade-off in choosing the bandwidth: the smaller the bandwidth, the smaller the bias, but the larger the bandwidth, the smaller the variance. Basically, there are two types of bandwidths: global (fixed) bandwidths and local (variable) bandwidths. The global bandwidth approach uses the same window width at each point W_0 , while the variable bandwidth approach changes the bandwidth according to the data density around W_0 . In other words, the variable bandwidth approach allows us to use a small bandwidth where the probability mass is dense and a larger bandwidth where the probability mass is sparse. As Fan and Gijbels (1992, p. 2013) put it, “A different amount of smoothing is used at different data locations.”

Fan and Gijbels (1992) suggest that it is advantageous to combine local regression with variable bandwidth. We use a simple adaptive variable bandwidth proposed by Fan and Gijbels (1996). In their procedure the size of the window $h_{k_n}(W_0)$ varies by the point W_0 ; $h_{k_n}(W_0)$ is chosen to include the same number of data points k_n closest to each W_0 to fit the local regression. The number k_n is determined by the sample size n . Essentially, we want k_n to become larger as the sample size grows but not too quickly.¹⁰ Our variable bandwidth is compatible with the suggestion of Silverman (1986) and others that one should use a subjective bandwidth choice.¹¹

¹⁰Fan and Gijbels (1996, theorem 4.2) prove that if $k_n \rightarrow \infty$ such that $k_n/n \rightarrow 0$ and $k_n/\log n \rightarrow \infty$, then the adaptive variable bandwidth h_{k_n} behaves asymptotically as $k/\{nf(w)\}$, where k is the number of the nearest neighbors, $f(w)$ is the density function of w_j , $j=1, n$ and n is the sample size. This bandwidth choice bears some resemblance to the k -nearest neighbor estimates of Härdle (section 3.2, 1990). However, Härdle’s estimator puts equal weight on all neighbors, while in our case the weight depends on how close the neighbor is to W_0 .

¹¹ Ruppert, Sheather, and Wand (1995) derive three optimal fixed (global) bandwidth selectors for local linear regression. We considered their preferred selector, the direct plug-in bandwidth selector (p. 1262), but it performed poorly in terms of producing matching estimates with large standard errors.

3.4 The Sampling Variance of the Matching Estimator

We follow the previous literature and use the bootstrap method to obtain standard errors for the matching estimators. An important decision is the choice of the number of bootstrap repetitions. We follow the procedure developed in Andrews and Buchinsky (2000, 2001). They propose a three-step method for choosing the number of bootstrap repetitions, pointing out that the number of bootstrap repetitions chosen by most empirical studies is usually less than needed. We find that commonly used numbers of bootstrap repetitions are much too small for the local cubic regression matching estimator, and that inference is affected by using an inappropriate number of repetitions. We describe the Andrews and Buchinsky procedure for calculating standard errors in Appendix A.

3.5 Allowing the Treatment Effect to Differ by Educational Group

The migration effect may depend on the level of schooling. This would occur, for example, if it is much easier for college graduates to search for a higher wage and find a job in a new location without moving there than it is for other educational groups. Let S denote schooling class and s denote a particular schooling level. Therefore, we now estimate

$$\Delta_s = E(Y_1 - Y_0 \parallel D = 1, S = s) = E(Y_1 \parallel D = 1, S = s) - E(Y_0 \parallel D = 1, S = s) \quad (3.8)$$

To obtain the first term in equation 3.8, we take the mean increase in wages for those in schooling class s who move. To obtain the second term, we again use matching and only match individual j to individual i if individual j is in individual i 's schooling class.¹² In our empirical work below, we find that it is important to allow the treatment effect to vary by educational group.

¹² In other words, the summation $\sum_{j=1}^n [\cdot]$ in (3.7) becomes $\sum_{j \in s} [\cdot]$.

3.6 Common Support Constraint, Balancing Tests and a Specification Test

The matching parameter is identified only over the portion of X 's support where each mover can find reasonable number of stayers in its neighborhood. To satisfy the condition in equation 3.3, we add a common support constraint, following the procedure proposed by Heckman, Ichimura, and Todd (1997).¹³ If the target trimming level is q , their procedure will trim between q percent and $2q$ percent of participants. The exact trimming level depends on the data structure; the closer the modes and shapes of the two distributions are, the closer the actual trimming is to q percent.¹⁴ Since this procedure trims participants only, it will not cause an extra boundary problem (in the context of local regression) when we estimate the treatment effect on the treated. Following previous work, we set $q = 5$. To test the sensitivity of our matching estimators to the trimming level, we also consider $q = 3$ and $q = 7$ for our baseline model. We find that our results are insensitive to the choice of trimming level.

For our model to be correctly specified, the conditioning variables X^* should be distributed identically across the treatment group and the matching sample. If they are, the propensity score balances the sample. We test whether this is satisfied for nearest neighbor matching via two types of tests, paired t -tests and joint F tests.¹⁵

Paired t -tests examine whether the mean of each element of X^* for the treatment group is equal to that for the matched sample. However, these tests are not able to detect differences between two distributions beyond the sample means. Since all matching methods require that the two distributions mimic each other at each quantile, instead of just exhibiting similar means, we also use a joint F test. The treatment group and matched sample are broken down into quartiles according to the estimated propensity scores.¹⁶ At each quartile, we test whether all elements of X^* are *jointly* different across the two groups. If a model fails to pass either the t -tests or the F

¹³ See also Smith and Todd (forthcoming) for details.

¹⁴ In the earlier version of this paper (Ham, Li, and Reagan 2001), we proposed a trimming procedure that will eliminate exactly q percent of the sample. Since our method trims both the participants and nonparticipants, it could exacerbate the boundary problem when estimating the treatment effect on the treated. For our sample, our procedure produced results that are very similar to those produced by the Heckman, Ichimura, and Todd procedure described above.

¹⁵ To conduct the balancing tests, we do not need to calculate standard errors for the nearest neighbor estimates. Abadie and Imbens (2004) show that one cannot use the bootstrap to calculate these standard errors. Abadie and Imbens (2002) provide an alternative means of calculating standard errors for nearest neighbor matching.

¹⁶ The number of intervals used in joint F tests depends on sample size. We have 378 movers, so we can only afford to break them down into quartiles. If larger samples are available, finer intervals, such as deciles, should be used.

tests, we add higher order terms or interaction terms until the variables are balanced across the two groups.

As a specification test, we examine whether migration has a significant “treatment effect” on annual wage growth on the first job by educational category. The idea here is twofold. First, wage growth on the first job is the pretreatment variable that is closest to our variable of interest: between-job wage growth by schooling level. Second, since this variable is pretreatment, any significant “treatment effect” for this variable can only reflect selection bias that matching fails to correct. This test is similar to that proposed for matching methods by Smith and Todd (forthcoming), except that they conduct the test over the entire sample, while we test it for each educational group.¹⁷

IV. An Economic Model of Migration, the Ignorable Treatment Assignment Assumption and Choice of Conditioning Variables

In this section we develop an optimizing model of migration for two reasons. First, it will aid us in determining the appropriateness of the assumption of Ignorable Treatment Assignment (ITA) for our empirical problem. Second, it will help us to choose the appropriate conditioning variables. One may feel that ITA is too strong an assumption to hold in a real-world empirical application. Alternatively, one may note that ITA will not hold exactly for our problem, since as Heckman, LaLonde and Smith (1999) discuss, it will not hold in a simple optimizing model if there are variables determining the returns that are observable to the individual making the investment decision but are unobservable to the econometrician.¹⁸ However, we would argue that for most empirical problems, the question is not whether a condition holds exactly, but rather whether the deviation from the condition is likely to be big or small. Thus we evaluate our economic model for several sensible sets of parameter values to investigate how serious we would expect the deviation from ITA to be. For some of these parameter values, i.e. when differencing and matching do a good job of removing the permanent effect in the outcome equation and the migration equation, we would expect the deviation to be relatively small. Ours is the first matching paper to undertake such an exercise, and this should be useful to other researchers since other decisions, such as participation in job training or college attendance, are consistent with a very similar economic model and researchers also use matching to analyze the effects of training and college attendance.

¹⁷ Note that we do not include education times annual wage growth on the first job in the propensity score, so this is not simply a balancing test.

¹⁸ See also Imbens (2004).

4.1 An Economic Model of Migration

We modify the Willis and Rosen (1979) model of education and apply it to the problem of migration choice. At the beginning of the period, all workers have quit their first job. They face a choice between accepting another job locally or moving to another labor market and accepting a job there. We assume that moving involves time costs and pecuniary costs. We also assume that switching jobs locally or in the other market involves search costs. To simplify the notation, we suppress the individual subscripts in this section. Net expected future earnings from changing jobs locally and across markets are $N_k = V_k - C_k$, $k = c, m$, where c denotes the initial labor market and m denotes the labor market to which an individual migrates. Further, V_k is the discounted present value of earnings in location k and C_k is the cost of changing jobs in k . We assume that the utility of location k equals

$$U_k = [\exp(Z_{1k}\eta_k + u_{1k})] \cdot N_k, \quad (4.1)$$

where Z_{1k} and u_{1k} reflect observed and unobserved variables, respectively, that indicate the value of living in k (e.g. amenities, proximity to one's family, childhood friends, spouse etc.), holding constant the net present value of income in the location. Workers choose to migrate if

$$[\exp(Z_{1m}\eta_m + u_{1m})] \cdot N_m > [\exp(Z_{1c}\eta_c + u_{1c})] \cdot N_c. \quad (4.2)$$

Denote the labor market-specific starting wages and wage growth rates as y_k^s and g_k , $k = c, m$, respectively. If the individual takes the local job, the wage at time t is

$$y_c(t) = y_c^s e^{g_c t}. \quad (4.3)$$

Assuming it takes M units of time to move, the wage at time t in the new location is

$$\begin{aligned} y_m(t) &= 0 && \text{for } t \leq M \\ &= y_m^s e^{g_m(t-M)} && \text{for } t > M. \end{aligned} \quad (4.4)$$

We make the following assumptions. First, workers face an infinite horizon. Second, the individual-specific discount rate, r , is constant over time, where $r > \max(g_c, g_m)$. Finally, the costs of changing jobs locally or across labor markets enter the net utility function exponentially. Under these assumptions, the net utility of changing jobs locally can be written as

$$U_c = [\exp(Z_{1c}\eta_c + u_{1c})] \left(\frac{y_c^S}{(r - g_c)} \right) [\exp(-Z_{2c}\lambda_c - u_{2c})], \quad (4.5)$$

where Z_{2c} and u_{2c} are observed and unobserved variables that reflect the costs of changing jobs locally and λ_c is a vector of weights. The net utility of changing jobs across labor markets can be written as

$$U_m = [\exp(Z_{1m}\eta_m + u_{1m})] \left(\frac{y_m^S}{(r - g_m)} \right) [\exp(-rM - Z_{2m}\lambda_m - u_{2m})], \quad (4.6)$$

where Z_{2m} , u_{2m} and λ_m are defined analogously to (4.5).

We define the migration decision equation as $I = \ln(U_m / U_c)$. An individual chooses to move only when $I > 0$. Substituting from equations (4.5) and (4.6) and taking a Taylor series approximation around the population mean values of $(\bar{g}_c, \bar{g}_m, \bar{r})$ yields

$$I = \alpha_0 + \ln y_m^S - \ln y_c^S + \alpha_1 g_m - \alpha_2 g_c + \alpha_3 r + Z_{1m}\eta_m - Z_{1c}\eta_c - (Z_{2m}\lambda_m - Z_{2c}\lambda_c) + u_{1m} - u_{1c} - (u_{2m} - u_{2c}) \quad (4.7)$$

where $\alpha_1 = 1/(\bar{r} - \bar{g}_m) > 0$, $\alpha_2 = 1/(\bar{r} - \bar{g}_c) > 0$, $\alpha_3 = -M - (\bar{g}_m - \bar{g}_c)/[(\bar{r} - \bar{g}_c)(\bar{r} - \bar{g}_m)]$.

We define the starting wage on the second job y_k^S , the wage growth g_k , and the discount rate r_k as follows. The equation for the starting wage on the new job is given by

$$\ln y_k^S = Z_{3k}\gamma_{3k} + u_{3k}, \quad k = c, m. \quad (4.8)$$

In (4.8) Z_{3k} and u_{3k} are observed and unobserved variables (to the researcher), respectively, which affect starting wages on the new job.¹⁹ To make it easier to keep track of the error decompositions that follow, we scale the following three equations by their respective coefficients in the migration equation (4.7). The growth rate of wages for those who move scaled by α_1 is given by

$$\alpha_1 g_m = Z_{4m} \gamma_{4m} + u_{4m}. \quad (4.9a)$$

The specific growth rate of wages for those who do not move scaled by α_2 is given by

$$\alpha_2 g_c = Z_{4c} \gamma_{4c} + u_{4c}. \quad (4.9b)$$

In (4.9a) and (4.9b) Z_{4k} and u_{4k} are observed and unobserved variables, respectively, which affect wage growth on the new job, and γ_{4k} is a vector of returns to Z_{4k} . Finally, the worker's scaled discount rate, r , is a function of family background variables Z_5 and an error term

$$\alpha_3 r = Z_5 \delta + u_5. \quad (4.10)$$

4.2 Error Decomposition

To facilitate our discussion of the appropriateness of ITA assumption, we decompose each of the five error terms defined in Section 4.1 into transitory and permanent components.

We assume

$$u_{jk} = \phi_{jk} \mu + e_{jk}, \quad j = 1, 4, k = c, m \quad (4.11a)$$

and

$$u_5 = \phi_5 \mu + e_5, \quad (4.11b)$$

where the ' e ' terms represent transitory errors that are specific to each equation and independent both of each other and of μ . The term μ is the common component across the five error terms, reflecting permanent factors such as the individual's personality, persistence, stability, etc. The interpretations of the idiosyncratic error terms vary across the equations. For example, the error

¹⁹ In reality there will be a component of the future wage which is unobservable to the individual. In an earlier draft we included such a forecast error, but omit it here to simplify the analysis.

term in the starting wage in the local labor market, e_{3c} , represents the (idiosyncratic) industry characteristics and size of firm in which the worker anticipates finding a job if he has not yet located a job locally. If, on the other hand, he has located a job locally, e_{3c} represents the idiosyncratic features of the job. Our goal in adopting this error structure is to capture the important features of the correlation between the errors from the migration decision equation and the outcome equation while keeping the structure simple enough to allow us to obtain interesting results.

4.3 Selection and Outcome Equations

Substituting equations (4.8) through (4.11b) into (4.7) and collecting terms yields the migration decision rule

$$I = X_1\theta_1 + \beta\mu + \left\{ \sum_{j=1,3,4} (e_{jm} - e_{jc}) - (e_{2m} - e_{2c}) + e_5 \right\} > 0, \quad (4.12)$$

where $\beta = \left\{ \sum_{j=1,3,4} (\phi_{jm} - \phi_{jc}) - (\phi_{2m} - \phi_{2c}) + \phi_5 \right\}$ and X_1 contains the unique elements of the Z_{jk} , $j=1,4$, $k=c,m$, and Z_5 . As will be discussed later, to come closer to the ITA condition, we use difference-in-difference (DID) matching, in which the outcome variable of interest is the starting wage on the second job minus the ending wage on the first job. We assume that the ending wage on the first job is determined by

$$\ln y^E = Z_6\gamma_6 + u_6, \quad (4.13)$$

where $u_6 = \phi_6\mu + e_6$.

Since the mean independence assumption 3.2B involves Y_0 only, we will focus on the outcome variable for the stayers in the following discussion regarding the identification assumption. For cross-sectional matching, the outcome is defined as in equation (4.8); for DID matching, the outcome is defined as

$$\ln y_c^S - \ln y^E = Z_{3c}\gamma_{3c} - Z_6\gamma_6 + (\phi_{3c} - \phi_6)\mu + e_{3c} - e_6. \quad (4.14)$$

We show in Section 4.4 that the DID approach helps to satisfy the ITA assumption.

4.4 Ignorable Treatment Assignment Assumption

In this section we focus on DID matching. The extension to cross-section matching is straightforward and omitted to save space. As a practical matter, in assessing the appropriateness of ITA we focus on the correlation between the selection equation (4.12) and the outcome equation (4.14).²⁰ The numerator of this term consists of the covariance between these error terms, and is equal to $(\phi_{3c} - \phi_6)\beta\sigma_\mu^2 - \sigma_{3c}^2$, where σ_μ^2 and σ_{3c}^2 are the variances of μ and e_{3c} respectively. Thus two factors contribute to this numerator: the permanent component μ and the transitory component e_{3c} . A sufficient condition for this correlation to be zero is that differencing completely removes the permanent component of the error in (4.14) and there are no transitory variables affecting the outcome that the individual knows but are unobservable to the econometrician. We expect that differencing diminishes the role of the permanent component, i.e. $|\phi_{3c} - \phi_6| < |\phi_{3c}|$, but does not allow us to ignore it. To come closer to ITA, we need to condition the migration decision and the outcome variable on additional variables X_2 that are correlated with the common component μ to reduce its effect. The migration decision equation now is given by

$$\begin{aligned} I &= X_1\theta_1 + X_2\theta_2 + \tilde{\varepsilon} > 0 \text{ or} \\ I &= X^*\theta^* + \tilde{\varepsilon} > 0, \end{aligned} \tag{4.15a}$$

where $\tilde{\varepsilon} = \beta\tilde{\mu} + \left\{ \sum_{j=1,3,4} (e_{jm} - e_{jc}) - (e_{2m} - e_{2c}) + e_5 \right\}$, and we have assumed for simplicity that conditioning on X_2 has not affected β or the idiosyncratic error terms. The outcome equation becomes²¹

²⁰ Choosing a different model specification will affect not only the covariance between the two error terms but also the variances. We focus on the correlation coefficient to have a comparable measure across specifications.

²¹ Here we abstract from the fact that X^* only contains variables, in addition to X_2 , that affect both migration and the wage growth.

$$\ln y_c^S - \ln y^E = X^* \gamma^* + \tilde{u}, \quad (4.15b)$$

where

$$\tilde{u} = (\phi_{3c} - \phi_6) \tilde{\mu} + e_{3c} - e_6.$$

In this case, the correlation coefficient between the error terms in (4.15a) and (4.15b) is

$$\rho_{MD} = \frac{(\phi_{3c} - \phi_6) \beta \sigma_{\tilde{\mu}}^2 - \sigma_{3c}^2}{\sqrt{(\beta^2 \sigma_{\tilde{\mu}}^2 + \sum_{j=1}^4 \sum_{k=c}^m \sigma_{jk}^2 + \sigma_5^2) \cdot \sqrt{(\phi_{3c} - \phi_6)^2 \sigma_{\tilde{\mu}}^2 + \sigma_{3c}^2 + \sigma_6^2}}} \quad (4.16)$$

where ρ_{MD} represents the correlation coefficient after matching on additional variables X_2 and differencing, and $\sigma_{\tilde{\mu}}^2$ is the variance of the new permanent component $\tilde{\mu}$. In an ideal world this correlation coefficient would be zero, but as a practical matter we would like it to be small. In other words, the larger the correlation, the less reasonable is the ITA assumption. Further, as a benchmark we also consider the correlation between the migration and outcome equations when we neither match nor difference, i.e. the correlation between (4.8) and (4.12)

$$\rho = \frac{\phi_{3c} \beta \sigma_{\mu}^2 - \sigma_{3c}^2}{\sqrt{(\beta^2 \sigma_{\mu}^2 + \sum_{j=1}^4 \sum_{k=c}^m \sigma_{jk}^2 + \sigma_5^2) \cdot \sqrt{\phi_{3c}^2 \sigma_{\mu}^2 + \sigma_{3c}^2 + \sigma_6^2}}} \quad (4.17)$$

We consider the evaluation of (4.16) and (4.17) for several sets of parameter values. We first normalize $E(e_{jk}^2) = E(e_5^2) = E(e_6^2) = \sigma_e^2 = 1$ and $\beta = \phi_{3c} = 1$. Given the normalizations, there are three crucial parameters that affect the magnitude of the correlation: i) σ_{μ}^2 - how large is the variance of the permanent component; ii) $\sigma_{\tilde{\mu}}^2$ - how large is the variance of the permanent component after conditioning on X_2 ; and iii) ϕ_6 - how large is the loading factor on the permanent component on the ending wage on job 1, i.e. how effective is differencing.²² We only

²² Changes in the scale of β can be subsumed into σ_{μ}^2 . There is the question of whether β should be positive or negative. We have focused on the case where $\beta > 0$ for the following reason. If those with lower discount rates (i.e. those that come from wealthier families) have unobserved factors that make their wages

consider the case where $\phi_6 < \phi_{3c}$, since given our normalizations, it implies that the variance in the starting wage for job 2 conditional on the job 1 variables is greater than the variance in the ending wage for job 1 conditional on the job 1 variables, and we would expect this latter condition to hold.²³

In Table 1 we select six sensible sets of values for these three influential parameters to show the roles of matching and differencing in achieving ITA. The first three columns contain the selected values for σ_μ^2 , $\sigma_{\tilde{\mu}}^2$ and ϕ_6 , and the last two columns present the values of ρ_{MD} (from 4.16) and ρ (from 4.17) corresponding to each set of parameters. Row 1 of Table 1 is our base case, where $\sigma_\mu^2 = 10$, $\sigma_{\tilde{\mu}}^2 = 5$ and $\phi_6 = .75$. We consider this case as reasonable since the variance of the permanent component is 10 times the variance of the idiosyncratic components, differencing effectively reduces the loading factor on the permanent component in the outcome variable by three-quarters, and conditioning on X_2 reduces the variance of the permanent component by one-half. All other cases represent different deviations from the base case, and we put in bold the parameter values that differ from the base case in the respective row. In our base case the argument for ITA looks quite good since $\rho_{MD} = 0.044$ while $\rho = 0.596$. The latter correlation indicates that this is a case where if we neither difference nor match on X_2 , we would expect selection to be important. In row 2 we show the effect of reducing the influence of the permanent component by cutting σ_μ^2 and $\sigma_{\tilde{\mu}}^2$ in half. Now $\rho_{MD} = -0.075$ from (4.16) while $\rho = 0.404$ from (4.17). Next, in row 3 we double σ_μ^2 and $\sigma_{\tilde{\mu}}^2$, and as a result $\rho_{MD} = 0.212$ and $\rho = 0.752$. In row 4 we show the result of less effective differencing by reducing ϕ_6 to 0.5 from our base case, which results in $\rho_{MD} = 0.222$.²⁴ In line 5 we show the effect of less effective matching by raising $\sigma_{\tilde{\mu}}^2$ to 7.5, estimating $\rho_{MD} = 0.137$. Finally, in line 6 we show the effect of less effective differencing and less effective matching by simultaneously reducing ϕ_6 to 0.5 and

higher and make α_3 negative (as is likely to be the case), then $\phi_5 > 0$. Since $\beta = \left\{ \sum_{j=1,3,4} (\phi_{jm} - \phi_{jc}) - (\phi_{2m} - \phi_{2c}) + \phi_5 \right\}$, it is the sum of a number of differences (which are likely to be small) plus a positive parameter, and thus we think $\beta > 0$ is reasonable. We should note that $\beta < 0$ would increase the absolute value of the correlation coefficients in (4.16) and (4.17).

²³ While we think the assumption that $\phi_6 < \phi_{3c}$ is reasonable, it is not innocuous. If we change the base case in line 1 of Table 1 by increasing ϕ_6 from .75 to 1.25, ρ_{MD} becomes -0.316 .

²⁴ In lines 4 through 6 ρ is unchanged from our base case.

raising σ_{μ}^2 to 7.5. Now $\rho_{MD} = 0.344$, which would leave many researchers reluctant to act as if ITA holds. If matching and differencing are quite effective, as in line 1, most economists would consider ITA a reasonable assumption. If matching and differencing are both less effective, as in line 6, ITA does not seem to be a reasonable assumption. (Lines 3 and 4 contain intermediate cases on which economists may differ.)²⁵ The upshot is that an economic model, *a priori*, does not rule out the use of matching as an econometric tool.

[Insert Table 1 here]

4.5 Choice of Conditioning Variables

The above model suggests that we should use all variables in the propensity score that are correlated with both the variables that an individual uses to make his migration decision *and* the outcome equation. Recall that X_2 contains variables that are not usually included in a reduced-form migration equation that may be correlated with the permanent component in the migration decision and outcome equation. We use the following variables in X_2 : the beginning wage on the first job, the ending wage on the first job, tenure of the first job and a dummy variable indicating home ownership while on job 1. Our overall conditioning variables in the matching procedure consist of the unique elements in X_1 (the variables usually thought to enter a migration decision) and X_2 . Variables such as age, education, professional status, marital status, race and living in an MSA²⁶ will directly affect wages and thus affect the migration decision. We would expect that home ownership would affect moving costs and would be correlated with the unobservables in wages. We would expect the wealth of the individual's parents to affect the discount rate and the migration decision, and we use the education of the individual's father to proxy family wealth. Whether this variable enters the wage equation or is correlated with the unobservables in the wage equation is an open question.²⁷ We include father's education in X_1 but experiment with excluding it from the propensity score.

²⁵ In an earlier version we had a forecast error in the realized wage if the individual stays, but omitted it here to simplify matters. If we put this error in the model, ρ_{MD} falls to 0.181 and 0.194 respectively in lines 3 and 4.

²⁶ Previous studies show that workers in cities earn more than their nonurban counterparts after controlling for earning capability. Glaeser and Mare (2001) suggest that the urban wage premium comes from living in the city, not from innate characteristics associated with urban residence.

²⁷ Willis and Rosen (1979) assume that father's education does not enter the wage equation, nor is it correlated with the error in the wage equation. However, others may find this assumption too strong.

We would argue that for our approach to be credible, there must be factors that affect the migration decision but not the outcome variable. Excluding the variables that only affect migration from the propensity score will help identification, since individuals with very similar propensity scores based on (4.15a) self-select into both the movers' and stayers' groups because they bear different moving costs.²⁸ Further, omitting them from the propensity score will not introduce bias. In our economic model, these are variables such as proximity to friends, families and spouses that only shift the location-specific present value of income in the utility function in (4.1); variables that only affect the cost of moving in (4.5) and especially (4.6), such as the distance of the move; and variables that only affect the discount rate in (4.10). For example, we would expect individuals who still lived in the county where they were born at age 14 to receive higher utility from staying, holding the net present value of income constant. We would not expect this variable to affect wages.

4.6 Comparison to Other Matching Studies

Finally, it is interesting to compare our approach to that used in the literature, which investigates matching as a means of estimating the effect of training on wages for those who undertake training e.g. Heckman, Ichimura, Smith and Todd (1998), Dehejia and Wahba (1999). The economic model behind these training studies is clearly very similar to our economic model, where individuals undertake training if it pays for them to do so. The matching literature concerning training considers the effect of training on those who undertake it (i.e. the treatment effect on the treated), which is comparable to our investigation of the gain to migration among those who migrate. Comparable to our differencing in (4.14), the training literature often works with the difference in post-training and pre-training wages as the variable of interest (e.g. Smith and Todd forthcoming). This literature tends to use lagged values of earnings and unemployment as the relevant conditioning variables to achieve ITA, analogous to our use of the work history and home ownership variables.

²⁸ Heckman and Navarro-Lozano (2004) stress the role that these type of variables play in identification.

V. Data Description

Our primary data source is the 1979-1996 waves of the NLSY79. The survey began in 1979 with a sample of 12,686 men and women born between 1957 and 1964. Annual interviews were conducted from 1979 to 1994, with biennial interviews thereafter.

The NLSY79 provides a comprehensive data set ideally suited for studying migration and job mobility together. First, the longitudinal aspects of the data make it possible to track the same individuals over time as they move across jobs and labor markets. Furthermore, the NLSY79 data files include detailed longitudinal records of the employment history of each respondent. Second, the confidential geocoding of the data allows us to obtain the exact latitude and longitude of the respondent's residence at the time of each interview. This, in turn, allows us to calculate a distance-based measure of migration and compare our results with more orthodox measures based on change of county or change of state. Our distance-based measure of migration corresponds more closely to the theoretical notion of changing local labor markets than do the alternative measures. A change-of-county definition of migration misclassifies as migrants individuals who move short distances across county lines but do not change labor markets. A change-of-state definition of migration misclassifies as stayers individuals who move hundreds of miles and change labor markets but remain in the same state.²⁹ Finally, the data focus on individuals at the outset of their work careers, a stage that exhibits the greatest moving and job changing.

In order to construct a sample suitable for empirical analysis, we introduce several selection criteria. The sample is limited to young men since the moving decisions of women are more complicated. Because our interest lies in postschooling labor market activity, we follow individuals from the time that they leave school. The longitudinal structure of the NLSY79 allows us to determine precisely when most workers make a permanent transition into the labor force. Conceptually, we define the working career as beginning the first time a respondent leaves formal schooling. To avoid counting summer breaks or other inter-term vacations as leaving school, we define a schooling exit as the beginning of the first non-enrollment spell lasting at least 12 consecutive months. Accordingly, respondents are excluded from the sample if the date of schooling exit cannot be clearly ascertained from the data. For example, respondents who are continuously enrolled throughout the observation period or who have incomplete or inconsistent schooling information are excluded from the sample.

²⁹ Kennan and Walker (2003) use a change-of-state definition of migration in their structural model. It would not be feasible for them to consider our definition of migration.

Of the 6,403 male respondents in the initial sample, 262 were deleted because they never left school, or because an exact date of school leaving could not be determined. Further, 576 individuals were deleted because they did not hold at least two civilian jobs. Another 116 were lost because they did not report hours and wages on at least two civilian jobs. We eliminated 50 respondents because they reported being fired from their jobs. We imposed this restriction because we wanted to concentrate on voluntary job transitions. Nine respondents were deleted because they were not interviewed during the duration of at least two civilian jobs.

We required that the respondent had at least two jobs that lasted at least 26 weeks, which resulted in the loss of 32 respondents. We also required that the respondent hold at least two jobs with average hours of at least 25 per week. This last restriction did not result in the loss of any respondents. To avoid extreme measurement error we required that the respondent report wages on at least two jobs of between \$1 and \$50 in 1990 dollars. This restriction resulted in the deletion of 120 respondents from the sample. Because we use a distance-based measure of migration, we required respondents to have valid residential location data at the time they report holding their two jobs. The requirement of valid location data for at least two jobs resulted in the loss of 1479 respondents. This is the largest single reason for sample deletion and reflects the difficulty of geocoding addresses between 1979 and 1989, when address information was sparser.

We deleted jobs that overlapped for more than 8 weeks. In this case we considered the respondent to be holding two jobs simultaneously and did not treat that as a job transition. For this reason we lost 132 respondents. We lost another 465 respondents who held two jobs satisfying all of the above criteria, but who reported an intervening job without location data. In this case we did not observe two consecutive jobs. Further, we lost 855 respondents who satisfied all of the above criteria but who experienced an interval of more than 13 weeks between two jobs. The between-job interval consisted of either a spell of unemployment, nonemployment or employment in part time jobs (defined as those with average hours less than 25 or lasting less than 26 weeks).³⁰ Finally we lost 229 respondents because information on the variables used in this analysis was lacking. Our final sample consists of 2078 male respondents.

To recap, we explore migration conditional on voluntarily quitting the first job. Our movers consist of men who quit their first job and move to a new location while our stayers consist of those who quit their first job but do not move. In this study, migration was defined to

³⁰ It is not possible to determine whether the unemployment is correlated with migration. The problem is that we can determine the dates of employment changes but only observe location at the interview dates. Thus someone who is in a new location at the time of the interview and has been unemployed for 6 months could have been unemployed for 6 months in the previous location and have just moved, or they could have been unemployed for 6 months in the new location.

have occurred if the respondent moved at least 50 miles or changed Metropolitan Statistical Area (MSA) and moved at least 20 miles.³¹ We focus on real wage growth between the first two jobs. We call the jobs “job 1” and “job 2” hereafter.

Table 2 presents descriptive statistics. The third column provides means for the whole sample, and the fourth and fifth columns show means for movers and stayers respectively. The last column presents the difference in means between movers and stayers. Over 18 percent of all voluntary job changes involved migration.

[Insert Table 2 here]

The individual characteristics shown in Table 2 are reported as of the end of job 1. The men in the sample are, on average, 26 years old at the time of the job change. The movers are slightly younger. African Americans are more likely than non-Hispanic whites to be stayers, while Hispanics are equally represented in both groups. On average, the movers have higher education and are more likely to be married. Prior to migration, movers are less likely to own a house or live in an MSA.

The NLSY79 provides detailed information on each respondent’s job history and on characteristics of each job, a feature that makes matching an appealing strategy for estimating the migration effect. On job 1, movers on average have higher starting and ending wages, have slightly longer tenure, and are more likely to have a professional job. Between job 1 and job 2, movers and stayers experience, on average, roughly the same wage gain (around 9 percent).

A characteristic such as whether the father has a college degree are likely to affect resources available to finance a move, and movers are more likely to report that their father had a college degree. As a proxy for ties to the local community, we use a variable that measures whether the respondent was residing at age 14 in the same county in which he was born. Not surprisingly, stayers are more likely than movers to have lived in their birth county at age 14.

Table 2 presents systematic differences between movers and stayers. Thus, there is reason to suspect, a priori, that selection will be a serious problem that must be addressed to estimate the effect of migration on the real wage growth for those who move.

³¹ Adjacent county centroids are typically about 25 miles apart, so a move of 50 miles roughly corresponds to a move two counties away.

VI. Empirical Results

6.1 Propensity Score Models

Table 3 reports probit estimates for the three models of the propensity score for the migration decision described in section IV. All models contain variables representing demographics, characteristics of job 1, and home ownership. The models differ in their inclusion of the father's education and same-county variables. From the literature it is not clear whether father's education affects only the resources to finance a move, in which case it does not belong in the propensity score, or it is also a proxy for unobserved earning ability, in which case it does belong. We believe that the same-county variable only represents psychic costs of moving, and its inclusion should not affect the final estimates.

Model I, our baseline model, contains the core variables and father's education. Model II contains only the core variables and omits father's education. Model III also contains father's education, the same-county variable and an interaction between the same-county variable and the professional occupation variable. (The interaction term is added to achieve balance between the movers and the matched sample.) By comparing the matching estimates of returns to migration in Models I and II, we investigate whether father's education affects only the resources to finance a move, since if it also affects the outcome variable, we would expect the results from the two models to differ. By comparing the matching estimates from Models I and III, we test the robustness of Model I to the inclusion of a variable that should not affect estimated returns to migration.

The demographic variables have the expected signs in all three models. Consistent with most migration studies, our results show that the probability of migration starts to decline at about age 25. Hispanics are more likely and African Americans less likely to move than are non-Hispanic whites, although the coefficient on the Hispanic dummy is statistically insignificant. Individuals with less schooling than a college degree are less likely to move than are those with a college degree. Married men are more likely to move than are unmarried men. Individuals residing in an MSA when they quit their first job are less likely to migrate than are those living in non-metropolitan areas. Men in professional occupations on job 1 are more likely to migrate. Homeownership has a negative and statistically significant effect on migration. The three work history variables (starting wage, ending wage and tenure of job 1) are not significant individually, but a likelihood ratio test shows they are jointly significant. On average, movers have higher hourly wages prior to migration than do stayers.

[Insert Table 3 here]

Although respondents whose fathers have a college education are more likely to move (see Model II), a comparison between Models I and II shows that none of the other coefficients are sensitive to inclusion of father's education. The Model III results indicate that living at age 14 in the county of birth reduces the probability of migration. The interaction between the professional occupation variable and the same-county variable, which is included only for purposes of balancing, alters only the coefficient on professional occupation. With this one exception, the coefficients are stable across the three models.

We show the distributions of the estimated propensity score for Model I in Figure 1. The top panel is a histogram plot for the movers and the bottom panel is the plot for the stayers. Most applications of matching on job training programs have shown that propensity score distributions for the treatment and comparison groups are very different in terms of the mode and empirical support. This poses a strong challenge for matching. In our case, the movers, on average, have a higher probability of migration than stayers, but the empirical support of the two distributions is very similar and the modes are quite close.

[Insert Figure 1 here]

6.2 Balancing and Specification Tests

Table 4 shows the results of the balancing tests of the three models. Panel A shows the paired t -statistics for the difference in the variable mean between movers and the matched sample of stayers. Panel B presents the joint F statistics for the difference in the means of all variables at each quartile of the propensity score. All the tests are conducted using the mover sample and the matched sample from nearest neighbor matching. We first discuss the t -tests in Panel A. Under all three models, the conditioning variables are well balanced.³² Matching does a good job with regard to pre-migration variables such as race, professional job dummy, and past wages that differ considerably between movers and stayers (see Table 2). The joint F tests in Panel B demonstrate

³² For Model I we have shown the balancing statistic for the same-county variable even though it is *not* included in the model. For Model II we have shown the balancing statistics for father's education and the same-county variable, even though these variables are *not* included in the model. These results basically confirm the probit results that both these variables affect the migration decision.

that the conditioning variables are well balanced jointly at each quartile of the estimated propensity score.

[Insert Table 4 here]

Table 5 presents our specification test. As discussed in Section 3.6, we examined the “treatment effect” of moving on annual wage growth on job 1 by educational category. Recall that since this variable is pre-treatment, any significant “treatment effect” can only reflect selection bias that matching fails to correct. The test statistics in Table 5 are based on local linear regression matching. None of these “effects” is significantly different from zero.³³

[Insert Table 5 here]

6.3 Estimates of the Migration Effects from the Baseline Model

Table 6 presents the matching estimates of the effect of migration for movers on wage growth from the baseline propensity score model (Model I). For all three estimators in Table 6, Panel A (with a $q = 5$ trimming level), we conduct 200, 300, and 1,100 bootstrap repetitions to illustrate the importance of choosing a sufficiently large number of repetitions in calculating standard errors. In Appendix A we present an algorithm for choosing the minimum required number of bootstrap repetitions based on the three-step method of Andrews and Buchinsky (2000, 2001). The minimum numbers are 248 and 1,074 for the local linear and local cubic estimators respectively.³⁴ Most of the literature uses at most 200 repetitions. For the local linear estimates, the standard errors from 200 repetitions are relatively close to those from 300 or 1,100 repetitions because 200 repetitions are not significantly less than the required minimum of 248. However, for the local cubic estimator, the standard errors from 200 or 300 repetitions are dramatically underestimated. For high school dropouts, the estimated standard error increases threefold when we increase the number of repetitions from 200 to 1,100. The large standard errors produced by

³³ The specification test does not reject the null hypothesis that the model is correctly specified when we use the local cubic estimator. We do not include these results given the problems we report below with this estimator. We use 350 repetitions, which is greater than the minimum required by the Andrews-Buchinsky method, to calculate standard errors.

³⁴ For each estimator, we calculate the minimum repetitions required for the overall effect. We then calculate the minimum repetitions for each education group separately. Finally, we take the maximum of the five numbers as our required number of repetitions.

local cubic matching indicate the problem of overparameterization (Fan and Gijbels, 1996). This problem is masked when standard errors are calculated using only 200 repetitions.

In what follows we focus solely on the local linear estimates with standard errors calculated from 300 bootstrap repetitions, which is more than the number of replications required by the Andrews-Buchinsky method. A 25% bandwidth gives us a wide enough window when we disaggregate the data by educational class.³⁵ When we do not disaggregate by education level, there is a quite small, and statistically insignificant, effect of migration. When we disaggregate by education, the effect of migration for high school dropouts is estimated to be -12%. This estimate is significantly different from zero at the 10 percent significance level but not at the 5 percent level. College graduates who migrate experience 10% greater wage growth, and this estimate is statistically significant at the 5 percent level. There is no statistically significant difference in wage growth from migration for job changers who have only a high school education or some college. Local cubic regression matching produces very large standard errors, which, as noted above, indicates the problem of overparameterization.

[Insert Table 6 here]

There are three issues worth pointing out with respect to the negative estimated effect for high school dropouts. First, we estimate a contemporaneous effect on wage growth of migration. Insignificant or negative contemporaneous effects do not necessarily imply that migration is an irrational decision from the perspective of the human capital approach. As noted in Section II, some previous studies have found that positive returns to migration often are not realized until five or six years after the original migration, and that the initial returns are negative. It is interesting to note that some of the previous studies found negative returns for the entire sample, while we find them only for high school dropouts. Migration may involve an assimilation process. A short-term loss in wage need not, and probably does not, imply a drop in life-time utility. In terms of the model in Section IV, the lifetime utility increases for migrants if the growth rate effect dominates a negative or zero initial wage gain. Of course, it may be the case that the model is not appropriate for dropouts. They could be insufficiently skilled to solve the

³⁵ To implement finer balancing matching, we first choose a variable bandwidth to give us a comparison group equal to 25% of the stayers. We then use only those in the group who are in the same educational category as the mover in question. Each mover gets far less than 25% of stayers in the local regression. We find that our results are not sensitive to a 1%-2% bandwidth change.

optimization problem, even approximately. Alternatively, they may not be able to see wages in the other location without visiting it.³⁶

Second, unlike most migration studies, our study estimates a migration effect that has netted out the effect of job changing, and thus our results do not imply that any group experiences a negative return to job changing. Third, it is possible that return and repeat migration are driving the negative returns for high school dropouts, and we do observe more repeat and return migration for high school dropouts than for other education categories.³⁷ To explore this possibility, we excluded those with repeat or return migration from the mover sample. This modification, however, did not change the negative migration effect for high school dropouts or the positive effect for college graduates.

Panels B and C of Table 6 present the estimates based on alternative trimming levels of $q = 3$ and $q = 7$ respectively. (We drop the local cubic estimator given its poor performance in Panel A.) The estimates are not sensitive to this change in the trimming level, except for a two-percentage point difference in the return to college graduates between $q = 3$ and $q = 5$. This difference may reflect the widespread finding in the matching literature that the right tail of the distribution of returns is more sensitive to the trimming level than are other parts of the distribution.

6.4 Robustness of the Treatment Effects to the Propensity Score Specification

Table 7 represents the migration effects estimated from the three alternative propensity score models. Panel A contains the estimates from our baseline model, repeated from Table 6, for 300 bootstrap repetitions and serves as a benchmark. Panel B reports estimates from Model II, in which we exclude from the propensity score the variable indicating whether the father has a college degree. If this variable affects the outcome variable as well as the migration decision, we would expect the results of Models I and II to differ. The estimated effects are almost identical under Models I and II. Thus it appears that father's education significantly affects the moving decision but does not provide extra information with regard to unobserved earning ability, after controlling for all the other individual characteristics and the lagged variables.

[Insert Table 7 here]

³⁶As one seminar participant put it, "College graduates can search and then move, while dropouts must move before they can search."

³⁷Return migration within two years is 36% for dropouts and 24% for the overall sample. Repeat migration within two years is 36% for dropouts and 22% for the whole sample.

Panel C of Table 7 reports estimates from Model III, in which we add the same-county variable. Again the estimates are very close to those from Models I and II. These results suggest that living at age 14 in the birth county is a migration cost variable and does not affect wage growth. In summary, the results from the two alternative models suggest that we have a well-specified propensity score model that is robust to alternative specifications.

6.5 Alternative Definitions of Migration

Our distance-based measure of migration made possible by the confidential geocoding of the NLSY79 data corresponds more closely to a change of labor markets than do two alternative definitions of migration commonly used in the literature: changing state of residence and changing county of residence. Table 8 presents summary statistics for all three definitions. The first column of Panel A shows the number of movers and stayers in our sample under each definition. The number of people who are considered movers differs substantially according to the definition used. There are only 258 movers (out of 2,078 job changers) who are movers when a move is defined as crossing a state line. In contrast, there are 542 movers when a move is defined as crossing a county line. The distance-based measure produces an intermediate number of movers (378).

Panel A of Table 8 also shows the average, minimum and maximum distances between consecutive locations for those classified as movers and stayers under each definition. The average distance for movers ranges from 379 miles under the change-of-county definition to 722 miles under the change-of-state definition. The average distance for movers under the distance-based measure is 535 miles. The average distances, however, mask the potential for misclassification inherent in the other two definitions.

Under the distance-based measure, the minimum distance between consecutive locations for movers is 20 miles, conditional on changing residence from one MSA to another. The maximum distance for stayers is 49 miles, conditional on not changing MSA. However, under the change-of-state definition of migration, the minimum distance for movers is 1 mile, and the maximum distance between consecutive locations for stayers is 668 miles. When a change-of-county definition is used, the minimum distance for movers is 1 mile and the maximum distance for stayers is 38 miles. Both the change-of-state definition and change-of-county definition incorrectly classify as movers those making short-distance changes in residence across a boundary. The change-of-state definition also incorrectly classifies as stayers individuals who make large-distance changes in residences.

Panel B of Table 8 shows the magnitude of the potential for misclassification of movers and stayers using definitions of migration based on crossing a state or county boundary. Row 1 describes individuals who are classified as movers under a distance-based measure but are classified as stayers under the change-of-state definition. These 136 individuals (36% of all movers under the distance-based measure) have an average distance of 120 miles between consecutive locations and a maximum distance of 668 miles. Row 2 describes individuals who are classified as stayers under the distance-based measure but are classified as movers under the change-of-state measure. These 16 individuals (less than .5% of all stayers under the distance-based measure) have an average distance between consecutive locations of 16 miles, a minimum distance of 1 mile and a maximum distance of 44 miles. The last row describes individuals who are classified as stayers under the distance-based measure but are classified as movers under the change-of-county definition.³⁸ These 164 individuals (10% of stayers under the distance-based measure) have an average distance between consecutive locations of 17 miles and a minimum distance of 1 mile.

[Insert Table 8 here]

In Table 9 we re-estimate all stages of the matching model using the two alternative definitions of migration. We do so to investigate the potential impact of misclassification, as described in Table 8, on the matching estimates. All results are based on Model I, our baseline model, with $q = 5$ trimming. Compared to our distance-based measure of migration, the alternative definitions yield smaller (in absolute value) and statistically insignificant estimates of the effect of migration on wage growth for dropouts and college graduates. None of the estimated effects is significant at even the 10 percent level. Our results raise the question of how previous estimated returns to migration in studies using different methodologies would change with a distance-based measure of migration.

[Insert Table 9 here]

³⁸ Not surprisingly, the change-of-county definition does not classify as stayers any movers under the distance-based measure.

VII. Conclusion

Our paper estimates the effect of U.S. internal migration for movers who quit their first job on real wage growth between the ending wage on their first job and the starting wage on their second job. Our analysis of migration differs from previous research in three important ways. First, we exploit the confidential geocoding in the National Longitudinal Surveys of Youth 1979 to obtain a distance-based measure of migration rather than defining migration as a movement across county lines or state lines. Second, we let the effect of migration on wage growth between the first and second jobs differ by schooling level. Third, we use propensity score matching to address selection issues and estimate the effect of migration on the wage growth of young men who move. We use an economic model to assess the practical validity of the assumption of Ignorable Treatment Assignment (which underlies all matching studies). We find that although one would expect ITA to be violated in principle, as long as matching and differencing do a relatively good job of eliminating the permanent component of the error in the outcome equation, as a practical matter this violation should be relatively minor and thus matching is a reasonable empirical strategy. The economic model also helps us choose which variables should be included in the propensity score. Matching is a “data hungry” estimation strategy, and our data set provides a rich array of variables on which to match. Specifically, we use variables on previous labor market history, family background, demographics, and homeownership.

We find a significant positive effect of migration on the wage growth of college graduates, and a marginally significant negative effect for high school dropouts. We do not find any significant effect for other educational groups or for the overall sample. Our results are robust to changes in the model specification. Our models pass balancing tests and a specification test. We find that better data matters; if we use a measure of migration based on moving across county lines or state lines, the significant effects of migration on the wage growth of college graduates and dropouts disappear. Finally, we provide useful information to applied researchers on the highest order of the polynomial when using local regression in the matching procedure, and on the number of bootstrap repetitions when calculating standard errors.

There are at least two avenues for future research. First, we could look at individuals five years after they quit their first jobs to measure the effect of migration on the wage growth after relocation. Second, we could consider migration effects for young women.

References

- Abadie, A. and Imbens, G. (2002), "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects," NBER Technical Working Paper no. 283.
- Abadie, A. and Imbens, G. (2004), "On the Failure of the Bootstrap for Matching Estimators," mimeo, October.
- Andrews, D. W. K. and Buchinsky, M. (2000), "A Three-Step Method for Choosing the Number of Bootstrap Repetitions," *Econometrica*, **67**, 23-51.
- Andrews, D. W. K. and Buchinsky, M. (2001), "Evaluation of a Three-Step Method for Choosing the Number of Bootstrap Repetitions," *Journal of Econometrics*, **103**, 345-386.
- Bartel, A. P. (1979), "The Migration Decision: What Role Does Job Mobility Play?" *American Economic Review*, **69**, 775-786.
- Baumann, R. W. and Reagan, P. B. (2002), "Employment and Earnings of Economically Disadvantaged Whites: Understanding Why Appalachians Are Particularly Disadvantaged" (Mimeo, Ohio State University).
- Borjas, G. J., Bronars, S. G. and Trejo, S. J. (1992a), "Assimilation and the Earnings of Young Internal Migrants," *The Review of Economics and Statistics*, **74**, 170-175.
- Borjas, G. J., Bronars, S. G. and Trejo, S. J. (1992b), "Self-Selection and Internal Migration in the United States," *Journal of Urban Economics*, **32**, 159-185.
- Dahl, G. B. (2002), "Mobility and the Returns to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, **70**, 2367-2420.
- Dehejia, R. H., and Wahba, S. (1999), "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, **94**, 1053-1062.
- Falaris, E. M. (1987), "A Nested Logit Migration Model with Selectivity," *International Economic Review*, **28**, 429-444.
- Fan, J. (1992), "Design Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, **87**, 998-1004.
- Fan, J. and Gijbels, I. (1992), "Variable Bandwidth and Local Regression Smoothers," *Annals of Statistics*, **20**, 2008-2036.
- Fan, J. and Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, Monographs on Statistics and Applied Probability 66 (London: Chapman & Hall)
- Frölich, M. (2004), "Finite Sample Properties of Propensity-Score Matching and Weighting Estimators," *Review of Economics and Statistics*, **86**, 77-90.

- Gabriel, P. E. and Schmitz, S. (1995), "Favorable Self-Selection and the Internal Migration of Young White Males in the United States," *Journal of Human Resources*, **30**, Summer, 460-471.
- Glaeser, L. E. and Mare, D. C. (2001), "Cities and Skills," *Journal of Labor Economics*, **19**, 316-342.
- Goss, E. P., and Schoening, N. C. (1984), "Search Time, Unemployment and the Migration Decision," *Journal of Human Resources*, **19**, 570-579.
- Greenwood, M. J. (1997) "Internal Migration in Developed Countries," *Handbook of Population and Family Economics*, vol. 1B, 648-720 (Amsterdam, New York: Elsevier).
- Hahn, J. (1998), "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, **66**, 315-331.
- Ham, J. C., Li, X. and Reagan, P. B. (2001) "Matching and Selection Estimates of the Effect of Migration on Wages for Young Men" (Mimeo, Ohio State University).
- Hanushek, E. A. (1973), "Regional Differences in the Structure of Earnings," *Review of Economics and Statistics*, **55**, 204-213.
- Härdle, W. (1990), *Applied Nonparametric Regression*, Econometric Society Monographs No. 19 (Cambridge: Cambridge University Press).
- Heckman, J., Ichimura, H. and Todd, P. (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies*, **64**, 605-654.
- Heckman, J., Ichimura, H. and Todd, P. (1998), "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, **65**, 261-294.
- Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998), "Characterizing Selection Bias Using Experimental Data," *Econometrica*, **66**, 1017-1098.
- Heckman, J., LaLonde R. and Smith, J.(1999), "The Economics and Econometrics of Active Labor Market Programs," in O. Ashenfelter and D. Card (eds.), *Handbook of the Labor Economics* (Amsterdam, New York: Elsevier).
- Heckman, J. and Navarro-Lozano, S. (2004), "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models," *Review of Economics and Statistics*, **86**, 30-57.
- Heckman, J. and Robb, R. (1985), "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*, Econometric Society Monographs series, no. 10 (Cambridge: Cambridge University Press).
- Hirano, K., Imbens, G. and Ridder, G. (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, **71**, 1161-1189.

- Hunt, J. C. and Kau, J. B. (1985), "Migration and Wage Growth: A Human Capital Approach," *Southern Economic Journal*, **51**, 697-710.
- Imbens, G. (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *Review of Economics and Statistics*, **86**, 4-29.
- Kennan, J. and Walker, J. (2003), "The Effect of Expected Income on Individual Migration Decisions" (Mimeo, University of Wisconsin).
- Lansing, J. B., and Mueller, E. (1967), "The Geographic Mobility of Labor," Survey Research Center, Ann Arbor, MI.
- Lechner, M. (2000), "An Evaluation of Public-Sector-Sponsored Continuous Vocational Training Programs in East Germany," *Journal of Human Resources*, Spring 2000, 347-375.
- Linneman, P. and Graves, P. E. (1983), "Migration and Job Change: A Multinomial Logit Approach," *Journal of Urban Economics*, **14**, 263-279.
- McCall, B. P. and McCall, J. J. (1987), "A Sequential Study of Migration and Job Search," *Journal of Labor Economics*, **5**, 452-476.
- Nakosteen, R. A. and Zimmer, M. A. (1980), "Migration and Income: The Question of Self-Selection," *Southern Economic Journal*, **46**, 840-851.
- Nakosteen, R. A. and Zimmer, M. A. (1982), "The Effects on Earnings of Interregional and Interindustry Migration," *Journal of Regional Science*, **22**, 325-341.
- Plane, D. A. (1993), "Demographic Influences on Migration," *Regional Studies*, **27**, 375-383.
- Polachek, S. W. and Horvath, F. W. (1977), "A Life Cycle Approach to Migration: Analysis of the Perspicacious Peregrinator," in Ehrenberg, R. G. (ed.), *Research in Labor Economics*, 103-149.
- Raphael, S. and Riker, D. A. (1999), "Geographic Mobility, Race, and Wage Differentials," *Journal of Urban Economics*, **45**, 17-46.
- Robinson, C. and Tomes, N. (1982) "Self-Selection and Interprovincial Migration in Canada," *Canadian Journal of Economics*, **15**, 474-502.
- Rosenbaum, P. R. and Rubin, D. B. (1983), "The Central Role of the Propensity Score in Observational Studies for Casual Effects," *Biometrika*, **70**, 41-55.
- Rosenbaum, P. R. and Rubin, D. B. (1985), "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *The American Statistician*, **39**, 33-38.
- Ruppert, D., Sheather, S. J. and Wand, M. P. (1995), "An Effective Bandwidth Selector for Local Least Squares Regression," *Journal of the American Statistical Association*, **90**, 1257-1270.

- Shaw, K. (1991), "The Influence of Human Capital Investment on Migration and Industry Change," *Journal of Regional Science*, **31**, 397-416.
- Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability 26 (London: Chapman & Hall).
- Sjaastad, L. A. (1962), "The Costs and Returns of Human Migration," *Journal of Political Economy*, **70**, 80-93.
- Smith, J. and Todd, P. (Forthcoming), "Does Matching Address Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*.
- Topel, R. H., and Ward, M. P. (1992), "Job Mobility and the Careers of Young Men," *Quarterly Journal of Economics*, **107**, 439-479.
- Tunali, I. (2000), "Rationality of Migration," *International Economic Review*, **41**, 893-920.
- Willis, R. and Rosen, S. (1979), "Education and Self-Selection," *Journal of Political Economy*, **87**, s7-s36.
- Yankow, J. J. (1999), "The Wage Dynamics of Internal Migration," *Eastern Economic Journal*, **25**, 265-278.
- Yankow, J. J. (2003), "Migration, Job Change, and Wage Growth: A New Perspective on the Pecuniary Return to Geographic Mobility," *Journal of Regional Science*, **43**, 483-516.
- Zhao, Z. (2004), "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence," *Review of Economics and Statistics*, **86**, 91-107.

Appendix A. Three-Step Method for Choosing the Number of Bootstrap Repetitions.

Andrews and Buchinsky (2000, 2001) propose a three-step method for choosing the number of bootstrap repetitions. We follow their procedure to set the proper number of bootstrap repetitions to calculate the standard errors for each parameter we estimate. The following is a special case in Andrews and Buchinsky (2001).

We first define the notations related to our problem following Andrews and Buchinsky (2001). θ is a scalar parameter, and λ is an unknown parameter of interest. In our case θ is the average treatment effect on the treated, and λ is the standard error of θ . B is the number of repetitions, and pdb denotes the measure of accuracy, which is the percentage deviation of the bootstrap quantity of interest based on bootstrap repetitions from the ideal bootstrap quantity for which $B = \infty$. The magnitude of B depends on both the accuracy required and the data. If we required the actual percentage deviation to be less than pdb with a specified probability $1 - \tau$, then the three-step method takes pdb and τ as given and provides a minimum number of repetitions B^* to obtain the desired level of accuracy. We use a conventional accuracy level, $(pdb, \tau) = (10, 0.05)$.

Step 1. Calculate initial number of repetitions B_1

The three-step method depends on a preliminary estimate ω_1 of the asymptotic variance ω of $B^{1/2}(\hat{\lambda}_B - \hat{\lambda}_\infty) / \hat{\lambda}_\infty$, where $\hat{\lambda}_B$ and $\hat{\lambda}_\infty$ are the bootstrap estimates from B and infinite repetitions respectively. Following Andrews and Buchinsky (2000, 2001), we set a starting value of $\omega_1 = 0.5$ in equation A.1 below. (The three-step method is not too sensitive to this starting value because it uses a finite sample estimate of ω in the last step.) Given this starting value, we then calculate

$$B_1 = \text{int} \left(\frac{10,000 * z_{1-\tau/2}^2 * \omega_1}{pdb^2} \right) \quad (\text{A.1})$$

where $z_{1-\tau/2}$ is $1 - \tau/2$ quantile of standard normal distribution. In our case $B_1 = 193$.

Step 2. Use the bootstrap results $\{\hat{\theta} : \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_{B_1}\}$ to update ω_1 to ω_B :

$$\mu_B = \frac{1}{B_1} \sum_{r=1}^{B_1} \hat{\theta}_r \quad (\text{A.2})$$

$$\gamma_B = \frac{1}{B_1 - 1} \sum_{r=1}^{B_1} (\hat{\theta}_r - \mu_B)^4 / se_B^4 - 3 \quad (\text{A.3})$$

where se_B is the standard deviation of $\{\hat{\theta} : \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_{B_1}\}$.

$$\text{Then } \omega_B = (2 + \gamma_B) / 4. \quad (\text{A.4})$$

Step 3. Calculate B_2 from

$$B_2 = \text{int} \left(\frac{10,000 * z_{1-\tau/2}^2 * \omega_B}{p d b^2} \right) \quad (\text{A.5})$$

Then the minimum number of repetitions is $B^* = \max(B_1, B_2)$.

Table 1. Roles of Matching and Differencing in Achieving Ignorable Treatment Assignment

Normalization: $E(e_{jk}^2) = E(e_5^2) = E(e_6^2) = \sigma_e^2 = 1$ $\beta = 1$ $\phi_{3c} = 1$

Case	Influential Parameters			Correlation Coefficient	
				Differencing and Conditioning on X_2	Neither Differencing Nor Conditioning on X_2
	σ_μ^2	$\sigma_{\tilde{\mu}}^2$	ϕ_6	ρ_{MD}	ρ
1. Base case	10	5	0.75	0.044	0.596
2. Low influence from permanent component	5	2.5	0.75	-0.075	0.404
3. High influence from permanent component	20	10	0.75	0.212	0.752
4. Less effective differencing	10	5	0.5	0.222	0.596
5. Less effective matching	10	7.5	0.75	0.137	0.596
6. Less effective differencing and less effective matching	10	7.5	0.5	0.344	0.596

Table 2. Variable Definitions and Descriptive Statistics

Variable Name	Variable Definition	Means			
		Whole Sample	Movers	Stayers	Difference
Migration Dummy					
Migrate	=1 if respondent moved at least 50 miles or changed MSA and moved at least 20 miles	0.18 (0.009)			
Individual Characteristics					
Age	Age in years	26.08 (0.096)	25.99 (0.185)	26.10 (0.110)	-0.106 (0.215)
Black	=1 if African American	0.23 (0.009)	0.15 (0.018)	0.25 (0.011)	-0.108 (0.021)
Hispanic	=1 if Hispanic	0.13 (0.007)	0.13 (0.017)	0.14 (0.008)	-0.009 (0.019)
Dropout	=1 if highest grade completed is less than 12	0.17 (0.008)	0.12 (0.017)	0.18 (0.009)	-0.06 (0.019)
High_school	=1 if highest grade completed is equal to 12	0.47 (0.011)	0.33 (0.024)	0.50 (0.012)	-0.166 (0.027)
Some_college	=1 if highest grade completed is greater than 12 and less than 16	0.18 (0.008)	0.18 (0.020)	0.18 (0.009)	0.002 (0.021)
College	=1 if highest grade completed is greater than 16	0.18 (0.008)	0.36 (0.025)	0.14 (0.008)	0.227 (0.026)
Married	=1 if married, spouse present	0.40 (0.011)	0.44 (0.026)	0.39 (0.012)	0.057 (0.028)
Home_Owner	=1 if own home on job 1	0.20 (0.009)	0.14 (0.018)	0.21 (0.010)	-0.068 (0.020)
MSA	=1 if reside in MSA at time of job 1	0.86 (0.008)	0.84 (0.019)	0.87 (0.008)	-0.029 (0.020)
Job 1 Characteristics					
log(startwage1)	Logarithm of starting wage on job 1 (\$1990)	2.02 (0.009)	2.13 (0.023)	1.99 (0.010)	0.141 (0.025)
log(endwage1)	Logarithm of ending wage on job 1 (\$1990)	2.09 (0.010)	2.21 (0.024)	2.07 (0.010)	0.142 (0.026)
Tenure	Tenure of job 1	2.60 (0.052)	2.66 (0.121)	2.59 (0.058)	0.070 (0.134)
Professional1	=1 if professional/managerial occupation on job 1	0.22 (0.009)	0.39 (0.025)	0.18 (0.009)	0.210 (0.027)
Job 2 Characteristics					
log(startwage2)	Logarithm of starting wage on job 2 (\$1990)	2.19 (0.010)	2.30 (0.026)	2.16 (0.011)	0.142 (0.029)
Family Background Characteristics					
Same_county	=1 if respondent resides at age 14 in the same county as county of birth	0.57 (0.011)	0.49 (0.026)	0.59 (0.012)	-0.096 (0.028)
Father_college	=1 if father was college graduate	0.14 (0.008)	0.25 (0.022)	0.12 (0.008)	0.132 (0.024)

Note: Sample size equals 2,078, and the sample consists of 378 movers and 1,700 stayers. Standard errors of mean are in parentheses.

Table 3. Propensity Score Coefficient Estimates

	Model I	Model II	Model III
Intercept	-5.86 (1.31)	-5.82 (1.30)	-5.86 (1.31)
Age/10.0	4.12 (0.99)	4.13 (0.98)	4.15 (0.99)
Age**2 /100.0	-0.82 (0.19)	-0.83 (0.19)	-0.83 (0.19)
Hispanic	0.05 (0.10)	0.03 (0.10)	0.04 (0.10)
Black	-0.24 (0.09)	-0.27 (0.09)	-0.21 (0.09)
Dropout	-0.50 (0.14)	-0.58 (0.13)	-0.52 (0.14)
High_school	-0.54 (0.11)	-0.60 (0.11)	-0.55 (0.11)
Some_college	-0.38 (0.11)	-0.42 (0.11)	-0.40 (0.11)
Married	0.17 (0.08)	0.16 (0.08)	0.17 (0.08)
MSA1	-0.32 (0.10)	-0.30 (0.10)	-0.26 (0.11)
Professional1	0.34 (0.09)	0.34 (0.09)	0.67 (0.24)
Home_Owner1	-0.53 (0.11)	-0.53 (0.11)	-0.54 (0.11)
log(startwage1)	0.18 (0.11)	0.18 (0.11)	0.18 (0.11)
Tenure	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)
log(endwage1)	0.06 (0.11)	0.07 (0.11)	0.06 (0.11)
Father_college	0.21 (0.10)		0.18 (0.10)
Same_county			-0.16 (0.07)
Same_county*Professional1			-0.37 (0.25)
Chi-square statistic*	8.08	8.72	8.12

Note: Values in the parentheses are standard errors.

* Chi-square statistics are from the likelihood ratio tests against the model without the three job 1 variables, starting wage, ending wage and tenure. Critical value at 5 percent significance level is 7.82.

Table 4. Balancing Tests

Panel A: <i>t</i>- tests						
	Model I		Model II		Model III	
	Difference	Paired <i>t</i> Statistics	Difference	Paired <i>t</i> Statistics	Difference	Paired <i>t</i> Statistics
Age	-0.0880	-0.2361	0.0176	0.4798	-0.0147	-0.3986
Hispanic	0.0117	0.4583	0.0147	0.5846	-0.0147	-0.5768
Black	0.0235	0.8726	0.0029	0.1123	0.0117	0.4645
Married	-0.0059	-0.1609	-0.0235	-0.6167	-0.0235	-0.6483
Father_college	0.0147	0.5020	-0.1026 ^b	3.7381**	0.0264	0.9761
MSA1	-0.0147	-0.5620	-0.0088	-0.3414	-0.0059	-0.2261
ProfessionalI	-0.0411	-1.3488	-0.0235	-0.8525	-0.0059	-0.1922
Home_OwnerI	0.0117	0.5158	-0.0205	-0.7773	-0.0205	-0.8679
log(startwageI)	-0.0005	-0.0172	-0.0222	-0.7169	-0.0014	-0.0476
Tenure	-0.0381	-0.2142	-0.0064	-0.0379	-0.0618	-0.3505
log(endwageI)	0.0023	0.0714	-0.0200	-0.6047	-0.0203	-0.6165
Same County	-0.1085 ^a	-2.8939**	-0.1085 ^b	-2.8939**	0.0147	0.4236

Panel B: *F* Tests^c

	Model I	Model II	Model III
1st quartile	0.40	0.84	0.86
2nd quartile	0.52	0.65	0.68
3rd quartile	0.64	0.81	1.29
4th quartile	1.60	0.90	0.99
Critical value at 5% level	$F(11, 74) = 1.95$	$F(10, 75) = 1.99$	$F(12, 73) = 1.92$

a. The Same County variable is not included in Model I.

b. The Father_college and Same County variables are not included in Model II.

c. Each *F* test is based on the variables included in the respective model.

** Significant at the 5% level.

Note: All tests based on nearest neighbor matching with $q = 5$ trimming.

**Table 5. Specification Tests: "Effect" of Migration on Wage Growth
on Job 1 by Education Group**

	Model I	Model II	Model III
High school dropouts	2.30% (3.61%)	2.33% (3.44%)	3.13% (3.67%)
High school graduates	-3.16% (3.41%)	-3.18% (3.38%)	-3.07% (3.28%)
Some college	4.26% (3.52%)	4.46% (3.45%)	2.93% (3.49%)
College graduates	-0.30% (2.30%)	0.01% (2.47%)	0.28% (2.47%)

Note: All tests based on local linear regression matching with $q = 5$ trimming. In the specification tests, the wage growth is standardized by job tenure, and standard errors are in parentheses. We use 350 repetitions, which is greater than the minimum required by the Andrews-Buchinsky method, to calculate standard errors. Since this variable is pre-migration, any significant "treatment effect" for this variable can only reflect selection bias that matching fails to correct for.

Table 6. Matching Estimates of the Effect of Migration on Wage Growth from Model I

Panel A: Trimming level $q = 5$					
Estimator	Overall	Dropouts	High_school	Some_college	College_Grads
Local linear 25% bandwidth	-0.56%	-12.46%	-4.73%	-0.75%	10.20%
(200 repetitions)	(2.35%)	(6.83%)	(3.79%)	(5.66%)	(5.03%)
[300 repetitions]	[2.32%]	[7.25%]	[3.65%]	[5.66%]	[5.06%]
{1100 repetitions}	{2.41%}	{7.25%}	{3.99%}	{5.56%}	{5.18%}
Local cubic 25% bandwidth	-1.64%	-12.51%	-4.90%	-4.48%	9.28%
(200 repetitions)	(3.48%)	(8.72%)	(5.28%)	(10.75%)	(5.81%)
[300 repetitions]	[3.62%]	[12.52%]	[5.89%]	[9.93%]	[5.62%]
{1100 repetitions}	{6.28%}	{29.26%}	{15.05%}	{8.74%}	{6.22%}
Panel B: Trimming level $q = 3$					
Local linear 25% bandwidth	0.63%	-12.46%	-4.73%	-0.70%	12.56%
(300 repetitions)	(2.27%)	(7.40%)	(3.83%)	(5.73%)	(4.88%)
Panel C: Trimming level $q = 7$					
Local linear 25% bandwidth	-0.88%	-12.46%	-4.73%	-0.75%	10.86%
(300 repetitions)	(2.44%)	(7.08%)	(3.95%)	(5.64%)	(5.29%)

Table 7. Matching Estimates of the Effect of Migration on Wage Growth Based on Three Alternative Models

Panel A: Model I					
Estimator	Overall	Dropouts	High_school	Some_college	College Grads
Local linear 25% bandwidth (300 repetitions)	-0.56% (2.32%)	-12.46% (7.25%)	-4.73% (3.65%)	-0.75% (5.66%)	10.20% (5.06%)
Panel B: Model II					
Local linear 25% bandwidth (300 repetitions)	-0.32% (2.46%)	-12.20% (7.10%)	-4.79% (3.96%)	0.03% (5.61%)	10.42% (5.35%)
Panel C: Model III					
Local linear 25% bandwidth (300 repetitions)	-0.02% (2.51%)	-12.80% (7.39%)	-4.12% (4.14%)	-1.48% (5.66%)	12.03% (5.16%)

Note: All three panels use q = 5 trimming level. See Table 3 for model specifications.

Table 8. Comparisons Between Movers and Stayers under Three Definitions of Migration

Panel A. Distance Between Consecutive Locations (Miles)					
Definition of Migration	Migration Status	N	Average Distance	Minimum Distance	Maximum Distance
Distance-Based Measure	Mover	378	535.43	20	3772
	Stayer	1700	3.82	0	49
Change-of-State Measure	Mover	258	722.28	1	3772
	Stayer	1820	12.38	0	668
Change-of-County Measure	Mover	542	378.68	1	3772
	Stayer	1536	2.37	0	38

Panel B. Misclassification of Movers and Stayers When Move is Defined as Change-of-State or Change-of-County Relative to a Distance-Based Measure					
		N	Average Distance	Minimum Distance	Maximum Distance
Change-of-State Measure	Undercounts of Movers	136	119.86	20	668
	Overcounts of Movers	16	16.00	1	44
Change-of-County Measure	Undercounts of Movers	0	-	-	-
	Overcounts of Movers	164	17.37	1	49

Table 9. Matching Estimates of the Effect of Migration on Wage Growth Based on Three Definitions of Migration

Panel A: Distance-Based Measure					
Estimator	Overall	Dropouts	High_school	Some_college	College Grads
Local linear 25% bandwidth (300 repetitions)	-0.56% (2.32%)	-12.46% (7.25%)	-4.73% (3.65%)	-0.75% (5.66%)	10.20% (5.06%)
Panel B: Change-of-County Measure					
Local linear 25% bandwidth (300 repetitions)	-1.98% (2.05%)	-7.35% (5.70%)	-6.23% (2.98%)	-0.48% (5.20%)	7.27% (4.81%)
Panel C: Change-of-State Measure					
Local linear 25% bandwidth (300 repetitions)	0.03% (2.91%)	-7.13% (8.49%)	-4.71% (5.41%)	-1.67% (6.86%)	8.59% (6.70%)

Note: This table uses the baseline model with a q=5 trimming level.

Figure 1: Distributions of Estimated Propensity Score

(Based on Model I)

