The Impact of Information on Migration Outcomes

Berna Demiralp

Old Dominion University

October 2010

Abstract

This paper presents a model of migration in which migration decisions are made with incomplete information regarding the destination. It explains the conditions under which greater access to information about the destination can lead to positive, negative or no effect on the probability of migration and the return to migration. The empirical section investigates the impact of social interactions, which convey information about destinations to individuals, on migration outcomes. The results suggest that greater information about destination increases one's probability of migration and return to migration when migration is defined by moving to another state within the past five years.

1 Introduction

Since Sjaastad's influential work, migration has been perceived as an investment in human capital (1962). Economists who study the return to this investment often examine the wage and earnings growth that are experienced by migrants as a result of their migration decisions. While the human capital theory of migration predicts that the present discounted value of lifetime earnings at the destination exceeds the present discounted value of lifetime earnings at home, it is silent on the direction of the more immediate wage or earnings growth due to migration. Empirical literature has also failed to produce a consensus on this topic as findings of positive, negative and insignificant returns to migration, calculated as contemporary wage or earnings growth, have been reported in the literature[1]. One of the factors that can account for such varying realizations in return to migration is the level of information individuals have about the destination and their opportunities in the destination labor markets prior to migration.

The purpose of this paper is to investigate how incomplete information regarding destination labor market conditions influences the decision to migrate and the associated return to migration. I start with the premise that individuals have incomplete information about their destination wages at the time of their migration decisions. Differences in the level of access to such information within the population can bring about differences in economic outcomes related to migration, such as the migration propensity and the return to migration. For example, individuals with better access to information about destination wages may more accurately estimate their post-migration outcomes and experience positive wage growth due to migration while those with higher uncertainty about their destination wages may overestimate their post-migration wages and thus experience negative wage growth. The question then is whether migrants with better information experience higher returns to their migration decisions. This paper addresses this question both theoretically and empirically. The theoretical model examines the implications of information on one's likelihood to migrate and wage growth due to migration. The empirical application recognizes that one of the most important channels through which people obtain information on various destinations is their social interactions with friends and family. Based on the theoretical framework, the empirical application then investigates the effect of social interactions on migration outcomes.

---

[1]Ham et al. provide an excellent review of the empirical literature on the contemporary wage and earnings change due to migration (2006). As summarized in their paper, Polachek and Horvath (1977), Borjas, Bronars and Trejo (1992), Tunali (2000) and Ham, Li and Reagan (2006) have found negative returns to migration while insignificant returns have been found by Bartel (1979), Hunt and Kau (1985), and Yankow (2003) for different migrant groups. Bartel (1979), Hunt and Kau (1985) and Yankow (2003) also report positive returns for other migrant subsamples.

It has been acknowledged in previous research that information plays an important role in people's migration decisions by directly affecting their expected benefits from migration. Many studies, which have empirically studied the impact of incomplete information on migration behavior, have concluded that information is a determinant of various migration-related outcomes, including migration propensity (Greenwood, 1975; DaVanzo, 1976; Allen, 1979), return migration (DaVanzo and Morrison, 1981; Allen, 1979), post-move earnings growth (Kau and Sirmans, 1977), and job search duration after the move (Gibbs, 1994). The theoretical foundation of most of this empirical work is rooted in the job search model. In their paper, Herzog, Hofler and Schlottmann (1985) emphasize the link between the job search model and migration under incomplete information, and they use the findings of the job search literature in developing a migration model with incomplete information. Their model assumes that greater labor market information increases actual post-migration wages; therefore, actual wages under incomplete information are less than the potential wages that people would earn under perfect information. Berninghaus and Seifert-Vogt (1987) present a migration model that is based on the sequential nature of the job search process. They assume that individuals compare income draws at various destinations and choose a destination based on the results of their comparisons. After the move, they conduct a search to find a job. A prediction of their model is that greater uncertainty about the destination labor market increases one's probability of migration.

In this paper, I present an alternative model of migration under incomplete information. The model is based on the assumption that there is a random component of destination wages that is not perfectly observable by the individual prior to migration. Furthermore, the individual does not know the population distribution of this random component, so he/she cannot use its population mean in predicting his/her post-migration wages. Instead, he/she receives a random sample of n draws from the population distribution of the random variable and uses the sample mean of n observations to predict his/her destination wages. The number of observations that the individual receives, n, increases with the level of his/her information about the destination. Since the variance of the sample mean distribution decreases with n, information affects the worker's migration decision by changing the spread of the distribution of the sample mean. The information acquisition process utilized in this model was initially proposed by Allen and Eaton, who studied the effect of information about a destination on the rate of migration to that destination (2005). In this paper, I focus on the role of incomplete information in bringing about the variation in the return to migration that is observed in the data.

The primary contribution of the theoretical section of this paper is to show that information about a des-

tination can have positive, negative or no effect on the probability of migration and the return to migration. Thus, the model can be used to explain a wide set of empirical findings regarding the relationship between information and migration outcomes. This flexibility of the model is based on the fact that it allows for both underprediction and overprediction of destination wages in the population as demonstrated in the following result. According to the implications of the theoretical framework, the effect of information on the probability of migration hinges on the difference between the population mean of wages at home and that at the destination. If the population mean of wages at the origin exceeds that of the wages at destination and the migration costs, increased information regarding destination labor market conditions is likely to change the migration decision of those who used to overestimate their post-migration wages. The withdrawal of these individuals from the migrant pool leads to a decrease in the probability of migration. If the population mean of wages at the origin is lower than that at the destination, net of migration costs, then access to more information is likely to increase the migration probability of those who used to underestimate their post-migration wages.

Regarding the return to migration, the model predicts that information about destination wages affects the return to migration through two channels, which I name as selection and heterogeneity effects. The selection effect reflects the impact of information on the return to migration through its effect on the composition of the migrant sample. A positive selection effect results when increased information leads to a withdrawal of overestimators from the migrant pool. As the migrant pool is made up of a smaller proportion of people who used to overestimate their destination wages, the average return to migration among migrants increases. A negative selection effect, on the other hand, is brought about when greater information generates a surge in migrants who used to underestimate their post-migration wages. As the migrant pool comprises of a greater proportion of people who used to underestimate their destination wages, the average return to migration among migrants decreases. The heterogeneity effect describes the effect of information on post-migration wage growth through its impact on the degree of heterogeneity among migrants. As a result of more information, variance within the migrant sample decreases. This decrease in migrant heterogeneity positively affects the return to migration when the migrant sample has an influx of overestimators as well as when it has a withdrawal of overestimators. The net effect of information on wage growth due to migration depends on the relative sizes of these two effects.

Another interesting implication of the theoretical model is its prediction that migrants on average overestimate their post-migration wages. This result provides an explanation for the observation made in earlier

research that "migration should select against those who underestimate the net returns to migration and attract those who overestimate them" (DaVanzo, 1983). The prevalence of overprediction of post-migration wages among migrants can explain the negative return to migration found in previous empirical research. In a setting of incomplete information, individuals, who overestimate their destination wages prior to migration, are more likely to experience negative returns to migration. The model also predicts that the expected value of the prediction error, the difference between the predicted and actual post-migration wages, in the entire population is zero. Therefore, the implication that migrants on average experience positive prediction error does not depend on a restrictive assumption such as a positive support for the prediction error in the population.

The theoretical model forms the basis for the empirical work, which investigates how information about destination labor markets affects individuals' decisions to migrate to another state within the U.S. and the wage gain associated with such moves. The econometric model is specified as a switching regression model where the migration decision determines the regime, and the wages are the economic outcomes of interest. In the empirical section, I draw on the results of previous research regarding the role of network externalities in migration, which indicate that people gain information about destinations through their interactions with family and friends who have already migrated (Radu (2008) provides a review of the literature on social interactions in migration). Based on this premise, I use state outmigration rate to identify the intensity of information regarding various destinations that is passed on to the residents of the home state. The econometric challenges involved in identifying the effect of group behavior (in this case, state outmigration rate) on individual behavior (individual's migration choice) have been documented in the economic literature (Manski, 1993). In addressing these challenges, I use the proportion of females with two or more children in a state whose first two children have different sexes as an instrumental variable in the analysis. The choice of this instrumental variable exploits a widely reported observation that parents prefer a balanced sex composition of their children[2]. The instrument is correlated with state outmigration rate because it is an indicator of family size at the state level. However, it is not correlated with an individual's migration decision.

The analyses are conducted using data from the March supplement to the 2005 Current Population Survey and the U.S. Census 2000. The estimation results yield that the state outmigration rate has a positive

---

[2]Westoff et al. (1963), Ben-Porath and Welch (1976) and Pebley and Westoff (1982) have found that parents with two boys or two girls are more likely to have a third child than parents with one boy and one girl. Williamson (1983) provides a review of studies on parent's preference regarding the sex of their children.

effect on one's probability of moving out of state with the last five years. Using the same definition of migration, I find that migrants whose home states have low outmigration rates experience a negative return to migration on average, and those coming from home states with high outmigration rates experience a positive average return to their migration decisions. This result suggests that information about destination labor markets may lead one to realize higher wage growth due to migration. When migration is defined by movement out of state within the past year, the effect of the state outmigration rate on the probability of migration becomes statistically insignificant. Furthermore, the return to migration is higher among migrants from states with low outmigration rates when the migrant sample is defined by state-to-state movement within the last year.

The paper is divided into two main sections. Section II presents the theoretical model and discusses its implications with respect to the probability of migration and the return to migration. Section III presents the econometric model, explains the empirical strategy and data and discusses the results of the analysis. Concluding remarks are given in Section IV.

2 Theoretical Model

2.1 The Framework

Individuals make a decision between moving ($M = 1$) and staying ($M = 0$) based on their current wages at the origin, their expected wages at the destination and their moving costs. They decide to move if they anticipate their wages at destination to be higher than the sum of their current wages and the moving costs. Consider individual i who has two wage alternatives: y_{1i} if he/she migrates and y_{0i} if he/she stays. These alternatives are given by

$$y_{0i} = \mu_0 + \upsilon_{0i} \tag{1}$$

$$y_{1i} = \mu_1 + \upsilon_{1i} + \varepsilon_i \tag{2}$$

υ_0 and υ_1 are normally distributed with means zero while ε follows a normal distribution with mean μ_ε in the population. Therefore, μ_0 and $\mu_1 + \mu_\varepsilon$ indicate the population means of wages at the origin and the destination, respectively, and υ_{0i} and υ_{1i} represent individual i's deviations from these means. While the individual can perfectly observe $\mu_0, \mu_1, \upsilon_{0i}$ and υ_{1i}, he/she cannot observe ε_i at the time of his/her migration decision. Furthermore, I assume that the individual has only partial information about the distribution of

ε. In particular, I apply the information acquisition process proposed by Allen and Eaton in their migration model to this framework (2005)[3].

The basic premise of Allen and Eaton's model, when applied to the current framework, is that the individual does not have any priors about the expected value of ε in the population. Therefore, he/she samples n random and independent draws from the population distribution of ε and uses the average of n draws as a predictor for the expected value of ε. Let $\overline{\varepsilon_i}$ be the average of the n random draws observed by individual i. Then, the individual anticipates his/her wages at the destination to be y_{1i}^e where

$$y_{1i}^e = \mu_1 + \overline{\varepsilon_i} + \upsilon_{1i} \tag{3}$$

The one-time moving cost faced by individual i is given by

$$c_i = \mu_c + \upsilon_{ci} \tag{4}$$

where both μ_c and υ_{ci} are known by the individual at the time of his/her migration decision. υ_c is normally distributed with mean zero. Then, the individual i's migration decision can be characterized as

$$
\begin{aligned}
M &= 1 \ \ if \ \ y_i^* > 0 \\
M &= 0 \ \ otherwise
\end{aligned}
\tag{5}
$$

where $y^* = y_{1i}^e - y_{0i} - c_i$.

Note that y_{1i} represents the *actual* post-migration wages while y_{1i}^e represents individual i's *anticipation* of his/her post-migration wages at the time of the migration decision. Furthermore, the random components, υ_0, υ_1, and υ_c are observed by the individual but are unknown to the researcher while ε is unknown to both the individual and the researcher. In order to be able to further pursue the implications of the model, I assume that υ_0, υ_1, υ_c and ε follow a multivariate normal distribution with covariance matrix, Σ[4].

---

[3]Allen and Eaton's model focuses on the effect that information about destination has on migration propensity (2005). The purpose of this paper includes studying the effect of information on the return to migration. This difference in focus brings about differences in the set-up of the two models. For example, in Allen and Eaton's model, all individuals in a given origin have identical expectations of their future earnings at origin. In the model presented here, individuals' expectation of their future home earnings is a random variable that follows a continuous probability distribution.

[4]$\Sigma = \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{0c} & \sigma_{0\varepsilon} \\ \sigma_{01} & \sigma_1^2 & \sigma_{1c} & \sigma_{1\varepsilon} \\ \sigma_{0c} & \sigma_{1c} & \sigma_c^2 & \sigma_{c\varepsilon} \\ \sigma_{0\varepsilon} & \sigma_{1\varepsilon} & \sigma_{c\varepsilon} & \sigma_\varepsilon^2 \end{bmatrix}$

Before moving on to the implications of the model, I would like to point out three observations. First, the size of the sample drawn from the ε distribution, n, indicates the individual's level of information regarding the destination labor market. Thus, the effect of information on migration-related outcomes, such as the migration propensity and the return to migration, can be studied by analyzing the effect of n on these outcomes. In order to observe the immediate impact of n in this model, one should note the characteristics of the distribution of the sample mean, $\bar{\varepsilon}$. Since ε is distributed normally in the population, $\bar{\varepsilon}$ also follows a normal distribution where $\bar{\varepsilon} \sim N(\mu_\varepsilon, \frac{\sigma_\varepsilon^2}{n})$. As individuals have access to more information about the destination labor market, represented by an increase in n, the variance of $\bar{\varepsilon}$ decreases and the individuals' $\bar{\varepsilon}$ draws become more concentrated around the actual population mean of ε, which is assumed to be μ_ε. As n increases to infinity, individuals move from a situation of incomplete information towards one of complete information, in which they know the expected value of ε in the population and can use it in predicting their post-migration wages.

Second, the characterization of incomplete information in this model differs from a standard migration model with uncertainty, in which the individual does not observe ε_i prior to migration but has complete information on the population distribution of ε. In that case, the individual does not need to sample ε since he/she uses the expected value of ε in the population to predict his/her destination wage as follows: $y_{1i}^e = \mu_1 + \upsilon_{1i} + \mu_\varepsilon$. Furthermore, in such a model, variance of ε, which indicates the level of uncertainty surrounding post-migration wages, has no effect on one's migration propensity since the individual knows μ_ε. In this model, variance of ε influences migration propensity as explained below.

Third, this model allows for both underprediction and overprediction of post-migration wages by individuals depending on their ε_i and $\bar{\varepsilon_i}$ draws. An individual overpredicts his/her post-migration wage if $\bar{\varepsilon_i} > \varepsilon_i$. In that case, the individual makes a positive prediction error since $\bar{\varepsilon_i} - \varepsilon_i > 0$. Similarly, an individual underpredicts his/her post-migration wage if $\bar{\varepsilon_i} < \varepsilon_i$. In that case, $\bar{\varepsilon_i} - \varepsilon_i < 0$, and the individual's prediction error is negative.

2.2 The Probability of Migration

Based on the migration decision characterized by Equation 5, the probability that a randomly chosen individual chooses to migrate equals the following:

$$P = \Pr(\eta > \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon) = 1 - \Phi(z) \tag{6}$$

where $\eta = v_1 + \overline{\varepsilon} - v_0 - v_c$, $z = \frac{\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon}{\sigma_\eta}$, and Φ is the cdf of a standard normal distribution. Then, the effect of n on the probability of migration is given by

$$\frac{\partial P}{\partial n} = -\phi(z) \cdot \frac{\partial z}{\partial \sigma_\eta} \cdot \frac{\partial \sigma_\eta}{\partial n} \tag{7}$$

As shown in the Appendix, the sign of $\frac{\partial P}{\partial n}$ depends on the sign of $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon$. In particular,

$$\frac{\partial P}{\partial n} \gtreqless 0 \ \ if \ \ \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon \lesseqgtr 0 \tag{8}$$

According to this result, as people have more information about the destination labor market conditions, the probability of migration moves in the direction of higher expected wages in population. If the population mean of wages at the origin (μ_0) is higher (lower) than the population mean of wages at the destination $(\mu_1 + \mu_\varepsilon)$ minus the moving cost (μ_c), then the probability of migration decreases (increases) as people become more informed about their labor market prospects at the destination. The intuition behind this result can be explained as follows: Suppose that the population mean of wages at origin is smaller than the population mean of wages at the destination minus the moving cost. Then as n increases and individuals' $\overline{\varepsilon}$ draws move closer to μ_ε, people who are likely to change their migration decisions are those who initially had low $\overline{\varepsilon}$ values. As their $\overline{\varepsilon}$ draws become closer to μ_ε after the increase in n, they become more likely to choose migration, thus leading to an increase in the migration rate. The opposite result holds when the population mean of wages at origin is greater than the population mean of wages at the destination minus the moving cost. In that case, those who are likely to change their migration decisions are those who initially had high $\overline{\varepsilon}$ values and decided to migrate. After n increases and their $\overline{\varepsilon}$ draws approach μ_ε, these individuals become less likely to choose migration, leading to a decrease in the probability of migration in the population.

Hypothesis 1 The individual's level of information regarding the destination labor market affects the probability of migration positively if $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0$ and negatively if $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0$.

2.3 The Return to Migration

The average return to migration, R, is defined as the difference between what migrants earn at the destination and what they would have earned at home had they stayed. Mathematically,

$$R = E(y_1|M=1) - E(y_0|M=1) \tag{9}$$

where $E(y_1|M=1)$ gives the expected value of migrants' wages at the destination, and $E(y_0|M=1)$ gives the expected value of migrants' wages at the origin. If we let $\omega = v_1 + \varepsilon$, $\rho_{\omega\eta} = Corr(\omega, \eta)$, $\lambda_1 = \frac{\phi(z)}{1-\Phi(z)}$ and $\rho_{0\eta} = Corr(v_0, \eta)$, then the average return to migration can be stated as

$$R = \mu_1 + \mu_\varepsilon - \mu_0 + \left(\rho_{\omega\eta}\sigma_\omega - \rho_{0\eta}\sigma_0\right)\lambda_1 \tag{10}$$

In order to examine how information affects the return to migration, I calculate $\frac{\partial R}{\partial n}$. As shown in the Appendix, this derivative can be expressed as

$$\frac{\partial R}{\partial n} = \left[\frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial n} - \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial n}\right] Var(v_1 - v_0) \tag{11}$$

when $\sigma_{0c} = \sigma_{1c} = \sigma_{0\varepsilon} = \sigma_{1\varepsilon}$ and $\sigma_{c\varepsilon} = 0$. As a result, the sign of $\frac{\partial R}{\partial n}$ is determined by the sign of $\left[\frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial n} - \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial n}\right]$.

The first term in the brackets describe the effect of n on R through its impact on the selection of the migrant sample. This selection effect is positive if $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0$ and negative if $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0$. Information about the destination has a positive effect on the selection of migrants when $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0$ because in that case the marginal individual, who changes his/her migration decision as a result of more information in that case, is one who had previously overestimated his/her destination wages. As a portion of the overestimators change their migration decisions from migration to staying, the migrant pool comprises of a smaller proportion of people who make negative prediction errors. As a result, the average post-migration wage among migrants increases, and the average return to migration rises. Therefore, a positive effect reflects the fact that as a smaller portion of migrants overpredict their destination wages, the return to migration rises.

On the other hand, information on the destination labor market has a negative effect on the selection of

the migrant sample when $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0$. In that case, the marginal individual, who changes his/her migration decision with more information, is one who had previously underestimated his/her post-migration wages. When more information is available, underestimators of post-migration wages are likely to choose migration, bringing about an injection of new migrants into the migrant pool and a higher migration rate. These new migrants are likely to come from the lower tail of the destination wage distribution because people from the higher tail would have most likely chosen migration initially even if they had underestimated their future wages. The injection of new migrants into the migrant pool from the lower tail of the destination wage distribution brings about a negative impact on the selection of the migrant sample.

The second term in the bracket in Equation 11 can be perceived as the effect of n on R through its impact on the heterogeneity of the migrant sample. If $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0$, the migrant sample consists of a greater proportion of people who have higher destination wages than anticipated, putting an upward pressure on R. If $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0$, the migrant sample consists of a smaller proportion of people who have lower destination wages than anticipated, again generating a positive effect on R. Therefore, this effect is unambiguously positive. As people have more information about the destination labor market, the migrant sample consists of a greater percentage of people from the upper tail of the y_1 distribution, bringing about an increase in the return to migration.

The net effect of n on R then depends on the sum of the effects on the selection of the migrant sample and the level of heterogeneity in the migrant sample. When $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0$, both effects are positive, generating a positive net effect of information on the return to migration. When $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0$, the selection effect is positive, and the heterogeneity effect is negative; thus the sign of $\frac{\partial R}{\partial n}$ depends on the relative sizes of the two effects. In sum,

$$
\begin{aligned}
\frac{\partial R}{\partial n} &< 0 \;\; if \;\; \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0 \;\; and \;\; \left| \frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial n} \right| > \left| \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial n} \right| \\
\frac{\partial R}{\partial n} &> 0 \;\; otherwise
\end{aligned}
\tag{12}
$$

Hypothesis 2 Information positively affects the return to migration if $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0$.

Hypothesis 3 If information negatively affects the return to migration, then it positively affects the migration rate since both conditions hold under $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0$.

Next, I consider how an increase in the variance of ε affects the average return to migration. Within the

framework of this model, σ_ε captures the uncertainty faced by an individual regarding his/her destination wages. Based on Equation 2, σ_ε directly affects the variance of wages at the destination. Therefore, the effect of σ_ε on the return to migration provides insight on how the variance of wages at destination impacts the average return to migration. It provides an explanation for what happens to the average return to migration when the wage distribution at the destination becomes more unequal.

Mathematically, the effect of σ_ε on the return to migration is given by[5]

$$\frac{\partial R}{\partial \sigma_\varepsilon} = \left[\frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial \sigma_\varepsilon} - \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} \right] Var(v_1 - v_0) \tag{13}$$

This net effect can also be decomposed into a selection and a heterogeneity effect. It can be shown that the effect of σ_ε on the heterogeneity among migrants, given by the second term in the brackets, is negative. As σ_ε goes up and the uncertainty that one faces in his/her destination earning increases, conditional on n, the individual's prediction of his/her post-migration wages becomes less precise, thus negatively affecting his/her return to migration. An increase in σ_ε also impacts the selection of migrants revealed by the first term in the brackets. The impact on selection can be positive or negative depending on the relative values of $\mu_0, \mu_c, \mu_1,$ and μ_ε. If $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0$, an increase in the variance of post-migration wages causes the sample of migrants to consist of a smaller proportion of people who overpredict their actual post-migration wages. In this case, the selection effect is positive. On the other hand, if $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0$, an increase in the variance of post-migration wages brings about a change in the sample of migrants so that a greater proportion of migrants overpredict their post-migration wages. As a result, the selection effect associated with an increase in σ_ε is negative. The net effect of a change in σ_ε on R depends on the following:

$$\begin{aligned}
\frac{\partial R}{\partial \sigma_\varepsilon} &> 0 \ if \ \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0 \ and \ \frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial \sigma_\varepsilon} > \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} \\
\frac{\partial R}{\partial \sigma_\varepsilon} &< 0 \ otherwise
\end{aligned} \tag{14}$$

The discussion above can be used to generate the following two hypotheses.

Hypothesis 4 If $\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0$, a higher variance of wages at the destination brings about a lower return to migration.

---

[5]See the Appendix for the derivation of $\frac{\partial R}{\partial \sigma_\varepsilon}$.

Hypothesis 5 If the variance of wages at the destination positively affects the return to migration, then

$$\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0.$$

Before starting the empirical application, I would like to discuss the implications of this model on the average prediction error among migrants. As shown in the Appendix, the model implies that the expected value of the prediction error among migrants ($E(y_1^e|M = 1) - E(y_1|M = 1)$) is positive. Therefore, the average migrant overestimates his/her destination wages. Based on this implication, the model provides an explanation for the observation made in earlier research that "migration should select against those who underestimate the net returns to migration and attract those who overestimate them" (DaVanzo, 1983). It is important to note that this theoretical result is not contingent on restrictive assumptions about the distribution of the prediction error in the population. In fact, the model allows for both over- and underestimation of destination wages, and it generates a positive prediction error among migrants even as the expected value of the prediction error in the population is zero ($E(\overline{\varepsilon_i} - \varepsilon_i) = 0$). This implication distinguishes the model presented here from earlier models by Herzog et al. (1985) and Daneshvary et al. (1992), which also conclude that migrants on average overestimate their destination wages. In these earlier models, however, the positive prediction error among migrants is contingent on the assumption that the prediction error has a positive support over the entire population. They assume that reservation wages are monotonically increasing over the level of information; thus everyone in the population underestimates their actual post-migration wages, leading to a positive expected value of the prediction error in the population. Furthermore, the model presented here implies that as n goes to infinity, λ_1 and σ_η approach constant values, $\frac{\sigma_\varepsilon^2}{n}$ approaches zero; hence the expected value of the prediction error among migrants ($E(y_1^e|M = 1) - E(y_1|M = 1)$) also approaches zero. Intuitively, as individuals approach having complete information, the average prediction error among migrants goes to zero.

The implication that the average prediction error among migrants is positive provides an explanation for the negative return to migration found in previous empirical research. As stated in the Introduction, several studies have found negative return to migration among migrants. For example, Tunali finds that about 75 percent of migrants in his sample realize negative returns to migration (2000). One of the explanations for the negative return to migration is that migrants overestimate their post-migration wages, only to realize after migration that their actual post-migration wages are less than their wages at the origin. By showing that the overestimation of destination wages is prevalent among migrants, the model presented here provides an

explanation for the negative return to migration within a human capital investment framework of migration decisions under incomplete information.

3   Empirical Application

3.1   Social Interactions in Migration Decisions

In this section, I aim to empirically examine the effect that greater access to information has on migration outcomes by estimating reduced forms of the behavioral model discussed above. A crucial part of this strategy lies in distinguishing between individuals facing different costs of information and thus different levels of access to information regarding destination labor markets. In other words, how to proxy for access to information about the destination?

One of the channels through which people obtain information about other regional labor markets is their social interaction with friends and neighbors who have already migrated to these regions. One can learn about various characteristics of the destination labor market, such as job openings, income distribution, average earnings, by talking to friends, family and neighbors who have migrated to the destination. This intuition implies that people with migrant friends have more access to information about destination labor markets, holding everything else constant. Therefore, one way to study the effect of information about the destination on migration outcomes is to study how one's migration choices are affected by others' (neighbors' and friends') migration choices. In other words, one should study the influence of social interactions on individuals' migration outcomes.

That social interactions can play an important role in migration decisions has been pointed out in previous literature[6]. Theoretical developments in this literature incorporate social dynamics in various aspects of the migration decision. For instance, Carrington et al. present a migration model in which moving costs are inversely related to the number of immigrants in the destination (1996). They use the results of their theoretical model to explain why the Great Black Migration from the South to the North took place during a time when the income gap between the two regions was narrowing. Spilimbergo and Ubeda specify social interaction in their migration model, based on the assumption that one's family and friend network at home might discourage him/her from migrating (2004). They find multiple equilibria and use the existence

---

[6]Radu (2008) provides a literature review of social interactions in migration research. For a broader review of neigborhood effects, see Ioannides and Topa (2010).

of multiple equilibria to explain why different groups have persistently exhibited different migration rates (e.g. White versus African-American, U.S. versus Europe). Social interactions are also implicit in migration models with network externalities (See Bauer, Epstein and Gang (2007) for more on network externalities in migration). On the empirical front, many studies show that having family or friends at the destination is positively related to the probability of migrating to that destination (e.g. Caces et al. 1985, Taylor 1986). More recently, Munshi investigates the role of Mexican migrant networks in determining Mexican migration into the U.S. (2003). Chen et al. empirically study the effect of social interactions on domestic migration in China (2010).

Empirical studies aimed at identifying social effects in migration, or in any context, face certain challenges. In order to better describe the nature of these challenges and explain how the present empirical strategy addresses them, I will use Manski's terminology and distinguish between three types of interaction (1993). First, one can study the *endogenous* social effects, which arise when an individual's own decision depends on the decisions of those in his/her reference group. The endogenous social effect in this study is the effect that the groups members' migration decisions have on the individual's migration decision, and it is the primary point of interest. However, an observed correlation between the group migration decisions and the individual's migration decision does not necessarily imply that the individual's migration behavior is affected by the group's migration choices. Such an observed correlation between individual and group outcomes may reflect *correlated* effects, which refer to similarities between individual and group outcomes that are due to the fact that the individual and the reference group share similar unobserved characteristics or similar institutional constraints. Even if the correlated effects are controlled, it may be impossible to separate the effect of the *group's migration choices* on the individual's migration decision (*endogenous* social effect) from the effect of the *group's characteristics* on the individual's migration decision in certain situations. The latter effect is the *contextual* effect, and this econometric problem is termed as the reflection problem by Manski (1993). The methodological challenge lies in empirically distinguishing between these three types of interactions.

In order to separately identify the endogenous social effects on one's migration decision, I use the instrumental variable approach. This method relies on finding an instrumental variable that is correlated with the group members' migration choices but does not affect the individual's migration decision except through its effect on the group migration decisions. The instrumental variable chosen for the analysis is the percentage of females with two or more children in the reference group whose first two children have

15

different sexes. This instrument is chosen based on two findings in the literature. First, it has been widely observed that parents with two children of the same sex are substantially more likely to have a third child compared to parents with one boy and one girl (Westoff et al. (1963), Ben-Porath and Welch (1976), Pebley and Westoff (1982)). This finding indicates that parents prefer a balanced sex composition of their offsprings. It further indicates that the sex mix of the first two children can be used as an instrument for future fertility among families with two or more children. Second, research on migration has shown that birth of children is one of several life-cycle considerations that play an important role in migration decisions (Greenwood, 1997). These two findings combined point to the conclusion that the sex composition of the first two children affects one's migration choice through its impact on the individual's family size. The implication is that at the group level, the percentage of people in the group whose first two children have different sexes is positively correlated with the group's migration propensity. Furthermore, since sex composition of children is a random phenomenon, the prevalence of same sex offsprings among members an of individual's group should have no effect on the individual's migration decision except through its impact on the group's migration behavior.

The observation that parents prefer equal numbers of sons and daughters has been exploited in several previous studies. Maurin and Moschion use the proportion of same sex sibling families in one's neighborhood to isolate the endogenous social effects in the labor force participation of mothers in France (2009). Angrist and Evans use a dummy variable for whether the two oldest children of a woman have the same sex as an instrumental variable to identify the effect of fertility on female labor force participation (1998). Chen et al. examine social effects in domestic migration in China using a similar identification strategy. They use China's one-child policy and use a dummy variable for whether the first born is a female as an instrument for the migration prevalence in one's village (2010).

3.2 Model Specification

Based on the theoretical framework, the econometric model can be expressed as a switching regression model with the following specification:

$$y_{0i} = X_i\beta_0 + u_{0i} \tag{15}$$

$$y_{1i} = X_i\beta_1 + u_{1i} \tag{16}$$

$$y_{1i}^e = X_i\beta_e + \beta_g g_i + u_{ei} \tag{17}$$

$$c_i = X_i\delta_x + Z_i\delta_z + u_{ci} \tag{18}$$

$$y_i^* = y_{1i}^e - y_{0i} - c_i = X_i\alpha_x + Z_i\alpha_z + \alpha_g g_i + \epsilon_i \tag{19}$$

$$M_i = 1[y_i^* \geq 0] = f(X_i, Z_i, g_i) \tag{20}$$

$$y_i = M_i y_{1i} + (1 - M_i)y_{0i} \tag{21}$$

y is the observed wages in the sample, X includes observable state level and individual level characteristics that affect earnings (human capital, geographic and occupation variables), and Z includes explanatory variables that affect the cost of moving but have no effect on earnings (number of children in the household, whether the individual owns his/her dwelling). g_i denotes the group members' migration choices and is specified as the outmigration rate in individual $i's$ state. The error terms, u_0, u_1 and ϵ, are assumed to be independent of X and Z, and they are assumed to follow a trivariate normal distribution with zero expectations and the positive definite covariance matrix given below:

$$
\begin{bmatrix}
\sigma_0^2 & \sigma_{01} & \sigma_{0\epsilon} \\
\sigma_{01} & \sigma_1^2 & \sigma_{1\epsilon} \\
\sigma_{0\epsilon} & \sigma_{1\epsilon} & \sigma_\epsilon^2
\end{bmatrix}
\tag{22}
$$

While X and Z contain exogenous explanatory variables, g is likely to be endogenous due to the reasons explained above. In order to account for the endogeneity of g, the migration equation is estimated using the instrumental variable approach.

The earnings equation is estimated using the two-step method developed by Heckman (1974, 1976, 1978) and Lee (1978, 1979). This estimation method takes into account the self-selection of migrants, which may result in migrants being systematically different than non-migrants in terms of unobservable characteristics (i.e. the estimation method allows for the condition that $Cov(u_0, \epsilon) \neq 0$ and/or $Cov(u_1, \epsilon) \neq 0$). In the first stage, the migration equation (Equation 20) is estimated using a probit regression, and the estimated parameters are used to generate the Inverse Mill's Ratio for every observation. In the second stage, the calculated Inverse Mill's Ratio is added to the earnings equation (Equation 21) as a regressor.

3.3 Data

The individual-level data including demographic, labor market and migration variables come from the March supplement to the 2005 Current Population Survey (CPS). The extraction of the data was performed using the IPUMS-CPS, which is an integrated set of the March CPS from 1962-2006 (King et al., 2004). One of the advantages of using the 2005 CPS is that it allows for the creation of two different migration variables. The first variable, *mig1*, indicates whether the individual is living in a different state than he/she did a year ago, and the second variable, *mig5*, indicates whether the individual is living in a different state than he/she did five years ago. Individuals who moved to another state and returned to their home state within the past year are considered non-migrants according to the *mig1* indicator. Similarly, individuals who have migrated and returned to their home states within the last five years are considered non-migrants according to the *mig5* definition. It is plausible to think that more return migration would occur within a five-year span compared to a one-year span. Since *mig1* and *mig5* generate samples of migrants that differ with respect to the proportion who may return to home state in the future, a comparison of results based on *mig1* and *mig5* enables the researcher to make inferences on the possible effects of "return migration" on several migration outcomes.

Log wages are used as the dependent variable in the earnings equation. As a result, the return to migration can be interpreted as the wage growth due to migration. The hourly wages are calculated by dividing the respondents' earnings by the number of hours worked, and the hourly wage variable included in the CPS is used whenever it is reported by the respondent.

The group migration decision is identified by the outmigration rate in the individual's state of origin. The state outmigration rates are obtained from the calculations performed by the U.S. Census Bureau using the U.S. Census 2000 (Franklin, 2003). These calculations are based on the number of people who reported having moved across states between 1995 and 2000. The state-level data are linked to the individual-level data by the non-migrant's state of residence and the migrant's state of origin.

The instrumental variable is computed by dividing the number of women aged 21-38 whose first two children have different same sexes by the number of women with 21-38 with at least two children in each state. The 1 percent sample of the U.S. Census 2000, extracted via IPUMS-USA, is used in creating this variable. Census does not allow one to link parents to children across households. Therefore, the age restriction is imposed in creating the instrumental variable so that it is more likely that the children living in

the household constitute the female respondent's only children.

The analysis sample is limited to civilians aged 15 or older who are in the labor force. It contains 96,333 observations. 2.58 percent of the sample report living in a different state than they did a year ago, and 8.15 percent report living in a different state than they did five years ago. The list of variables used in the analysis as well as the average characteristics of migrants and stayers are presented in Table 1. These descriptive statistics reveal that migrants on average are younger, less experienced and more educated than stayers. In addition, a smaller proportion of movers are married and own their homes compared to stayers, and people who migrate have on average fewer number of children in the household. Migrants and stayers are also quite different with respect to their employment status. Migrants are more likely to be unemployed during the week of the interview than stayers, and they are less likely to be self-employed after the move relative to stayers. Compared to the baseline residence in the West, migrants are more likely to reside in South and Northeast and less likely to reside in the Midwest. It is not surprising that on average the home states of migrants have higher outmigration rates than those of stayers. The statistically significant differences between average characteristics of migrants and stayers remain when *mig1* is used as the migration indicator.

3.4 Results

3.4.1 The Probability of Migration

Table 2 presents the maximum likelihood estimation results of the migration equation. Specifications (I) and (II) refer to the estimation of ordinary probit regression with and without the outmigration rate in the state of origin as an explanatory variable. In specification (III), the outmigration rate is instrumented using the percentage of women in the state, aged 21-38 with two or more children, whose two oldest children have different sexes. All specifications are estimated for two dependent variables: *mig1* (indicator for the state-to-state migration within the past year) and *mig5* (indicator for the state-to-state migration within the past five years).

The coefficient estimates reported in column 1 indicate that age positively affects the probability of migration while experience has a negative impact on one's probability of migration. Men are more likely to move than women; however race and marital status do not significantly affect the decision to migrate. Number of children and owning a house negatively influence one's likelihood to migrate. This result is consistent with the notion that the number of children and owning a house generate higher moving costs and

thus hinder migration. People who report being unemployed are more likely to have migrated from another state within the past year, suggesting that unemployment can be more prevalent among recent migrants than non-migrants. These results also indicate that the propensity to migrate does not differ significantly between educational groups or across occupational categories.

When *mig5* is used as the dependent variable in specification (I), the above mentioned coefficients retain their signs and significance, and their magnitudes become larger. Thus, the factors determining migration seem to have an even stronger effect on the migration decision when migration is defined by movement across states within the past five years. The difference between the results using *mig1* and *mig5* may be driven by the fact that the *mig1* migrant sample is likely to consist of a greater proportion of individuals who will eventually return to their states of origin. In other words, these results may provide indirect evidence that return migrants are systematically different than permanent migrant. Furthermore, several variables, which exhibit statistically insignificant effects when *mig1* is used as the dependent variable, have statistically significant effects when *mig5* is used as the dependent variable. The coefficient on high school is significantly negative, indicating that high school graduate are less likely to migrate compared to those who have dropped out of high school. In addition, married people have a higher likelihood to move to another state compared to non-married individual, and those living in center city exhibit a lower likelihood of migration when *mig5* is the dependent variable.

The outmigration rate in the state of origin has a positive and statistically significant effect on the probability of migration under both definitions of migration, *mig1* and *mig5*, when the regression is specified as an ordinary probit (Specification (II)). When the outmigration rate is instrumented in Specification (III), the outmigration rate has a statistically insignificant effect in the regression with *mig1*, and it has a positive and significant effect at the 10 percent level in the regression with *mig5*. These result suggests that greater access to information about the destination, proxied by the outmigration rate in the home state, does not change one's probability of migration when migration is defined as movement out of own's state of origin with the last year. When migration is defined by movement out of own's home state within the last five years, greater access to information about the destination seem to increase one's probability of migration. As reported in the bottom panel of Table 2, the Durban-Wu-Hausman test cannot reject the null hypothesis, indicating that the instrumental variable estimation is necessary to obtain unbiased estimates of the causal relationship. In addition, the F-test of the excluded instrument yields F-statistics that are greater than 10, providing evidence for the instrument strength.

3.4.2 The Return to Migration

The wage regression results with *mig1* as the migration indicator reveal that age, college education, being male and married positively affect wages of migrants (Table 3). The concavity of the age-wage profile is more prominent within the stayer subsample as the coefficient on age squared is significantly negative. Experience has a negative and statistically significant effect on the wages of stayers although the effect increases with experience. Similar to the results of the migrant sample, being male, married and college educated leads to higher wages among stayers. When *mig5* is used as the migration indicator, experience has a negative effect on the wages of migrants, and production and service workers have lower wages among migrants. The other mentioned coefficient estimates retain their signs and statistical significance in both the migrant and stayer regressions when *mig5* is used. The coefficient of the added regressor, the estimated Inverse Mill's Ratio, is positive for movers and negative for stayers, indicating that movers experience a positive self-selection effect, and stayers experience a negative self-selection effect. This result suggests that migrants are selected predominantly from the upper tail of the wage distribution while stayers come predominantly from the lower tail of the wage distribution in the population.

Next, I calculate the average return to migration by taking the difference between the wages earned by migrants and the wages that they would have earned at the origin had they chosen not to migrate. The latter is calculated using the coefficient estimates for stayers presented in Table 3. Standard errors are calculated using bootstrap with 1000 repetitions. Since log(wage) is used as the migration outcome in the analysis, the return to migration can be interpreted as the wage growth due to migration. According to the results in Table 4, when migration is defined by movement across states within the past year (*mig1*), the average return to migration among migrants, measured as the difference in log wages, is 0.035. When migration is defined by movement across states within the past five years, the average return to migration among migrants is 0.0073. The median return to migration is very close to the mean under both definitions of migration. In both cases, a significant portion of the migrant sample realize negative returns to migration (30 percent of the migrants based on *mig1* and 42 percent of the migrants based on *mig5*).

Table 4 also presents the average return to migration by different levels of educational attainment. Results show that the average return to migration is highest for high school dropouts. High school graduates gain the lowest wage growth as a result of migration compared to other educational groups. College graduates experience higher returns to migration than high school graduates and lower returns to migration than

high school dropouts. One of the factors that can bring about such a result is the variation in migration costs across educational categories. In particular, higher migration costs among people with low education can generate the observed negative relationship between educational attainment and the return to migration. If high school dropouts face higher migration costs compared to college graduates, then they would require higher return to migration to justify migrating and incurring the associated high migration costs. The observed higher return to migration among high school dropouts can also arise if interstate differences in returns to skills vary across different skill categories. The results in Table 4 are consistent with the hypothesis that interstate wage differentials are higher in jobs that are mostly occupied by high school dropouts. If that is the case, when workers with less than a high school degree migrate to another state, they exploit the large interstate variation in wages and potentially experience a large wage increase due to migration. On the other hand, if jobs that are filled by college graduates have less variation in wages across states, then migrant with college degrees experience smaller return to migration.

The theoretical model presented in this paper has several implications regarding the effect that information about the destination has on the return to migration. In order to empirically test these implications, I compare the average return to migration across migrants based on the outmigration rates in their states of origin. The cut-off migration rates used in the definition of low, middle and high outmigration rate categories correspond to the 25th percentile the 75th percentile observation in the migrant sample. The findings presented in Table 5 suggest that the observed correlation between the outmigration rate in the state of origin and the return to migration is negative among individuals who have moved out of state within the last year. When the migrant sample is restricted to individuals who are living in a different state than they were five years ago, the average return to migration increases with the outmigration rate at the home state. In fact, individuals from states with outmigration rates less than 7.28 percent on average experience a negative return to migration. Although these observed relationship do not necessarily imply a causal effect of outmigration rate on the return to migration, they suggest that information on destinations may be playing a key role in bringing about the negative return to migration found in the current and previous studies.

Finally, I investigate the relationship between income inequality in the destination and the return to migration among migrants. To that end, I calculate the average return to migration for different migrant samples based on the income inequality measures in their destination states. The theoretical model implies that the income inequality in the destination plays an important role in an individual's migration decision as it partially determines the uncertainty surrounding post-migration wages. I use two measures of income

inequality in classifying states into different categories of income inequality: 80:20 income ratio and 95:20 income ratio. 80:20 income ratio is the ratio of income at the 80th percentile to that at the 20th percentile in the income distribution, and the 95:20 income ratio is the ratio of income at the 95th percentile to that at the 20th percentile. These income ratios are calculated by Bernstein et al. (2006) for each state using the 2001-2003 Current Population Survey data. The cut-off ratios used in categorizing destination states into low, middle and high income inequality states, correspond to the 25th percentile and the 75th percentile observation in the migrant sample.

The results presented in Table 6 suggest a positive relationship between income inequality at the destination state and the return to migration. Migrants, who have migrated to states with high income inequality, experience larger wage growth due to migration. Within the framework of the theoretical model, these results provide support for the hypothesis that greater uncertainty about destination wages leads to greater return to migration.

4 Conclusion

The purpose of this paper is to investigate, theoretically and empirically, the impact that information about destination labor markets has on an individual's migration decision and his/her return to migration. To that end, the paper presents a theoretical framework, in which individuals make migration decisions under incomplete information about the destination. The model's implications reveal that information regarding the destination labor market can have positive, negative or no effect on one's probability of migration as well as his/her return to migration. Thus, the model can be used to explain a wide set of empirical findings regarding the relationship between information and migration outcomes.

According to the theoretical model, the effect of information on the probability of migration depends on the population mean of the home and destination wages as well as the moving costs. If the population mean of wages at the origin exceeds the population mean of the wages at destination less the mean moving cost, greater information about the destination labor market conditions leads to lower probability of migration. Otherwise, the probability of migration increases with greater access to such information. Furthermore, information about the destination affects the return to one' migration decision through a selection and a heterogeneity effect. The selection effect reflects the impact of information on the return to migration through its effect on the way that the migrant sample is selected. The heterogeneity effect describes the

effect of information on the return to migration through its impact on the degree of heterogeneity within the migrant sample. While the heterogeneity effect is always positive, the selection effect can be either positive or negative. The net effect of information on the return to migration depends on the sum of these two effects.

The econometric model is specified as a switching regression model where the migration decision determines the regime, and the wages are the economic outcomes of interest. The migration and wage equations are estimated using data from the 2005 Current Population Survey and the U.S. Census 2000. The level of information that an individual has about destination labor markets is indicated by the outmigration rate in one's home state. This specification is based on the premise that people gain information about destinations through their interactions with friends and relatives, who have already moved to those destinations. In order to identify the effect that migration at the state level has on an individual's migration behavior, I use an instrumental variable approach. The instrumental variable used in the analysis is the proportion of females with two or more children in one's home state whose first two children have different sexes. Previous research has shown that this variable is a predictor of future fertility. Since it predicts family size at the state level, this variable is also related to the state migration rate while it is not correlated with an individual's migration choice.

The estimation results show that the state outmigration rate positively affects the probability of migration when migration is defined by movement out of state within the past five years. This result suggests that increased access to information about destinations increases one's probability of migration to another state. Furthermore, the average return to migration is negative among migrants, who come from states with low outmigration rate and positive for those, coming from states with high outmigration rates. According to these results, the lack of information about destinations may explain why some individuals experience a decline in wages as a result of migration. When the migrant sample is defined by out-of-state migration within the past year, the outmigration rate has a statistically insignificant impact on the probability of migration, and the return to migration is higher among individuals coming from states with low outmigration rates.

In conclusion, this paper underscores, both theoretically and empirically, that the information structure available to an individual at the time of his/her migration decision is an important determinant of migration and economic outcomes related to migration. Future research should investigate the extent to which differences in the level of information across migrants with different demographic and socioeconomic characteristics can explain the variation in migration outcomes observed across these groups.

24

References

Allen, Jeremiah. "Information and Subsequent Migration: Further Analysis and Additional Evidence," *Southern Economic Journal,* (Apr. 1979), 45: 1274-1284.

Allen, Jeremiah and B. Curtis Eaton. "Incomplete Information and Migration: The Grass is Greener Across the Higher Fence," *Journal of Regional Science*, (2005), 45: 1-19.

Angrist, Joshua D. and William N. Evans. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, (1998), 88(3): 450-477.

Bartel, A. P. "The Migration Decision: What Role Does Job Mobility Play?" *American Economic Review*, (1979), 69, 775-786.

Bauer, Thomas, Gil S. Epstein and Ira N. Gang. "The Influence of Stocks and Flows on Migrants' Location Choices," in S. Polachek and K. Tatsiramos (eds.), *Research in Labor Economics*, (2007), Emerald Group Publishing Limited, 26: 199-229.

Ben-Porath, Yoram and Finis Welch. "Do Sex Preferences Really Matter?" *Quarterly Journal of Economics*, (1976), 90(2): 285-307.

Berninghaus, Siegfried and Hans G. Seifert-Vogt. "International Migration under Incomplete Information, " *Schweizerische Zeitschrift fur Volkswirtschaft und Statistik*, (1987), 123: 199-218.

Bernstein, Jared, Elizabeth McNichol and Karen Lyons. "Pulling Apart: A State-by-State Analysis of Income Trends," Economic Policy Institute/Center on Budget and Policy Priorities manuscript, (Jan. 2006).

Borjas, G. J., Bronars, S. G. and S. J. Trejo. "Assimilation and the Earnings of Young Internal Migrants," *Review of Economics and Statistics*, (1992), 74: 170-175.

Caces, F., F. Arnold, J. T. Fawcett and R. W. Gardner. "Shadow Household and Competing Auspices: Migration Behavior in Philippines," *Journal of Development Economics*, (1985), 17: 5-25.

Carrington, William J., Enrica Detragiache and Tara Vishwanath. "Migration with Endogenous Moving Costs," *American Economic Review*, (Sep. 1996), 86: 909-930.

Chen, Yuyu, Ginger Zhe Jin and Yang Yue. "Peer Migration in China," National Bureau of Economic Research Working Paper 15671, (January 2010).

Daneshvary, N., H.W. Herzog, R.A. Hofler and A. Schlottmann. "Job Search and Immigration Assimilation: An Earnings Frontier Approach," *Review of Economics and Statistics*, (1992), 74: 482-492.

DaVanzo, Julie. "Differences between Return and Nonreturn Migration: An Econometric Analysis," *International Migration Review*, (Spring 1976), 13-27.

DaVanzo, Julie and Peter A. Morrison. "Return and Other Sequences of Migration in the U.S.," *Demography,* (Feb. 1981), 18: 85-101.

DaVanzo, Julie. "Repeat Migration in the United States: Who Moves Back and Who Moves On?" *Review of Economics and Statistics*, (1983), 65: 552-559.

Franklin, Rachel S. "Domestic Migration Across Regions,Divisions, and States: 1995 to 2000," Census 2000 Special Reports, (Aug. 2003).

Gibbs, R.M. "The Information Effects of Origin on Migrants Job Search Behavior," *Journal of Regional Science*, (1994), 34: 163-178.

Greenwood, Michael. "Research on Internal Migration in the United States: A Survey," *Journal of Economic Literature*, (1975), 13: 397-433.

Greenwood, Michael J. "Internal Migration in Developed Countries," in M.R. Rosenzweig and O. Stark (eds.), *Handbook of Population and Family Economics*, (1997), Elsevier, Amsterdam, pp. 647-720.

Ham, John C., Xianghong Li and Patricia B. Reagan. "Propensity Score Matching, a Distance-Based Measure of Migration, and the Wages of Young Men," manuscript, (2006).

Heckman, J.J. "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, (1974), 42: 679-94.

Heckman, J.J. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simpler Estimator for Such Models," *Annals of Economic and Social Measurement*, (1976), 475-92.

Heckman, J.J. "Dummy Endogenous Variables in a Simultaneous Equations System," *Econometrica*, (1978), 46: 931-59.

Herzog, Henry W., Richard A. Hofler and Alan M. Schlottmann. "Life on the Frontier: Migrant Information, Earnings and Past Mobility," *Review of Economics and Statistics*, (1985), 67: 373-382.

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

Ioannides, Yannis M. and Giorgio Topa. "Neighborhood Effects: Accomplishments and Looking Beyond Them," *Journal of Regional Science*, (2010), 50(1); 343-362.

Kau, James B. and C. F. Sirmans. "The Influence of Information Costs and Uncertainty on Migration: A Comparison of Migrant Types," *Journal of Regional Science,* (1977), 17: 89-96.

King, Miriam, Steven Ruggles, Trent Alexander, Donna Leicach and Matthew Sobek. *Integrated Public Use Microdata Series, Current Population Survey; Version 2.0.* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004.

Lee, L.F. "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables," *International Economic Review*, (1978), 19: 415-33.

Lee, L.F. "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables," *Econometrica*, (1979), 47: 977-96.

Manski, C. F. "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, (1993), 60: 531-542.

Maurin, Eric and Julie Moschion. "The Social Multiplier and Labor Market Participation of Mothers," *American Economic Journal: Applied Economics*, (2009), 1(1): 251-272.

Munshi, K. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *Quarterly Journal of Economics*, (2003), 118(2): 549-99.

Pebley, Anne R. and Charles F. Westoff. "Women's Sex Preferences in the United States: 1970-1975," *Demography*, (1982), 19: 177-189.

Polachek, S. W. and Horvath, F. W. "A Life Cycle Approach to Migration: Analysis of the Perspicacious Peregrinator," in Ehrenberg, R. G. (ed.), *Research in Labor Economics*, (1977), 103-149.

Radu, Dragos. "Social Interactions in Economic Models of Migration: A Review and Appraisal," *Journal of Ethnic and Migration Studies*, (2008), 34: 531-548.

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

Spilimbergo, Antonio, and Luis Ubeda. "A Model of Multiple Equilibria in Geographic Labor Mobility," *Journal of Development Economics*, (2004), 73: 107-123.

Taylor, J. Edward. "Differential Migration, Networks, Information and Risk," in O. Stark (ed.), *Migration, Human Capital and Development*, (1986), JAI Press, Greenwich, CT.

Tunali, I. "Rationality of Migration," *International Economic Review*, (2000), 41: 893-920.

Westoff, Charles F., Robert G. Potter and Phillip C. Sagi. The Third Child: A Study in the Prediction of Fertility, (1963), Princeton University Press, Princeton, NJ.

Yankow, J. J. "Migration, Job Change, and Wage Growth: A New Perspective on the Pecuniary Return to Geographic Mobility," *Journal of Regional Science*, (2003), 43: 483-516.

Appendix

A1. The Probability of Migration

The probability that a randomly chosen individual chooses to migrate is given by

$$P = \Pr(\eta > \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon) = 1 - \Phi(z) \tag{23}$$

where $\eta = \upsilon_1 + \bar{\varepsilon} - \upsilon_0 - \upsilon_c$, $z = \frac{\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon}{\sigma_\eta}$, and Φ is the cdf a standard normal distribution. The derivative of P with respect to n is

$$\frac{\partial P}{\partial n} = -\phi(z) \cdot \frac{\partial z}{\partial \sigma_\eta} \cdot \frac{\partial \sigma_\eta}{\partial n} \tag{24}$$

First, I will simplify the expression for σ_η. By definition, σ_η can be stated as

$$\sigma_\eta = \sqrt{Var(\upsilon_1 + \bar{\varepsilon} - \upsilon_0 - \upsilon_c)} \tag{25}$$

$$Var(\upsilon_1 + \bar{\varepsilon} - \upsilon_0 - \upsilon_c) =$$

$$Var(\upsilon_1) + Var(\bar{\varepsilon}) + Cov(\upsilon_1, \bar{\varepsilon}) - Cov(\upsilon_1, \upsilon_0) - Cov(\upsilon_1, \upsilon_c)$$

$$- Cov(\upsilon_0, \bar{\varepsilon}) - Cov(\upsilon_c, \bar{\varepsilon}) + Cov(\upsilon_0, \upsilon_c) + Var(\upsilon_0) + Var(\upsilon_c)$$

Since $Cov(\upsilon_1, \bar{\varepsilon}) = Cov(\upsilon_0, \bar{\varepsilon}) = Cov(\upsilon_c, \bar{\varepsilon}) = 0$, the expression for $Var(\upsilon_1 + \bar{\varepsilon} - \upsilon_0 - \upsilon_c)$ can be written as

$$Var(\upsilon_1 + \bar{\varepsilon} - \upsilon_0 - \upsilon_c) =$$

$$= Var(\upsilon_1) + Var(\bar{\varepsilon}) - Cov(\upsilon_1, \upsilon_0) - Cov(\upsilon_1, \upsilon_c) + Cov(\upsilon_0, \upsilon_c) + Var(\upsilon_0) + Var(\upsilon_c)$$

$$= \sigma_1^2 + \sigma_{\bar{\varepsilon}}^2 + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c} = \sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c}$$

Then, σ_η is equals

$$\sigma_\eta = \left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c} \right)^{0.5} \tag{26}$$

As a result, the third factor in the expression of $\frac{\partial P}{\partial n}$ is

$$\frac{\partial \sigma_\eta}{\partial n} = 0.5 \left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c} \right)^{-0.5} \cdot \left(-\frac{\sigma_\varepsilon^2}{n^2} \right) < 0 \tag{27}$$

The second factor in the expression of $\frac{\partial P}{\partial n}$ can be written as

$$\frac{\partial z}{\partial \sigma_\eta} = \left(-\frac{\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon}{\sigma_\eta^2} \right) \tag{28}$$

Since $\phi(z) > 0$ and $\frac{\partial \sigma_\eta}{\partial n} < 0$,

$$\begin{aligned}
\frac{\partial P}{\partial n} &< 0 \;\; if \;\; \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0 \\
\frac{\partial P}{\partial n} &> 0 \;\; if \;\; \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0
\end{aligned} \tag{29}$$

The derivative of P with respect to σ_ε is

$$\begin{aligned}
\frac{\partial P}{\partial \sigma_\varepsilon} &= -\phi(z) \cdot \frac{\partial z}{\partial \sigma_\eta} \cdot \frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} \\
&= -\phi(z) \cdot \left(-\frac{\mu_0 + \mu_c - \mu_1 - \mu_\varepsilon}{\sigma_\eta^2} \right) \cdot 0.5 \left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c} \right)^{-0.5} \cdot \left(\frac{2\sigma_\varepsilon}{n} \right)
\end{aligned} \tag{30}$$

The third factor in the above expression is

$$\frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} = 0.5 \left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c} \right)^{-0.5} \cdot \left(\frac{2\sigma_\varepsilon}{n} \right) > 0 \tag{31}$$

As a result,

$$\begin{aligned}
\frac{\partial P}{\partial \sigma_\varepsilon} &> 0 \;\; if \;\; \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0 \\
\frac{\partial P}{\partial \sigma_\varepsilon} &< 0 \;\; if \;\; \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0
\end{aligned} \tag{32}$$

A2. The Return to Migration

The return to migration, R, can be expressed as

$$R = E(y_1|M = 1) - E(y_0|M = 1) \tag{33}$$

The two terms in the expression for R can be expressed as follows:

$$E(y_1|M = 1) = \mu_1 + E(\omega|\eta > \mu_0 + \mu_c - \mu_1) = \mu_1 + \mu_\varepsilon + \rho_{\omega\eta}\sigma_\omega\lambda_1 \tag{34}$$

where $\omega = \upsilon_1 + \varepsilon$, $\rho_{\omega\eta} = Corr(\omega, \eta)$ and $\lambda_1 = \frac{\phi(z)}{1-\Phi(z)}$. Similarly,

$$E(y_0|M = 1) = \mu_0 + E(\upsilon_0|\eta > \mu_0 + \mu_c - \mu_1) = \mu_0 + \rho_{0\eta}\sigma_0\lambda_1 \tag{35}$$

where $\rho_{0\eta} = Corr(\upsilon_0, \eta)$. Then, the return to migration can be stated as

$$R = \mu_1 + \mu_\varepsilon - \mu_0 + \left(\rho_{\omega\eta}\sigma\omega - \rho_{0\eta}\sigma_0\right)\lambda_1 \tag{36}$$

The following expressions are needed to further simplify the equation for R.

$$\rho_{0\eta} = Corr(\upsilon_0, \eta) = \frac{Cov(\upsilon_0, \eta)}{\sigma_0\sigma_\eta} \tag{37}$$

$$
\begin{aligned}
Cov(\upsilon_0, \eta) &= Cov(\upsilon_0, \upsilon_1 + \bar{\varepsilon} - \upsilon_0 - \upsilon_c) \\
&= Cov(\upsilon_0, \upsilon_1) + Cov(\upsilon_0, \bar{\varepsilon}) - \sigma_0^2 - Cov(\upsilon_0, \upsilon_c)
\end{aligned} \tag{38}
$$

Since $Cov(\upsilon_0, \bar{\varepsilon}) = 0$,

$$Cov(\upsilon_0, \eta) = \sigma_{01} - \sigma_0^2 - \sigma_{0c} \tag{39}$$

$$\rho_{0\eta}\sigma_0\lambda_1 = \frac{\sigma_{01} - \sigma_0^2 - \sigma_{0c}}{\sigma_0\sigma_\eta}\sigma_0\lambda_1 = \frac{\lambda_1}{\sigma_\eta}\left(\sigma_{01} - \sigma_0^2 - \sigma_{0c}\right) \tag{40}$$

31

$$\rho_{\omega\eta} = Corr(\omega, \eta) = Corr(v_1 + \varepsilon, v_1 + \overline{\varepsilon} - v_0 - v_c) = \frac{Cov(v_1 + \varepsilon, v_1 + \overline{\varepsilon} - v_0 - v_c)}{\sigma_\omega \sigma_\eta} \tag{41}$$

$$Cov(v_1 + \varepsilon, v_1 + \overline{\varepsilon} - v_0 - v_c) =$$

$$Var(v_1) + Cov(v_1, \overline{\varepsilon}) - Cov(v_1, v_0) - Cov(v_1, v_c) + Cov(\varepsilon, v_1) + Cov(\varepsilon, \overline{\varepsilon}) - Cov(\varepsilon, v_0) - Cov(\varepsilon, v_c)$$

Since $Cov(v_1, \overline{\varepsilon}) = Cov(\varepsilon, \overline{\varepsilon}) = 0$,

$$Cov(v_1 + \varepsilon, v_1 + \overline{\varepsilon} - v_0 - v_c) = \sigma_1^2 - \sigma_{10} - \sigma_{1c} + \sigma_{1\varepsilon} - \sigma_{0\varepsilon} - \sigma_{\varepsilon c} \tag{42}$$

Then,
$$\rho_{\omega\eta}\sigma_\omega\lambda_1 = \frac{\sigma_1^2 - \sigma_{10} - \sigma_{1c} + \sigma_{1\varepsilon} - \sigma_{0\varepsilon} - \sigma_{\varepsilon c}}{\sigma_\omega \sigma_\eta} \sigma_\omega \lambda_1 = \frac{\lambda_1}{\sigma_\eta}\left(\sigma_1^2 - \sigma_{10} - \sigma_{1c}\right) \tag{43}$$

If $\sigma_{0c} = \sigma_{1c} = \sigma_{0\varepsilon} = \sigma_{1\varepsilon}$ and $\sigma_{c\varepsilon} = 0$, the expression for R can be further simplified to

$$R = \mu_1 + \mu_\varepsilon - \mu_0 + \frac{\lambda_1}{\sigma_\eta} Var(v_1 - v_0) \tag{44}$$

Then, the derivative of R with respect to n is given by

$$\frac{\partial R}{\partial n} = \left[\frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial n} - \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial n}\right] Var(v_1 - v_0) \tag{45}$$

In order to be able to sign this derivative, one has to consider the derivatives of $\frac{\partial \lambda_1}{\partial n}$ and $\frac{\partial \sigma_\eta}{\partial n}$.

$$\frac{\partial \lambda_1}{\partial n} = \frac{\partial \lambda_1}{\partial z} \cdot \frac{\partial z}{\partial \sigma_\eta} \cdot \frac{\partial \sigma_\eta}{\partial n} \tag{46}$$

The three factors in this equation have the following signs:

$$\frac{\partial \lambda_1}{\partial z} > 0 \tag{47}$$

As shown above,

$$
\begin{aligned}
\frac{\partial z}{\partial \sigma_\eta} &> 0 \ if \ \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0 \\
\frac{\partial z}{\partial \sigma_\eta} &< 0 \ if \ \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0
\end{aligned} \tag{48}
$$

and

$$
\frac{\partial \sigma_\eta}{\partial n} = 0.5 \left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c} \right)^{-0.5} \cdot \left(-\frac{\sigma_\varepsilon^2}{n^2} \right) < 0 \tag{49}
$$

Therefore,

$$
\begin{aligned}
\frac{\partial \lambda_1}{\partial n} &> 0 \ if \ \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0 \\
\frac{\partial \lambda_1}{\partial n} &< 0 \ if \ \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0
\end{aligned} \tag{50}
$$

Then,

$$
\begin{aligned}
\frac{\partial R}{\partial n} &< 0 \ if \ \mu_0 + \mu_c - \mu_1 < 0 \ and \ \left| \frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial n} \right| > \left| \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial n} \right| \\
\frac{\partial R}{\partial n} &> 0 \ otherwise
\end{aligned} \tag{51}
$$

Equation 44 can also be used to calculate the derivative of R with respect to σ_ε.

$$
\frac{\partial R}{\partial \sigma_\varepsilon} = \left[\frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial \sigma_\varepsilon} - \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} \right] Var(v_1 - v_0) \tag{52}
$$

As shown above, $\frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} = 0.5 \left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c} \right)^{-0.5} \cdot \left(\frac{2\sigma_\varepsilon}{n} \right) > 0$.

Furthermore,

$$
\frac{\partial \lambda_1}{\partial \sigma_\varepsilon} = \frac{\partial \lambda_1}{\partial z} \cdot \frac{\partial z}{\partial \sigma_\eta} \cdot \frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} \tag{53}
$$

$$
\frac{\partial \lambda_1}{\partial z} > 0 \tag{54}
$$

$$
\frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} > 0 \tag{55}
$$

$$\frac{\partial z}{\partial \sigma_\eta} > 0 \ \ if \ \ \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0$$

$$\frac{\partial z}{\partial \sigma_\eta} < 0 \ \ if \ \ \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon > 0 \tag{56}$$

Therefore,

$$\frac{\partial R}{\partial \sigma_\varepsilon} > 0 \ \ if \ \ \mu_0 + \mu_c - \mu_1 - \mu_\varepsilon < 0 \ \ and \ \ \frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial \sigma_\varepsilon} > \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon}$$

$$\frac{\partial R}{\partial \sigma_\varepsilon} < 0 \ \ otherwise \tag{57}$$

A3. The Prediction Error

The prediction error is defined as the difference between the predicted post-migration earnings and the actual post-migration earnings. The expected value of the prediction error in the entire population is

$$E(y_1^e - y_1) = E(\mu_1 + \upsilon_1 + \bar{\varepsilon} - \mu_1 - \upsilon_1 - \varepsilon) \tag{58}$$

Since $E(\varepsilon) = E(\bar{\varepsilon}) = \mu_\varepsilon$,

$$E(y_1^e - y_1) = 0 \tag{59}$$

The expected value of the prediction error among migrants is $E(y_1^e - y_1 | M = 1)$. This expression depends on $E(y_1^e | M = 1)$, which can be expressed as follows:

$$E(y_1^e | M = 1) = \mu_1 + E(\upsilon_1 + \bar{\varepsilon} | \eta > \mu_0 + \mu_c - \mu_1) \tag{60}$$

Let $a = \upsilon_1 + \bar{\varepsilon}$. Then,

$$E(y_1^e | M = 1) = \mu_1 + \mu_\varepsilon + \rho_{a\eta} \sigma_a \lambda_1 \tag{61}$$

The following calculations are used in simplifying the expression for $E(y_1^e | M = 1)$.

$$\rho_{a\eta} = \frac{Cov(\upsilon_1 + \bar{\varepsilon}, \upsilon_1 + \bar{\varepsilon} - \upsilon_0 - \upsilon_c)}{\sigma_a \sigma_\eta} \tag{62}$$

$$Cov(v_1 + \bar{\varepsilon}, v_1 + \bar{\varepsilon} - v_0 - v_c) =$$

$$Var(v_1) + Cov(v_1, \bar{\varepsilon}) - Cov(v_1, v_0) - Cov(v_1, v_c) + Cov(v_1, \bar{\varepsilon}) + Var(\bar{\varepsilon}) - Cov(v_0, \bar{\varepsilon}) - Cov(v_c, \bar{\varepsilon})$$

Since $Cov(v_1, \bar{\varepsilon}) = Cov(v_0, \bar{\varepsilon}) = Cov(v_c, \bar{\varepsilon}) = 0$,

$$Cov(v_1 + \bar{\varepsilon}, v_1 + \bar{\varepsilon} - v_0) = \sigma_1^2 - \sigma_{01} - \sigma_{1c} + \frac{\sigma_\varepsilon^2}{n} \tag{63}$$

and

$$E(y_1^e | M = 1) = \mu_1 + \mu_\varepsilon + \frac{\sigma_1^2 - \sigma_{01} - \sigma_{1c} + \frac{\sigma_\varepsilon^2}{n}}{\sigma_\eta} \lambda_1 \tag{64}$$

Then, the prediction error among migrants can be stated as

$$E(y_1^e | M = 1) - E(y_1 | M = 1) =$$

$$\mu_1 + \mu_\varepsilon + \frac{\lambda_1}{\sigma_\eta} \left(\sigma_1^2 - \sigma_{01} - \sigma_{1c} + \frac{\sigma_\varepsilon^2}{n} \right) - \mu_1 - \mu_\varepsilon - \frac{\lambda_1}{\sigma_\eta} \left(\sigma_1^2 - \sigma_{01} - \sigma_{1c} \right) = \frac{\lambda_1}{\sigma_\eta} \left(\frac{\sigma_\varepsilon^2}{n} \right) > 0$$

Note that as $n \to \infty$, $\frac{\sigma_\varepsilon^2}{n} \to 0$ and $E(y_1^e | M = 1) - E(y_1 | M = 1) \to 0$.

5 Tables

Table 1: Variable Definitions and Descriptive Statistics

Variable Name	Variable Definition	Mean (Standard Deviation)			
		Whole Sample	Migrants	Stayers	(Stayers - Migrants)
Mig1	=1 if R moved to a different state within the past year	0.0258 (0.0007)			
Mig5	=1 if R moved to different state within past five years	0.0815 (0.0011)			
Log(wage)	Log of hourly wage	2.6985 (0.0029)	2.6894 (0.0104)	2.6993 (0.0030)	0.0100 (0.0108)
Age	Age in years	40.7180 (0.0531)	36.3390 (0.1687)	41.1065 (0.0554)	4.7675 (0.1776)
Experience	Age - years of schooling - 6	23.3505 (0.0529)	18.4958 (0.1700)	23.7812 (0.0551)	5.2854 (0.1787)
High school	=1 if highest degree earned is high school diploma	0.5906 (0.0019)	0.5179 (0.0070)	0.5971 (0.0020)	0.0792 (0.0072)
College	=1 if highest degree earned is bachelor's or higher	0.2949 (0.0018)	0.3857 (0.0068)	0.2868 (0.0018)	-0.0989 -(0.0071)
Male	=1 if male	0.5321 (0.0019)	0.5495 (0.0069)	0.5305 (0.0020)	-0.0190 (0.0072)
White	=1 if white	0.8257 (0.0015)	0.7917 (0.0057)	0.8287 (0.0015)	0.0370 (0.0059)
Married	=1 if married	0.5950 (0.0019)	0.5132 (0.0070)	0.6023 (0.0020)	0.0890 (0.0073)
Number of children	Number of children in the household	0.8393 (0.0041)	0.6991 (0.0134)	0.8518 (0.0043)	0.1527 (0.0140)
Own house	=1 if household owns home	0.7380 (0.0017)	0.4988 (0.0070)	0.7593 (0.0018)	0.2605 (0.0072)
Center city	=1 if respondent lives in central city in metro area	0.8348 (0.0014)	0.8468 (0.0048)	0.8337 (0.0014)	-0.0131 (0.0050)
Self-employed	=1 if respondent is self-employed	0.0998 (0.0012)	0.0785 (0.0038)	0.1017 (0.0012)	0.0232 (0.0040)

Unemployed	=1 if employment status is unemployed	0.0408 (0.0008)	0.0612 (0.0035)	0.0390 (0.0008)	-0.0222 (0.0036)
Manager	=1 if R works in management occupation	0.3430 (0.0018)	0.3833 (0.0068)	0.3394 (0.0019)	-0.0439 (0.0070)
Service	=1 if R works in a service occupation	0.4167 (0.0019)	0.4086 (0.0068)	0.4175 (0.0020)	0.0089 (0.0071)
Production	=1 if R works in production occupation	0.1321 (0.0013)	0.1082 (0.0042)	0.1342 (0.0014)	0.0260 (0.0044)
Northeast	=1 if respondent lives in northeast US	0.1886 (0.0015)	0.1516 (0.0049)	0.1919 (0.0016)	0.0403 (0.0051)
Midwest	=1 if respondent lives in midwest US	0.2330 (0.0016)	0.1879 (0.0051)	0.2370 (0.0016)	0.0490 (0.0054)
South	=1 if respondent lives in southern US	0.3526 (0.0019)	0.4106 (0.0070)	0.3475 (0.0020)	-0.0631 (0.0072)
Outmigration Rate	Outmigration rate at R's state of origin (%)	8.7358 (0.0076)	9.4762 (0.0502)	8.6701 (0.0068)	-0.8061 -(0.0506)
Sample Size		96,333	88,094	8,239	

Notes: Sampling weights are used in the calculation of the statistics presented in this table.

Mig5 migration indicator is used in the classification of migrants; therefore, migrants moved to a different state within the past five years.

Table 2: Probit Estimates of the Migration Equation

	I		II		III	
Dependent Variable	**Mig1**	**Mig5**	**Mig1**	**Mig5**	**Mig1**	**Mig5**
Age	0.0969***	0.1437***	0.0990***	0.1446***	0.0980***	0.1397***
	(0.0202)	(0.0135)	(0.0202)	(0.0134)	(0.0205)	(0.0152)
Age squared	-0.0009***	-0.0011***	-0.0009***	-0.0011***	-0.0009***	-0.0011***
	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)
Experience	-0.0768***	-0.1116***	-0.0781***	-0.1114***	-0.0789***	-0.1058***
	(0.0157)	(0.0104)	(0.0158)	(0.0104)	(0.0156)	(0.0127)
Experience squared	0.0009***	0.0010***	0.0009***	0.0010***	0.0008***	0.0010***
	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)
High school	-0.0649	-0.1302***	-0.0805	-0.1399***	-0.0689	-0.1454***
	(0.0607)	(0.0399)	(0.0610)	(0.0400)	(0.0643)	(0.0390)
College	0.0195	-0.0702	-0.0028	-0.0802	0.0060	-0.0840
	(0.1024)	(0.0657)	(0.1032)	(0.0656)	(0.1025)	(0.0639)
Male	0.0571**	0.0675***	0.0580**	0.0671***	0.0557**	0.0654***
	(0.0246)	(0.0160)	(0.0247)	(0.0160)	(0.0251)	(0.0160)
White	0.0215	-0.0156	0.0396	0.0025	0.0215	0.0241
	(0.0293)	(0.0194)	(0.0296)	(0.0195)	(0.0440)	(0.0282)
Married	0.0070	0.0530***	0.0104	0.0533***	0.0107	0.0504***
	(0.0282)	(0.0186)	(0.0283)	(0.0186)	(0.0279)	(0.0185)
Number of children	-0.0393***	-0.0469***	-0.0377***	-0.0440***	-0.0407***	-0.0382***
	(0.0119)	(0.0074)	(0.0120)	(0.0074)	(0.0125)	(0.0099)
Own house	-0.6157***	-0.5714***	-0.6215***	-0.5773***	-0.6032***	-0.5681***
	(0.0248)	(0.0166)	(0.0249)	(0.0166)	(0.0601)	(0.0273)
Center city	-0.0195	-0.0557***	0.0064	-0.0277	-0.0460	0.0260
	(0.0322)	(0.0198)	(0.0326)	(0.0200)	(0.0960)	(0.0562)
Self-employed	-0.0105	-0.0236	-0.0107	-0.0238	-0.0092	-0.0241
	(0.0485)	(0.0271)	(0.0489)	(0.0272)	(0.0481)	(0.0265)
Unemployed	0.2704***	0.1963***	0.2773***	0.1940***	0.2655***	0.1886***
	(0.0452)	(0.0356)	(0.0453)	(0.0349)	(0.0552)	(0.0348)
Manager	0.0022	0.0525*	0.0083	0.0565*	0.0010	0.0614**
	(0.0473)	(0.0300)	(0.0474)	(0.0299)	(0.0491)	(0.0294)
Service	0.0130	0.0336	0.0201	0.0368	0.0156	0.0389
	(0.0427)	(0.0278)	(0.0427)	(0.0279)	(0.0433)	(0.0272)
Production	0.0109	-0.0209	0.0224	-0.0071	0.0033	0.0131
	(0.0481)	(0.0316)	(0.0481)	(0.0318)	(0.0592)	(0.0362)

Northeast	-0.1262***	-0.1559***	-0.0672*	-0.0937***	-0.1559	-0.0004
	(0.0357)	(0.0230)	(0.0363)	(0.0233)	(0.1608)	(0.0942)
Midwest	-0.0569*	-0.1008***	0.0451	0.0091	-0.1326	0.1912
	(0.0332)	(0.0214)	(0.0345)	(0.0224)	(0.3198)	(0.1756)
South	0.0752**	0.0848***	0.1289***	0.1509***	0.0275	0.2508***
	(0.0298)	(0.0194)	(0.0302)	(0.0197)	(0.1926)	(0.0958)
Outmigration rate			0.0512***	0.0592***	-0.0417	0.1563*
			(0.0041)	(0.0026)	(0.1676)	(0.0922)
Instrumental Var.	No	No	No	No	Yes	Yes
Pseudo-R^2	0.0841	0.0949	0.0730	0.0844		
Durbin-Wu-Hausman test (p-value)					0.000	0.000
F test of excluded instrument(F-stat.)					73.01	68.29

Notes: Robust standard errors of means are in parentheses. Probit regressions are weighted.

*=significant at 10%, **=significant at 5%, ***=significant at 1%.

Table 3: OLS Estimates of the Wage Equation

	Migration Indicator is Mig1		Migration Indicator is Mig5	
	Migrants	**Stayers**	**Migrants**	**Stayers**
Age	0.0663**	0.1309***	0.0785***	0.1298***
	(0.0272)	(0.0045)	(0.0148)	(0.0047)
Age squared	-0.0001	-0.0006***	-0.0002	-0.0005***
	(0.0003)	(0.0000)	(0.0002)	(0.0000)
Experience	-0.0310	-0.0809***	-0.0408***	-0.0795***
	(0.0213)	(0.0033)	(0.0109)	(0.0036)
Experience squared	-0.0003	0.0001**	-0.0002	0.0000
	(0.0003)	(0.0000)	(0.0001)	(0.0000)
High school	0.0328	-0.0264**	0.0547	-0.0269**
	(0.0925)	(0.0125)	(0.0462)	(0.0131)
College	0.2871**	0.0774***	0.2539***	0.0599***
	(0.1489)	(0.0207)	(0.0742)	(0.0217)
Male	0.2990***	0.2594***	0.2552***	0.2565***
	(0.0370)	(0.0053)	(0.0193)	(0.0056)
White	0.0682*	0.0403***	0.0243	0.0455***
	(0.0372)	(0.0061)	(0.0203)	(0.0064)
Married	0.1076***	0.0956***	0.1590***	0.1106***
	(0.0361)	(0.0056)	(0.0185)	(0.0058)
Manager	0.2268***	0.1915***	0.2337***	0.1929***
	(0.0731)	(0.0096)	(0.0344)	(0.0101)
Service	-0.0065	-0.0897***	-0.0607**	-0.0897***
	(0.0638)	(0.0085)	(0.0300)	(0.0090)
Production	0.0430	-0.0670***	-0.0738**	-0.0665***
	(0.0771)	(0.0093)	(0.0357)	(0.0098)
Constant	0.4023	-0.5485***	0.5159***	-0.2703***
	(0.3961)	(0.0674)	(0.2087)	(0.0669)
Inv. Mill's Ratio	0.1354**	-0.2201***	0.0626***	-0.1121***
	(0.0578)	(0.0096)	(0.0205)	(0.0069)
R^2	0.27	0.31	0.30	0.31

Notes:

1. Robust standard errors of means are in parentheses.

2. Regressions are weighted.

Table 4: Average Return to Migration by Education

	Migrants Moved Within Last Year		Migrants Moved Within Last Five Years	
	Mean	**Standard Error**	**Mean**	**Standard Error**
Entire sample	0.0351	(0.0014)	0.0073	(0.0005)
Education Categories				
Less than highschool	0.0994	(0.0040)	0.0290	(0.0020)
Highschool	0.0211	(0.0017)	-0.0037	(0.0007)
College	0.0360	(0.0025)	0.0166	(0.0007)

Notes: Sample includes migrants. Return to migration is calculated by each migrant's post-migration wage minus his/her imputed pre-migration wage. Standard errors are calculated from 1000 bootstrap repetitions.

3

Table 5: Average Return to Migration by Migration Rates in State of Origin

	Migrants Moved Within Last Year		Migrants Moved Within Last Five Years	
	Mean	**Error**	**Mean**	**Error**
By Outmigration Rate in the Home State				
Outmigration rate less than or equal to 7.28	0.0488	(0.0030)	-0.0088	(0.0009)
Outmigration rate between 7.28 and 11.22	0.0373	(0.0020)	0.0045	(0.0006)
Outmigration rate greater than or equal to 11.22	0.0155	(0.0028)	0.0312	(0.0010)

Notes: Sample includes migrants. Return to migration is calculated by each migrant's post-migration wage minus his/her imputed pre-migration wage. Standard errors are calculated from 1000 bootstrap repetitions. State outmigration rates used in the analysis are obtained from the calculations performed by the U.S. Census Bureau using the U.S. Census 2000 (Franklin, 2003).

Table 6: Average Return to Migration by Income Inequality in the Destination State

	Last Year		Last Five Years	
	Mean	**Error**	**Mean**	**Error**
By 80:20 Income Ratio in the Destination State				
80:20 Income Ratio less than or equal to 6.4	0.0356	(0.0024)	0.0066	(0.0009)
80:20 Income Ratio between 6.4 and 7.6	0.0336	(0.0019)	0.0072	(0.0007)
80:20 Income Ratio greater than or equal to 7.6	0.0407	(0.0040)	0.0095	(0.0013)
By 95:20 Income Ratio in the Destination State				
95:20 Income Ratio less than or equal to 10.3	0.0356	(0.0024)	0.0068	(0.0008)
95:20 Income Ratio between 10.3 and 13	0.0309	(0.0023)	0.0074	(0.0007)
95:20 Income Ratio greater than or equal to 13	0.0433	(0.0032)	0.0079	(0.0011)

Notes: Sample includes migrants. Return to migration is calculated by each migrant's post-migration wage minus his/her imputed pre-migration wage. Standard errors are calculated from 1000 bootstrap repetitions. Income ratios used in the analysis are calculated by Bernstein et al. using 2001-2003 Current Population Survey (Bernstein, 2006).