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

John C. Ham
Xianghong Li
Patricia B. Reagan

Department of Economics and CHRR
Ohio State University

First Draft: August 2001
This Draft: November 2003

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 address selection issues. An economic model helps us determine the variables included in the propensity score. 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, William Greene, 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, McGill/University of Montreal, McMaster, Minnesota (Industrial Relations), Ohio State, Pompeu Fabra, Rutgers, Toronto, UC Berkeley, UC Davis, UC San Diego, USC, 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 through grant SES0136925. We are responsible for all errors. This paper in no way represents the views of the NSF.

I. Introduction

Internal migration is an important economic phenomenon in the United States. Between March 1999 and March 2000, about 43 million Americans moved, and more than 67 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 wage gain from U.S. internal migration.

We contribute to the 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 return to 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 measure 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

[1] See http://www.census.gov/prod/2001pubs/p20-538.pdf.
[2] To the best of our knowledge Yankow (1999) is the only other author who allows migration effects to differ by education.

overestimates migration by about 43%. Further, we are not able to find a significant migration effect for college graduates or high school dropouts by using the alternative measures of moving across state or county.

Selection bias can be a severe problem in migration studies, and we use the propensity score matching method (Rosenbaum and Rubin 1983) to address this problem. A theoretical model helps us determine the appropriate variables on which to base the matching, and the NLSY79 data provide a rich set of such variables. By matching movers and stayers *within* each education group (based on the estimated propensity score) rather than matching individuals across the entire sample, we implement what Rosenbaum and Rubin call finer balancing. We obtain stable and sensible results from this approach. 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 procedure for choosing the number of bootstrap repetitions is quite helpful, and in our case suggests a higher number of repetitions than the number often used by applied researchers. In the case of one estimator, using too small a number of bootstrap repetitions gives misleading results.

Our paper is organized as follows. After reviewing the migration and matching literature, we present our econometric model. We then describe the theoretical model that we use to guide our choice of the conditioning variables in the matching procedure. The next two sections describe our data and our empirical results. The paper ends with a summary of our conclusions.

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 costs 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, the

[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.

determinants of migration and 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 lifecycle. Geographic mobility peaks during the early to mid-twenties and declines thereafter with age 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 propensity is affected by migration costs, both pecuniary and nonpecuniary. 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. Thus, psychic costs entailed in moving away from friends and family play a role in deterring migration.

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 returns to migration. Estimates of the average contemporaneous returns can be 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 low 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.

---

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

[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 experiences wage reductions after moving, a minority realizes 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.

The sign and significance of the migration effect depend on the sample chosen and on how researchers address three critical questions. First, what is the definition of migration being 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 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 the 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 Schmitz (1995) also find favorable self-selection. However, Hunt and Kau (1985) and Borjas, Bronars, and Trejo (1992a) find no evidence of self-selection. Given the mixed evidence on the importance of sample selection, we thought it appropriate to consider matching as an alternative to dealing with the selection issue.[7]

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

[7] The empirical literature on matching as an alternative means of addressing selection, literature that has grown extremely quickly in recent years, especially since the first draft of our paper in 2001, is too extensive to review here. Some of the earlier papers are Dehejia and Wahba (1999), Heckman, Ichimura, and Todd (1997) and Lechner (1999, 2000). These studies all consider matching in the context of job training.

III. Econometric Model

3.1 Estimating the Effect of Moving

Our goal is to estimate the effect of internal migration on between job wage growth for those who quit their first job. Following the notation in the evaluation literature, let $D=1$ if an individual moves and $D=0$ otherwise. Also we define the outcome with moving ($D=1$) as Y_1 and the outcome for those who do not move ($D=0$) 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 effect of treatment on the treated (i.e., the effect of migration on those who migrate)

$$\Delta = E(Y_1 - Y_0 \| D=1) = E(Y_1 \| D=1) - E(Y_0 \| D=1) \ . \qquad (3.1)$$

We observe the first term on the righthand side of equation 3.1. However, we do not observe the second term on the righthand 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* (Rosenbaum and Rubin 1983) or *selection on observables* (Heckman and Robb 1985). Formally, this condition is written as

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

where X^* is a vector denoting 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. This is known as the conditional independent assumption (CIA). In empirical work X^* usually contains pretreatment variables and time-invariant individual characteristics. In the next section we use a theoretical model and the variables available to us to argue that our rich data makes CIA plausible and thus matching is a suitable estimation approach for our problem.

Matching identification also requires that

$$0 < \Pr(D=1 \| X^*) < 1. \quad (3.3)$$

This common support condition requires that at each level of X^*, the probability of observing both the participants and nonparticipants is positive.[8] (This condition can be enforced by adding a common support constraint – we discuss this in Section 3.6)

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 general equilibrium effects. Since our 378 migrants are from a sample randomly drawn in the 50 states, SUTVA is reasonable in our problem.

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. It 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

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

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

$$(Y_1, Y_0) \perp D \| P\left(X^*\right) \quad (3.5)$$

and

$$0 < \Pr\left(D=1 \| P\left(X^*\right)\right) < 1 \ . \quad (3.6)$$

If CIA 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.

[8] Since we focus on the treatment effect on the treated, we only need the second part of the inequality.

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. Then 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 $E(Y_{0i} \| D=1)$ by 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. Heckman, Ichimura, and Todd (1997, 1998), and Heckman, Ichimura, Smith, and Todd (1998) incorporate local regression into matching. For each observation $i \; (i = 1, \ldots, N_1)$ 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 $\hat{m}\left(\hat{P}\left(X_i^*\right)\right)$ to approximate $E(Y_{0i} \mid D=1)$. Within the window, the closer the $\hat{P}(X_j^*)$ is to $\hat{P}(X_i^*)$, the greater the weight the observation j gets in estimating $\hat{m}\left(\hat{P}\left(X_i^*\right)\right)$.

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:

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

where $K(\bullet)$ 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 $(w_j, z_j) = \left(\hat{P}(X_j^*), Y_{0j}\right)$. For each mover i ($i = 1, N_1$), we run a local regression at its estimated propensity score $\hat{P}\left(X_i^*\right)$ and estimate $\hat{m}\left(\hat{P}\left(X_i^*\right)\right)$. 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 the boundary regions and in the highly clustered regions). There is no corresponding increase, however, in the asymptotic variance. (Their result implies that kernel regression is dominated by local linear regression, at least asymptotically.) 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 (2001) 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 and discuss it 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. Frolich's results suggest that local linear matching performs reasonably well when the comparison group is large relative to the treatment group (say a ratio of the comparison group to the treatment group 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.

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; 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. The variable bandwidth approach changes the bandwidth according to the data density around W_0. In other words, it 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, pp 2013) put it, "A different amount of smoothing is used at different data locations."

Fan and Gijbels (1992) suggest 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 give the same number of data points k_n that are closest to 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.[9] Our variable bandwidth is compatible with the suggestion of Silverman (1986) and others that one should use a subjective bandwidth choice.[10]

3.4 The Sampling Variance of the Matching Estimator

We follow the previous literature and use the bootstrap to obtain standard errors for the matching estimators. An important decision facing a researcher 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.

---

[9] Fan and Gijbels (1996, theorem 4.2) prove that if $k_n \to \infty$ such that $k_n / n \to 0$ and $k_n / \log n \to \infty$, 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.

[10] 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.

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 their 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. We now want to estimate

$$\Delta_s = E(Y_1 - Y_0 \| D=1, S=s) = E(Y_1 \| D=1, S=s) - E(Y_0 \| D=1, S=s). \qquad (3.8)$$

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 but only match individual j to individual i if individual j is in individual $i's$ schooling class.[11] As noted earlier, 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 the movers with the stayers in the same educational group based on the estimated propensity score. In our empirical work below, we find that it is important to allow the treatment effect to vary by educational group.

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 condition in equation 3.3, we add a common support constraint, following the procedure proposed by Heckman, Ichimura, and Todd (1997).[12] 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

---

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

[12] See also Smith and Todd (forthcoming) for details.

trimming is to q percent.[13] 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.

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 the difference 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 the 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.[14] At each quartile, we test if all elements of X^* are *jointly* different across the two groups. If a model fails to pass either the *t*-tests or the *F* tests, we add higher order terms or interaction terms until the variables are balanced across the two groups (Smith and Todd, forthcoming).

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 finer balancing 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.[15]

---

[13] In the earlier version of this paper (Ham, Li, and Reagan 2001), we propose 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. Our procedure produced results for our sample that are very similar to those produced by the Heckman, Ichimura, and Todd procedure described above.

[14] 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.

[15] 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.

IV. Using an Economic Model of Migration to Choose the Conditioning Variables to Satisfy the Conditional Independence Assumption

The Conditional Independence Assumption (CIA) underlies all matching estimators. It is a strong identification assumption since it assumes that all selection bias related to the outcome has been removed after controlling for X^*. Thus, matching depends crucially on choosing conditioning variables X^* that satisfy CIA. This "data hungry" identification strategy requires a rich data set that contains *all* premigration variables related to *both* the treatment decision and the outcome (Rosenbaum and Rubin 1985, Lechner 1999).

In order to determine the variables for the propensity score, and to investigate the identification of our model, we modify the Willis and Rosen (1979) model of education to 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**.** Let X_{ki} denote a vector of observed individual human capital variables and local labor market conditions. We use $k = c$ to denote the initial labor market and $k = m$ to denote the labor market to which an individual migrates. Let ε_{1ki} *and* ε_{2ki} represent unobserved individual components (such as motivation and unmeasured ability) of the initial wage and growth rate respectively in earnings capabilities in location k. At the beginning of the period, expected future earnings in each location k are given by

$$V_{ki} = v\left(X_{ki}, \varepsilon_{1ki}, \varepsilon_{2ki}\right),\ k = c, m. \qquad (4.1)$$

The cost of changing jobs is given by

$$C_{ki} = C_k\left(W_i, u_{ki}\right),\ k = c, m, \qquad (4.2)$$

where W_i is a vector of variables that affects the relative costs of changing jobs locally and across locations. The vector W_i includes tastes for migration measured by family background variables, such as whether the individual was living in his county of birth at age 14. Further, u_{ci}

and u_{mi} are unobserved error terms. Net expected future earnings from changing jobs locally and across markets are

$$V_{ki} - C_{ki},\ k = c, m\ . \qquad (4.3)$$

Workers choose to migrate if

$$V_{mi} - C_{mi} > V_{ci} - C_{ci}\ . \qquad (4.4)$$

To determine the appropriate variables for the CIA to hold, we adopt the following structure. The starting wage on the second job, y_{ki}^{S}, is

$$\ln y_{ki}^{S} = X_{1ki}\gamma_{1k} + \varepsilon_{1ki}, \quad k = c, m, \qquad (4.5)$$

where X_{1ki} is a vector consisting of a subset of the elements in X_{ki}, and γ_{1k} is a vector of returns to the X_{1ki} variables. The labor market specific growth rates of wages on the second job are

$$g_{ki} = X_{2ki}\gamma_{2k} + \varepsilon_{2ki},\ k = c, m, \qquad (4.6)$$

where X_{2ki} is a vector consisting of a subset of the elements in X_{ki}, and γ_{2k} is a vector of returns to the X_{2ki} variables. Finally, the worker's discount rate, r_i, is a function of family background variables T_i:

$$r_i = T_i\delta + \varepsilon_{3i}. \qquad (4.7)$$

If the individual takes the local job, the expected wage at time t (where t is normalized to zero at the time the worker quits his first job) is

$$y_{ci}(t) = y_{ci}^{S} e^{g_{ci}t}\ . \qquad (4.8)$$

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

$$\begin{aligned} y_{mi}(t) &= 0 && \text{for } t \le M \\ &= y_{mi}^{S} e^{g_{mi}(t-M)} && \text{for } t > M. \end{aligned} \tag{4.9}$$

We make the following assumptions. First, gross utility is the present value of wages. Second, workers face an infinite horizon. Third, the discount rate, r_i, is constant for each individual, where $r_i > \max(g_{si}, g_{mi})$. Finally, the costs of changing jobs or migrating enter the net utility function exponentially. Under these assumptions, the net utility of changing jobs locally can be written as[16]

$$U_{ci} = V_{ci} - C_{ci} = \left(y_{ci}^{S} \big/ (r_i - g_{ci}) \right) e^{-W_i \lambda_c - u_{ci}} , \tag{4.10}$$

where λ_c is a vector that weights individual characteristics to reflect the costs of local job changing. The net utility of changing jobs across labor markets can be written as

$$U_{mi} = V_{mi} - C_{mi} = \left(y_{mi}^{S} \big/ (r_i - g_{mi}) \right) e^{-r_i M} e^{-W_i \lambda_m - u_{mi}} , \tag{4.11}$$

where λ_m is a vector of weight parameters reflecting migration cost.

We define $I_i = \ln\left(U_{mi} / U_{ci}\right)$. Substituting from equations 4.10 and 4.11 and taking a Taylor series approximation around the population mean values of $\left(\overline{g}_s, \overline{g}_m, \overline{r}\right)$ yields

$$I_i = \alpha_0 + \ln y_{mi}^{S} - \ln y_{ci}^{S} + \alpha_1 g_{mi} + \alpha_2 g_{ci} + \alpha_3 r_i + \alpha_4 W_i + u_{ci} - u_{mi} \tag{4.12}$$

[16] We assume that individuals choose sensible parameters for y_k^S *and* g_k, $k = c, m$, and then act as if there is no uncertainty. This seems reasonable given our data and empirical model.

where $\alpha_1 = 1/(\bar{r} - \bar{g}_m) > 0, \alpha_2 = -1/(\bar{r} - \bar{g}_c) > 0, \alpha_3 = -M + (\bar{g}_m - \bar{g}_c) / [(\bar{r} - \bar{g}_c)(\bar{r} - \bar{g}_m)]$ and $\alpha_4 = \lambda_c - \lambda_m$.

Substituting equations 4.5, 4.6, and 4.7 into 4.12 yields the migration decision rule

$$I_i = \theta Z_{1i} + \varepsilon_i^* > 0 , \qquad (4.13)$$

where

$\varepsilon_i^* = \varepsilon_{1mi} - \varepsilon_{1ci} + \alpha_1 \varepsilon_{2mi} - \alpha_2 \varepsilon_{2ci} + \alpha_3 \varepsilon_{3i} + u_{ci} - u_{mi}$, and Z_{1i} contains the unique elements of X_{ki}, W_i and T_i. We assume that all error components in ε_i^* are correlated, since if these correlations are zero, there is no selection problem. Consider now our outcome equations. Since we implement difference-in-difference matching, we want to look at 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

$$\ln y_i^E = X_{4i} \gamma_4 + \varepsilon_{4i} . \qquad (4.15)$$

For those who move, our outcome variable is

$$\ln y_{mi}^S - \ln y_i^E = X_{1m} \gamma_{1m} - X_4 \gamma_4 + \varepsilon_{1mi} - \varepsilon_{4i} \text{ , and} \qquad (4.16)$$

for those who stay, the outcome variable is

$$\ln y_{ci}^S - \ln y_i^E = X_{1c} \gamma_{1c} - X_4 \gamma_4 + \varepsilon_{1ci} - \varepsilon_{4i} . \qquad (4.17)$$

Having set up the migration decision equation and the two outcome equations, we are ready to choose our conditioning variables, X_i^*, to achieve the CIA. Concerning this choice there are three points we want to make.

First, we cannot simply use Z_{1i} as the conditioning variables to achieve the CIA since there is no reason to assume ε_i^* is independent of $(\varepsilon_{1mi} - \varepsilon_{4i})$ and $(\varepsilon_{1ci} - \varepsilon_{4i})$, a necessary condition for CIA to hold. We would expect that taking the difference between the starting wage on the second job and the ending wage on the first job would help us to meet CIA but would not allow us to completely satisfy it. Instead we assume that we have other variables Z_{2i} such that

$$\varepsilon_i^* = f(Z_{2i}) + \tilde{\varepsilon}_i \text{, where} \quad \tilde{\varepsilon}_i \perp (\varepsilon_{1ki} - \varepsilon_{4i}), k = m, c \,. \tag{4.18}$$

Of course, the question arises as to what to include in Z_{2i}. We think of ε_i^* as containing unobservable traits such as ability, motivation, and inclination towards turnover. We will use the following variables in Z_{2i} (beginning wage, ending wage, and tenure of the first job) as proxies for the individual specific traits listed above.

Second, we will include in X_i^* the migration decision variables (Z_{1i} in equation 4.13) to the extent that they also affect the outcome, or are correlated with the error term in the outcome equations. Variables such as age, education, professional status, marital status, and race will directly affect wages. We would expect that home ownership would affect moving costs and would be correlated with unobservables in the wage equation. We would expect background variables indicating the wealth of the individual's parents to affect the discount rate. Whether these background variables enter the wage equation or are correlated with the unobservables in the wage equation is an open question.[17] We include them in X_i^* but experiment with excluding them from the propensity score. We also include in X_i^* a dummy variable indicating if an individual lived in an MSA during the first job period.[18]

Third, we will not use a Z_{1i} variable in X_i^* if it affects only moving costs but not wages, since excluding such a variable will not introduce selection bias into the outcome evaluation. Furthermore, excluding it from the propensity score will help the matching identification. For example, we would expect individuals who, at age 14, lived in the same county where they were born to have higher physic costs of moving, but we would not expect this variable to affect wages. As Heckman and Navarro-Lozano (2003) point out, this type of variables do play an important role in identification. Variation in this variable will sort individuals with similar earning capabilities into the mover and stayer categories due to different moving costs.

[17] Willis and Rosen (1979) assume that family background 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.

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

In summary, the variables in the propensity score are: labor market history variables, human capital variables, family background variables, and home ownership. Although the CIA assumption is not testable, a credible case for its holding in our economic problem can be made.

V. Data Description

Our primary data sources are 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, and biennially 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.[19] 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 (see Table 1). 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 begins 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 nonenrollment spell lasting at least 12 consecutive months. Accordingly, respondents are excluded from the sample if the

[19] 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.

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.

[Insert Table 1 here]

Of the 6,403 male respondents in the initial sample, 184 individuals were deleted because they never reported a job. Another 78 were deleted because they never left school, or an exact date of school leaving could not be determined. Another 215 were deleted because they never reported a civilian job. We eliminated an additional 3,569 observations because we could not identify a "clean" job-to-job transition spanning two consecutive interviews with valid address data. For a job to be considered, it had to last at least 6 months and require at least 25 hours per week. A "clean" job-to-job transition is defined as one with no more than 2 months of nonemployment between the jobs and no more than 3 months overlap while both jobs are being held. Finally, each job had to span the date of an interview. In order to identify migrants, the two jobs had to overlap with at least two consecutive interviews for which we were able to match the respondent's address to a latitude and longitude. After excluding another 279 respondents who did not have valid data for all the variables used in this analysis, [20] we reached a final sample of 2,078 men as Table 1 shows.

We explore migration conditional on voluntarily quitting the previous 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 have occurred if the respondent moved at least 50 miles or changed Metropolitan Statistical Area (MSA) and moved at least 20 miles.[21] We focus on wage growth between the first two jobs with a "clean" job-to-job transition as just defined. We call them job 1 and job 2 thereafter.

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]

[20] If an hourly wage on the first job or the second job was reported as greater than $50 dollars or less than $3 (in real terms) it was considered invalid.
[21] Adjacent county centroids are typically about 25 miles apart, so a move of 50 miles roughly corresponds to a move two counties away.

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 to identify 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).

Family background characteristics are hypothesized to affect the costs of changing jobs across labor markets relative to the costs of changing jobs locally. Characteristics such as whether the father has a college degree are likely to affect resources available to finance a move. 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 a return to migration.

VI. Empirical Results

6.1 Propensity Score Models

Table 3 reports the probit estimates of 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. The same-county variable, we believe, 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. Model III 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. By comparing the matching estimates in Models I and III, we test the robustness of Model I to 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 are less likely, to move than 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 nonmetropolitan areas. Men in professional occupations on job 1 are more likely to migrate. Homeownership has a negative and statistically significant effect on migration, as does living in an urban area. 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 all of the other coefficients are not 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 to 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 on 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 with 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.[22] Matching does a good job with regard to those 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 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 since this variable is pre-treatment, any significant "treatment effect" can only reflect selection bias that finer matching fails to correct. Panel A reports test statistics based on local linear regression matching and Panel B reports those based on nearest neighbor matching. None of these "effects" is significantly different from zero.

[Insert Table 5 here]

[22] 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. This basically confirms the probit results that both these variables affect the migration decision.

6.3 Estimates of the Migration Effects from the Baseline Model

Table 6 presents the matching estimates of the effect of migration on wage growth from the baseline propensity score model (Model I). *Recall that all estimates represent the effect of treatment on the treated (i.e., what migrants gained by moving over what they would have made had they not migrated).*[23] 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 218, 248 and 1,074 for the nearest neighbor, local linear, and local cubic estimators respectively.[24] Most of the literature uses at most 200 repetitions. For the nearest neighbor and local linear estimators, 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 minimums of 218 and 248 respectively. 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 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.

We focus here on the local linear estimates with standard errors calculated from 300 bootstrap repetitions. A 25% bandwidth gives us a wide enough window when we disaggregate the data by educational class[25]. 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

[23] Note that this is not the unconditional effect that a selection model would estimate.

[24] 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.

[25] 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% to 2% bandwidth change.

wage growth from migration for job changers who have only a high school education or some college. Nearest neighbor matching provides noisier estimates of returns to migration because the procedure uses the data less efficiently than local linear matching. 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 things 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 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 (to the extent the difference is significant) 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 optimization problem, even approximately. Alternatively, they may not be able to see wages in the other location without visiting it.[26]

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.[27] 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 and positive effect for college graduates.

---

[26]As one seminar participant put it, "college graduates can search and then move, while dropouts must move before they can search."

[27] The return migration within two years is 36% for dropouts and 24% for the overall sample. The repeat migration within two years is 36% for dropouts and 22% for the whole sample.

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.) For both cases (and for the results presented later) we use 300 bootstrap repetitions, more than the required minimum from the three-step Andrews-Buchinsky method. In each case, nearest neighbor matching still produces imprecise estimates. The estimates from local linear matching 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. The effects estimated with the local linear estimator are almost identical under Model I and Model II. Recall that under Model II, the father's college degree variable is not balanced between the movers and the matched sample of stayers (see Table 4). Since father's education is significant in Model I and since migration effects are similar regardless of whether this variable is included in the propensity score, it appears that father's education significantly affects the moving decision but does not provides extra information with regard to unobserved earning ability, after controlling for all the other individual characteristics and the lagged variables. The nearest neighbor matching estimates are very noisy under both Model I and II and, not surprisingly, the point estimates vary substantially across the two models.

[Insert Table 7 here]

Panel C of Table 7 reports estimates from Model III, in which we add the same-county variable. This variable is not balanced in Models I and Model II (see Table 4). Further, its coefficient is significant in Model III (see Table 3). Again the estimates from the local linear estimator are very close to those from Model I and Model II. Taken together 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 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 defined as 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, which is conditional on changing residence from one MSA to another. The maximum distance for stayers is 49 miles, which is 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.[28] 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, and $q = 5$ trimming. Once again we use 300 bootstrap repetitions, which is larger than the minimum numbers calculated from the three-step method. 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. Specifically, 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]

VII. Conclusion

Our paper estimates the effect of U.S. internal migration, for those who quit their first job, on the 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 aspects. First, we

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

exploit the confidential geocoding in the National Longitudinal Surveys of Youth 1979 to obtain a distance-based measure of migration rather than define 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. An economic model helps us determine which variables should be included in the propensity score used in the matching procedure. Matching is a "data hungry" estimation strategy, and our data set provides a rich array of variables on which to match. This, in turn, makes the Conditional Independence Assumption, which underlies all matching, quite plausible in our case. 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 and matching method. Our models pass balancing tests and a specification test. We find that better data matters: if we use either 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 jobs to measure the effect of migration on the wage growth after relocation. Second, we could consider migration effects for young women.

References

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. (Forthcoming), "Program Evaluation as a Decision Problem", *Journal of Econometrics*.

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. (2001), "Nonparametric Covariate Adjustment: Paired-Matching versus Local Polynomial Matching" (Mimeo, University of St. Gallen).

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*, Summer 1995, 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).

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 Nonparmetric 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. and Navarro-Lozano, S. (2003), “Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models”, (wp 9497, NBER).

Heckman, J. and Robb, R. (1985), “Longitudinal Analysis of Labor Market Data”, Econometric Society Monographs series, no. 10 (Cambridge, New York, and Sydney: Cambridge University Press).

Hunt, J. C. and Kau, J. B. 1985. “Migration and Wage Growth: A Human Capital Approach”, *Southern Economic Journal*, **51**, 697-710.

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. (1999), “Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification”, *Journal of Business and Economic Statistics*, January 1999, 74-90.

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.

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 notation following Andrews and Buchinsky (2001). 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 conventional accuracy level $\left(pdb,\tau\right) = (10, 0.05)$.

Step 1. Calculate initial number of repetitions B_1

Set a starting value $\omega_1 = 0.5$ in equation A.1 below. This is the specification for calculating standard errors based on asymptotics, but this method is not sensitive to the starting value.

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

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 $\left\{\hat{\theta} : \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_{B_1}\right\}$ to update ω_1 to ω_B:

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

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

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

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

Step 3. Calculate B_2 from

$$B_2 = \text{int}\left({10{,}000 * z^2_{1-\tau/2} * \omega_B}/{pdb^2} \right) \quad \text{(A.5)}$$

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

Table 1. Sample Selection Criteria and Resulting Sample Size

Criteria	Sample Size
Male respondents in NLSY79	6,403
Number after deletion because respondent reported no job	6,219
Number after deletion because information was missing on the week first left school for at least 12 months	6,141
Number after deletion because no civilian job observations	5,926
Number after deletion because no "clean" job-to-job transition with location information*	2,357
Number after deletion because lacked information on variables used in this analysis was lackingor because the reported hourly wage was less than 3 dollars or more than 50 dollars	2,078

Source: National Longitudinal Surveys of Youth 1979, 1979-1996 waves.

* The job must last at least 6 months and require at least 25 hours per week. A "clean" job-to-job transition is defined as one involving a quit on the first job and no more than 2 months of nonemployment between the jobs and no more than 3 months overlap while both jobs are being held.

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 (0.39)			
Individual Characteristics					
Age	Age in years	26.08 (4.37)	25.99 (3.59)	26.10 (4.53)	-0.106 (0.215)
Black	=1 if African American	0.23 (0.42)	0.15 (0.35)	0.25 (0.44)	-0.108 (0.021)
Hispanic	=1 if Hispanic	0.13 (0.34)	0.13 (0.33)	0.14 (0.34)	-0.009 (0.019)
Dropout	=1 if highest grade completed is less than 12	0.17 (0.38)	0.12 (0.33)	0.18 (0.39)	-0.06 (0.019)
High_school	=1 if highest grade completed is equal to 12	0.47 (0.50)	0.33 (0.47)	0.50 (0.50)	-0.166 (0.027)
Some_college	=1 if highest grade completed is greater than 12 and less than 16	0.18 (0.39)	0.18 (0.39)	0.18 (0.38)	0.002 (0.021)
College	=1 if highest grade completed is greater than 16	0.18 (0.38)	0.36 (0.48)	0.14 (0.34)	0.227 (0.026)
Married	=1 if married, spouse present	0.40 (0.49)	0.44 (0.50)	0.39 (0.49)	0.057 (0.028)
Home_Owner	=1 if own home on job 1	0.20 (0.40)	0.14 (0.35)	0.21 (0.41)	-0.068 (0.020)
MSA	=1 if reside in MSA at time of job 1	0.86 (0.35)	0.84 (0.37)	0.87 (0.34)	-0.029 (0.020)
Job 1 Characteristics					
log(startwage1)	Logarithm of starting wage on job 1 ($1990)	2.02 (0.42)	2.13 (0.45)	1.99 (0.41)	0.141 (0.025)
log(endwage1)	Logarithm of ending wage on job 1 ($1990)	2.09 (0.44)	2.21 (0.47)	2.07 (0.43)	0.142 (0.026)
Tenure	Tenure of job 1	2.60 (2.37)	2.66 (2.35)	2.59 (2.37)	0.070 (0.134)
Professional1	=1 if professional/managerial occupation on job 1	0.22 (0.42)	0.39 (0.49)	0.18 (0.39)	0.210 (0.027)
Job 2 Characteristics					
log(startwage2)	Logarithm of starting wage on job 2 ($1990)	2.19 (0.47)	2.30 (0.51)	2.16 (0.45)	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 (0.50)	0.49 (0.50)	0.59 (0.49)	-0.096 (0.028)
Father_college	=1 if father was college graduate	0.14 (0.35)	0.25 (0.43)	0.12 (0.32)	0.132 (0.024)

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

Table 3. Propensity Score Coefficient Estimates

	Model I	Model II	Model III
Intercept	-5.86	-5.82	-5.86
	(1.31)	(1.30)	(1.31)
Age/10.0	4.12	4.13	4.15
	(0.99)	(0.98)	(0.99)
Age2 /100.0**	-0.82	-0.83	-0.83
	(0.19)	(0.19)	(0.19)
Hispanic	0.05	0.03	0.04
	(0.10)	(0.10)	(0.10)
Black	-0.24	-0.27	-0.21
	(0.09)	(0.09)	(0.09)
Dropout	-0.50	-0.58	-0.52
	(0.14)	(0.13)	(0.14)
High_school	-0.54	-0.60	-0.55
	(0.11)	(0.11)	(0.11)
Some_college	-0.38	-0.42	-0.40
	(0.11)	(0.11)	(0.11)
Married	0.17	0.16	0.17
	(0.08)	(0.08)	(0.08)
MSA1	-0.32	-0.30	-0.26
	(0.10)	(0.10)	(0.11)
Professional1	0.34	0.34	0.67
	(0.09)	(0.09)	(0.24)
Home_Owner1	-0.53	-0.53	-0.54
	(0.11)	(0.11)	(0.11)
log(startwage1)	0.18	0.18	0.18
	(0.11)	(0.11)	(0.11)
Tenure	0.02	0.02	0.02
	(0.02)	(0.02)	(0.02)
log(endwage1)	0.06	0.07	0.06
	(0.11)	(0.11)	(0.11)
Father_college	0.21		0.18
	(0.10)		(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 model without the three job 1 variables, starting wage, ending wage and tenure. Critical value at 5 percent significant level is 7.82.

Table 4. Balancing Tests (Nearest Neighbor Matching)

Panel A: *t*- tests

	Model I		Model II		Model III	
	Difference	Paired *t* Statistics	Difference	Paired *t* Statistics	Difference	Paired *t* 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	3.7381**	0.0264	0.9761
MSA1	-0.0147	-0.5620	-0.0088	-0.3414	-0.0059	-0.2261
Professional1	-0.0411	-1.3488	-0.0235	-0.8525	-0.0059	-0.1922
Home_Owner1	0.0117	0.5158	-0.0205	-0.7773	-0.0205	-0.8679
log(startwage1)	-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(endwage1)	0.0023	0.0714	-0.0200	-0.6047	-0.0203	-0.6165
Same County	-0.1085	-2.8939**	-0.1085	-2.8939**	0.0147	0.4236

Panel B: *F* Tests

	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	$F(11, 74) = 1.95$	$F(10, 75) = 1.99$	$F(12, 73) = 1.92$

Note: All tests based on nearest neighbor matching with q = 5 trimming.
** Significant at the 5% level.

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

Panel A: Local Linear Regression Matching

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

Panel B: Nearest Neighbor Matching

	Model I	**Model II**	**Model III**
High school dropouts	**4.32%**	**5.41%**	**1.28%**
	(5.34%)	(5.34%)	(5.59%)
High school graduates	**-6.74%**	**-2.41%**	**-4.50%**
	(4.38%)	(4.33%)	(4.16%)
Some college	**8.74%**	**1.25%**	**5.01%**
	(5.34%)	(5.52%)	(5.26%)
College graduates	**1.76%**	**2.54%**	**0.90%**
	(3.45%)	(3.77%)	(3.42%)

Note: All tests based on matching with q = 5 trimming. In the specification tests, the wage growth is standardized by job tenure, and in parentheses are standard errors. Since this variable is pre-migration, any significant “treatment effect” for this variable can only reflect selection bias that finer balancing matching fails to correct.
** Significant at the 5% level.

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%}
Nearest neighbor (one)	**-0.84%**	**-12.23%**	**2.46%**	**-8.34%**	**5.28%**
(200 repetitions)	(3.53%)	(10.91%)	(5.98%)	(8.94%)	(7.37%)
[300 repetitions]	[3.58%]	[11.54%]	[5.84%]	[8.87%]	[7.20%]
{1100 repetitions}	{3.66%}	{11.10%}	{5.90%}	{8.76%}	{7.35%}
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

Estimator	Overall	Dropouts	High_school	Some_college	College Grads
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%)
Nearest neighbor (one)	**0.52%**	**-12.23%**	**2.46%**	**-8.76%**	**9.07%**
(300 repetitions)	(3.48%)	(11.56%)	(5.93%)	(8.81%)	(6.82%)

Panel C: Trimming level q = 7

Estimator	Overall	Dropouts	High_school	Some_college	College Grads
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%)
Nearest neighbor (one)	**-1.20%**	**-12.23%**	**2.46%**	**-8.34%**	**5.00%**
(300 repetitions)	(3.68%)	(11.05%)	(5.99%)	(8.92%)	(7.48%)

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	**-0.56%**	**-12.46%**	**-4.73%**	**-0.75%**	**10.20%**
(300 repetitions)	(2.32%)	(7.25%)	(3.65%)	(5.66%)	(5.06%)
Nearest neighbor (one)	**-0.84%**	**-12.23%**	**2.46%**	**-8.34%**	**5.28%**
(300 repetitions)	(3.58%)	(11.54%)	(5.84%)	(8.87%)	(7.20%)

Panel B: Model II

Local linear 25% bandwidth	**-0.32%**	**-12.20%**	**-4.79%**	**0.03%**	**10.42%**
(300 repetitions)	(2.46%)	(7.10%)	(3.96%)	(5.61%)	(5.35%)
Nearest neighbor (one)	**-2.21%**	**-19.08%**	**-4.56%**	**4.25%**	**4.07%**
(300 repetitions)	(3.91%)	(11.22%)	(5.81%)	(8.81%)	(8.08%)

Panel C: Model III

Local linear 25% bandwidth	**-0.02%**	**-12.80%**	**-4.12%**	**-1.48%**	**12.03%**
(300 repetitions)	(2.51%)	(7.39%)	(4.14%)	(5.66%)	(5.16%)
Nearest neighbor (one)	**-2.09%**	**-9.01%**	**-1.73%**	**-9.78%**	**5.94%**
(300 repetitions)	(3.76%)	(10.87%)	(6.32%)	(9.75%)	(7.84%)

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	Minimum	Maximum
Distance-Based Measure	Mover	378	535	20	3772
	Stayer	1700	4	0	49
Change-of-State Measure	Mover	258	722	1	3772
	Stayer	1820	12	0	668
Change-of-County Measure	Mover	542	379	1	3772
	Stayer	1536	2	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	Minimum	Maximum
Change-of-State Measure	Undercounts of Movers	136	120	20	668
	Overcounts of Movers	16	16	1	44
Change-of-County Measure	Undercounts of Movers	0	-	-	-
	Overcounts of Movers	164	17	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	**-0.56%**	**-12.46%**	**-4.73%**	**-0.75%**	**10.20%**
(300 repetitions)	(2.32%)	(7.25%)	(3.65%)	(5.66%)	(5.06%)

Panel B: Change-of-County Measure

Estimator	Overall	Dropouts	High_school	Some_college	College Grads
Local linear 25% bandwidth	**-1.98%**	**-7.35%**	**-6.23%**	**-0.48%**	**7.27%**
(300 repetitions)	(2.05%)	(5.70%)	(2.98%)	(5.20%)	(4.81%)

Panel C: Change-of-State Measure

Estimator	Overall	Dropouts	High_school	Some_college	College Grads
Local linear 25% bandwidth	**0.03%**	**-7.13%**	**-4.71%**	**-1.67%**	**8.59%**
(300 repetitions)	(2.91%)	(8.49%)	(5.41%)	(6.86%)	(6.70%)

Note: The baseline model , q = 5 trimming level.

Figure 1: Distributions of Estimated Propensity Score

(Based on Model I)