# The Impact of Information on Migration Outcomes

Demiralp, Berna

Old Dominion University

June 2009

# The Impact of Information on Migration Outcomes

Berna Demiralp
Old Dominion University

June 2009

**Abstract**

This paper presents a model of migration in which migration decisions are made with incomplete information on the labor market conditions at destination. It provides an explanation for how differences in the level of information about the destination can bring about differences in economic outcomes related to migration, such as the migration propensity and the return to migration. The implications of the model show the conditions under which information positively and negatively affects these outcomes. Thus, the model can be used to explain a wide set of empirical findings regarding the relationship between information and migration outcomes. 2005 CPS data are used to estimate the econometric model. The estimation results suggest that increased access to information regarding destination labor markets increases one's likelihood to migrate to another state. Furthermore, the findings suggest that people who have more information regarding the destination at the time of their migration decision on average experience higher returns to migration.

## 1 Introduction

Since Sjaastad's influential work, migration has been perceived as an investment in human capital (1962). Economists who have studied the return to this investment have often looked at the wage and earnings growth that are experienced by migrants as a result of their migration decisions. 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; however, 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 the contemporaneous change in wage and earnings of migrants that is brought about by migration. Findings of positive, negative and insignificant returns to migration, calculated as contemporary wage and earnings growth, exist in the literature, begging for the question of why different migrants experience different returns to their migration decisions[1].

In this paper, I present a model of migration under incomplete information as an alternative framework in which to study the consequences of migration. I start with the premise that information about labor market conditions at the destination is an important determinant of migration. Differences in the level of access to such information in the population can bring about differences in economic outcomes related to migration, such as the migration propensity and the return to migration. For example, in an environment with incomplete information regarding post-migration wages, negative wage growth can reflect overestimation of migration wages among migrants. The purpose of this paper is to investigate the role of incomplete information in migration decisions and more specifically the effect of increased information on migration outcomes. To that end, I present a model of migration under incomplete information, discuss its theoretical implications and provide an empirical analysis of how access to information about destination labor markets affects the rate of state-to-state migration within the U.S. and the return to migration among migrants.

Economists acknowledge that information plays an important role in people's migration decisions by directly affecting their expected benefits from migration. Many studies have empirically investigated the impact of incomplete information on migration behavior and 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 use the findings of the job search literature in developing their 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 comparison. 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.

[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.

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 to destination wages that are not perfectly observable by the individual. The individual does not know the population distribution of this random component, so she cannot use its population mean in calculating her prediction of her post-migration wages. The model is based on an information acquisition process initially proposed by Allen and Eaton (2005). According to this process, the individual receives a sample of $n$ draws from the population distribution of the random variable, so in effect, she draws a sample mean. She uses the sample mean of the $n$ observations to calculate the expected value of her post-migration wages. The number of observations that she receives, $n$, increases with the level of her information about the destination. Since the variance of the distribution of sample mean decreases with $n,$ in this model, information affects the worker's migration decision by changing the spread of the distribution from which she draws the sample mean. The model allows for both underprediction and overprediction of destination wages in the population. Allen and Eaton have used this information acquisition process to explain the effect of information about a destination to the rate of migration to that destination. In this paper, I focus on the role of incomplete information in bringing about the variation in the return to migration observed in the data.

The model presented in this paper predicts 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 post-migration wages 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.

Furthermore, the implications of the model reveal that increased information about destination labor markets can have both positive and negative effects 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. The effect of information on the probability of migration hinges on the difference between the expected value of home wages and the expected value of destination wages in the population. If the population mean of wages at origin exceeds that of the wages at destination, 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 migration rate. If the population mean of wages at origin is lower than the population mean of wages at destination, access to more information is likely to change the migration decisions of those who used to underestimate their post-migration wages. As these individuals decide to migrate under increased information, the migration rate increases.

Increased information regarding destination labor market conditions affects the return to migration through two channels, which I name as composition and scale effects due to their resemblance of the composition and scale effects of Borjas' model (1987). The composition effect reflects the impact of information on the return to migration through its effect on the composition of the migrant sample, conditional on the rate of migration. As a result of more information, the migrant sample consists of a greater proportion of people with high destination wages, and this change in the composition of migrants has a positive effect on the return to migration, conditional on the size of the migrant sample. The scale effect describes the effect of information on the return to migration through its impact on the size of the migrant sample. A positive scale effect exists 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 scale effect, on the other hand, is brought about when information brings about 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 theoretical model forms the basis for the empirical work which investigates how information about the destination labor market affects the rate of state-to-state migration within the U.S. and the wage gain associated with such moves. I use data from the March supplement to the 2005 Current Population Survey in the analysis. I draw on the results of previous research regarding the role of network externalities in migration. These results indicate that one of the channels through which people obtain information about other regional labor markets is their interaction with their friends and neighbors. One can learn about the job prospects in another state by talking to her friends and neighbors who have already migrated to or migrated from that state. Based on this statement, I assume that people who live in states with high gross migration rates have greater access to information about labor market conditions in other states since residents of such states are more likely to come into contact with people who have moved to or from other states.

In the estimation of the migration probability, the use of the state gross migration rate as a proxy for the residents' level of information about destination labor markets presents a potential problem of endogeneity because the gross migration rate can be correlated with unobservables such as local economic conditions which are also likely to affect a resident's migration decision.

Therefore, I use the median age in the state of origin as an instrumental variable for the state's gross migration rate in estimating the migration probit regression. The median age in the state is likely to affect the migration rate in that state since age is a determinant of migration; however, it should have no effect on an individual resident's migration decision when the decision is conditional on the resident's own age. The instrumental variable probit regression results indicate that increased access to information regarding destination labor markets increases one's likelihood to migrate to another state.

In order to address the potential endogeneity of the gross migration rate in the wage equation, I use a modified version of the Heckman's two-step estimator in estimating the wage equation for migrants and stayers. The estimation results indicate that increased access to information about destination labor markets has no significant effect on the wages of migrants and a significantly negative effect on the wages of stayers. In addition, I use the parameter estimates of the wage equation to calculate the return to migration among migrants, which is the difference between their reported post-migration wages and the estimated wage they would have earned had they decided to stay. The average return to migration is greater for individuals who have migrated from states with high gross migration rates. Furthermore, comparative static exercises show that increasing the gross migration rate leads to an increase in the return to migration, conditional on the original subsample of migrants. I also find that the return to migration is highest among high school dropouts and lowest among individuals with some college.

The paper is divided into several 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 and explains the empirical strategy employed in this study. Section IV discusses the dataset, and Section V contains the results of the empirical analysis. Concluding remarks are given in Section VI.

# 2 Theoretical Model

## 2.1 The Framework

Individuals make a decision between moving ($M = 1$) and staying ($M = 0$) based on their current wages, 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 she migrates and $y_{0i}$ if she does not migrate. These alternatives are given by

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

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

where $\mu_0$ and $\mu_1$ represent the population mean of wages at the origin and the population mean of the wages at the destination, respectively, and $\upsilon_{0i}$ and $\upsilon_{1i}$ represent the deviations from the mean that are observed by the individual. While the individual perfectly observes $\mu_0$, $\mu_1$, $\upsilon_{0i}$ and $\upsilon_{1i}$, she does not observe $\varepsilon_i$ prior to migration, and as such $\varepsilon_i$ can be interpreted as the random component of one's post-migration wages from the perspective of the individual at the time of the migration decision.

The information acquisition process presented in this model was originally proposed by Allen and Eaten in their migration model (2005)[2]. I assume that the individual does not have complete information in the sense that she does not know the distribution of $\varepsilon$ in the population. Instead, she observes $n$ number of random and independent draws from the population distribution $\varepsilon$ and uses the average of the $n$ draws as a predictor of $\varepsilon_i$. Let $\overline{\varepsilon_i}$ be the average of the $n$ random draws observed by individual $i$. Then, the individual anticipates her wages at the destination to be $y_{1i}^e$ where

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

The information acquisition process in this model is based on the assumption that information about the population mean of the random term, $\varepsilon$, is costly. $n$, the size of the sample drawn from the population distribution of $\varepsilon$, is a decreasing function of the cost of acquiring information. Individuals who face high costs of acquiring information about the labor market conditions at the destination have low values of $n$, and those, who face low costs of acquiring information and thus can obtain more about the destination labor market, have high values of $n$.

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

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

where both $\mu_c$ and $\upsilon_{ci}$ are known by the individual at the time of her migration decision. Then, individual $i$'s migration decision can be characterized as

$$\begin{aligned} M &= 1 \;\; if \;\; y_{1i}^e - y_{0i} - c_i > 0 \\ M &= 0 \;\; otherwise \end{aligned} \tag{5}$$

[2] Allen and Eaton's model focuses on explaining how the migration propensity changes with information about the destination. 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.

Note that $y_{1i}$ represents the actual post-migration wages while $y_{1i}^e$ represents individual $i$'s anticipation of her post-migration wages at the time of the migration decision. Furthermore, the random components, $\upsilon_0$, $\upsilon_1$, and $\upsilon_c$, are observed by the individual but are unknown to the researcher while $\varepsilon$ 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 $\upsilon_0$, $\upsilon_1$, $\upsilon_c$, and $\varepsilon$ are jointly normal with zero means and the covariance matrix

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

Since $\varepsilon$ is distributed normally in the population, the sample mean, $\overline{\varepsilon}$, also follows a normal distribution where $\overline{\varepsilon} \sim N(0, \frac{\sigma_\varepsilon^2}{n})$. In addition, since $n$ is inversely related to the variance of $\overline{\varepsilon}$, access to more information about the destination labor market, represented by an increase in $n$ in this model, decreases the variance of $\overline{\varepsilon}$ that individuals use in calculating their anticipated post-migration wages. As $n$ increases, individuals' $\overline{\varepsilon}$ draws become more concentrated around the actual population mean of $\varepsilon$, which is assumed to be zero. Thus, individuals move from a situation of incomplete information towards one of complete information, in which they know the expected value of $\varepsilon$ in the population and can use it in predicting their post-migration wages.

## 2.2 The Probability of Migration

From the perspective of the researcher, the probability that a randomly chosen individual chooses to migrate is given by

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

where $\eta = \upsilon_1 + \overline{\varepsilon} - \upsilon_0 - \upsilon_c$, $z = \frac{\mu_0+\mu_c-\mu_1}{\sigma_\eta}$, and $\Phi$ is the cdf of a standard normal distribution.

Since $n$ captures the individual's level of information regarding the destination labor market conditions, 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. Based on Equation 6, 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 $\mu_0 + \mu_c - \mu_1$. In particular,

$$\begin{aligned} \frac{\partial P}{\partial n} &> 0 \;\; if \;\; \mu_0 + \mu_c - \mu_1 < 0 \\ \frac{\partial P}{\partial n} &< 0 \;\; if \;\; \mu_0 + \mu_c - \mu_1 > 0 \end{aligned} \tag{8}$$

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 home wages is higher than the population mean of wages at destination minus the moving cost, then the probability of migration decreases as people become more informed. If the population mean of wages at home is lower than the population mean of destination wages minus the moving cost, then access to more information has a positive effect on the probability of migration.

Before explaining the intuition behind this result, one should note that this model allows for both underprediction and overprediction of post-migration wages by individuals depending on their $\varepsilon_i$ and $\overline{\varepsilon_i}$ draws. In particular, an individual overpredicts her post-migration wages if $\overline{\varepsilon_i} > \varepsilon_i$. In that case the individual makes a positive prediction error since $\overline{\varepsilon_i} - \varepsilon_i > 0$. Similarly, the individual underpredicts her post-migration wages if $\overline{\varepsilon_i} < \varepsilon_i$. In that case the $\overline{\varepsilon_i} - \varepsilon_i < 0$, and the individual's prediction error is negative.

The mathematical intuition behind Equation 8 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 zero, 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 zero 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 zero, these individuals become less likely to choose migration, leading to a decrease in the probability of migration in the population.

The incomplete information in this model, characterized as the individuals' lack of information regarding the population distribution of $\varepsilon$, has unique implications on the migration rate that cannot be captured by a standard model in which the individuals know the population distribution of $\varepsilon$. If the individuals know the expected value of $\varepsilon$ in the population, then $y^e_{1i} = \mu_1 + \upsilon_{1i} + E(\varepsilon) = \mu_1 + \upsilon_{1i}$ under the current assumption that $E(\varepsilon)$ is zero. Consequently, the individual chooses to migrate as long as $\mu_1 + \upsilon_{1i} > \mu_0 + \mu_c + \upsilon_{0i} + \upsilon_{ci}$, and the probability of migration is given by $P = \Pr(\upsilon_1 - \upsilon_0 - \upsilon_c > \mu_0 + \mu_c - \mu_1)$. In such a model, the variance of $\varepsilon$, which reflects the level of uncertainty surrounding post-migration wages, has no effect on the individual's migration propensity.

**Hypothesis 1** The individual's level of information regarding the destination labor market positively affects the probability of migration if $\mu_0 + \mu_c - \mu_1 < 0$. Otherwise, its effect on the probability of migration is negative.

## 2.3 The Return to Migration

The return to migration, $R$, can be 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 wages of migrants at the "new" location, and $E(y_0|M = 1)$ gives the expected value of wages of migrants at "home." If we let $\omega = \upsilon_1 + \varepsilon$, $\rho_{\omega\eta} = Corr(\omega, \eta)$ and $\lambda_1 = \frac{\phi(z)}{1-\Phi(z)}$, then the expected value of post-migration wages of migrants can be stated as

$$E(y_1|M = 1) = \mu_1 + E(\omega|\eta > \mu_0 + \mu_c - \mu_1) = \mu_1 + \rho_{\omega\eta}\sigma_\omega\lambda_1 \tag{10}$$

Similarly, the expected value of pre-migration wages of migrants can be expressed as

$$E(y_0|M = 1) = E(\mu_0 + \upsilon_0|\eta > \mu_0 + \mu_c - \mu_1) = \mu_0 + \rho_{0\eta}\sigma_0\lambda_1 \tag{11}$$

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

$$R = \mu_1 - \mu_0 + \left(\rho_{\omega\eta}\sigma_\omega - \rho_{0\eta}\sigma_0\right)\lambda_1 = \mu_1 - \mu_0 + A\frac{\lambda_1}{\sigma_\eta} \tag{12}$$

where $A = \sigma_1^2 - 2\sigma_{10} + \sigma_0^2 - \sigma_{1c} + \sigma_{0c} + \sigma_{1\varepsilon} - \sigma_{0\varepsilon} - \sigma_{c\varepsilon}$.

Before investigating effect of information on the return to migration, I would like to discuss the implications of the model on the average prediction error among migrants. As shown in the Appendix, the expected value of the prediction error among migrants is positive.

$$E(y_1^e|M = 1) - E(y_1|M = 1) = \frac{\lambda_1}{\sigma_\eta}\left(\frac{\sigma_\varepsilon^2}{n}\right) > 0 \tag{13}$$

Therefore, the average migrant overestimates her destination wages. Based on this implication, this 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, the positive prediction error among migrants is contingent on the assumption that the prediction error has a positive support over the entire population. These models 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, $\lambda_1$ and $\sigma_\eta$ approach constant values, $\frac{\sigma_\varepsilon^2}{n}$ approaches zero, and hence the expected value of the prediction error among migrants ($E(y_1^e|M = 1) - E(y_1|M = 1)$) 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 the migration that their actual post-migration wages are less than their wages at the origin. By showing that the overestimation of post-migration wages is prevalent among migrants, the model presented here provides an explanation for the negative return to migration within a human capital investment approach to migration when individuals have incomplete information regarding destination labor market conditions.

Next I turn to the question of how the level of information about the destination labor market affects the return to migration. Do migrants who have a better knowledge of the labor market conditions at the destination experience a higher return to their migration decisions? The impact of information on the return to migration is summarized by the effect of $n$ on $R$. If $\sigma_{0c} = \sigma_{1c} = \sigma_{0\varepsilon} = \sigma_{1\varepsilon}$ and $\sigma_{c\varepsilon} = 0$, then this derivative can be expressed as

$$\frac{\partial R}{\partial n} = \left[\frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial n} - \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial n}\right] Var(\upsilon_1 - \upsilon_0) \tag{14}$$

As shown in the Appendix, the sign of this derivative 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 described the effect of $n$ on $R$ through its impact on the selection and the size of the migrant sample. I will call this effect the "scale effect" since in a mathematical sense, it is similar to the "scale effect" described in Borjas' migration model (1987). The direction of the scale effect is inherently tied to the direction of the migration rate and the conditions that determine it. The scale effect is positive under the same conditions when $\frac{\partial P}{\partial n} < 0$ , and it is negative under the same conditions when $\frac{\partial P}{\partial n} > 0$ (i.e. $\frac{\partial \lambda_1}{\partial n} > 0$ when $\mu_0 + \mu_c - \mu_1 > 0$, and $\frac{\partial \lambda_1}{\partial n} < 0$ when $\mu_0 + \mu_c - \mu_1 < 0$). As mentioned above, when $\mu_0 + \mu_c - \mu_1 > 0$, the marginal individual who changes her migration decision as a result of more information is one who had previously overestimated her destination wages. As overestimators update their predictions of their post-migration wages, a portion of them are likely to change their migration decision from migration to staying, leading to a withdrawal of overestimators from the pool of migrants. As the migrant pool comprises of a smaller proportion of people who make a negative prediction error, the average post-migration wages in the migrant subsample increases, and the average return to migration rises. Therefore, a positive scale effect reflects the fact that as a smaller portion of migrants overpredict their destination wages, the return to migration is positively affected by an increase in $n$.

On the other hand, the scale effect is negative when the population mean of home wages plus the moving costs is less than the population mean of destination wages ($\mu_0 + \mu_c - \mu_1 < 0$). In that case, the marginal individual who changes her migration decision is one who had underestimated her post-migration wages. When more information is available, underestimators of future wages are likely to choose to migrate, 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 most likely have 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 scale effect.

The second term in the brackets can be perceived as the effect of $n$ on $R$ through its impact on the composition of migrants, holding the migration rate constant. When $n$ increases, the variances of $\overline{\varepsilon}$ and $y_1^e$ become smaller while the variance of $y_1$ remains constant. Then, if $\mu_0 + \mu_c - \mu_1 > 0$, the migrant sample consists of a greater proportion of people who have higher destination wages than anticipated. This change in the composition of migrants puts an upward pressure on $R$. If $\mu_0 + \mu_c - \mu_1 < 0$, the migrant sample consists of a smaller proportion of people who have lower destination wages than anticipated. Such a change in the composition of the migrant sample also has a positive effect on $R$. Therefore, the composition effect is unambiguously positive. Intuitively, 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 scale and composition effects. When $\mu_0 + \mu_c - \mu_1 > 0$, both effects are positive, generating a positive effect of information on the return to migration. When $\mu_0 + \mu_c - \mu_1 < 0$, the composition effect is positive, and the scale effect is negative; thus the sign of $\frac{\partial R}{\partial n}$ depends on which effect dominates. In that case, information increases the return to migration if $\frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial n} < \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial n}$. In sum,

$$\begin{aligned} \frac{\partial R}{\partial n} &< 0 \quad if \quad \mu_0 + \mu_c - \mu_1 < 0 \;\; and \;\; \left| \frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial n} \right| > \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial n} \\ \frac{\partial R}{\partial n} &> 0 \quad otherwise \end{aligned} \tag{15}$$

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

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

Next, I consider how an increase in the variance of $\varepsilon$ affects the return to migration. Within the framework of this model, $\sigma_\varepsilon$ captures the uncertainty faced by an individual regarding her destination wages. Based on Equation 2, $\sigma_\varepsilon$ directly affects the variance of wages at the destination. Therefore, the effect of $\sigma_\varepsilon$ 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 $\sigma_\varepsilon$ on the return to migration is given by[3]

$$\frac{\partial R}{\partial \sigma_\varepsilon} = \left[ \frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial \sigma_\varepsilon} - \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} \right] Var(v_1 - v_0) \tag{16}$$

This net effect can also be decomposed into a composition and scale effect. It can be shown that the composition effect, given by the second term in the bracket, is negative. As $\sigma_\varepsilon$ goes up and the uncertainty that one faces in her destination earning increases, conditional on $n$, the individual's prediction of her post-migration wages becomes less precise, thus negatively affecting her return to migration. An increase in $\sigma_\varepsilon$ also impacts the selection of migrant and the migration rate revealed by the scale effect (the first term in the brackets). The scale effect can be positive or negative depending on the relative values of $\mu_0$, $\mu_c$, and $\mu_1$. If $\mu_0 + \mu_c - \mu_1 < 0$, 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 scale effect is positive. On

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

the other hand, if $\mu_0 + \mu_c - \mu_1 > 0$, 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 scale effect associated with an increase in $\sigma_\varepsilon$ is negative. The net effect of a change in $\sigma_\varepsilon$ on $R$ depends on the sum of the composition and scale effects. In particular,

$$\begin{aligned}\frac{\partial R}{\partial \sigma_\varepsilon} &> 0 \; if \;\; \mu_0 + \mu_c - \mu_1 < 0 \;\; and \;\; \frac{1}{\sigma_\eta} \cdot \frac{\partial \lambda_1}{\partial \sigma_\varepsilon} > \left| \frac{\lambda_1}{\sigma_\eta^2} \cdot \frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} \right| \\ \frac{\partial R}{\partial \sigma_\varepsilon} &< 0 \; otherwise \end{aligned} \tag{17}$$

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

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

**Hypothesis 5** If the variance of wages at the destination positively affects the return to migration, then $\mu_0 + \mu_c - \mu_1 < 0$.

Hypothesis 4 can be combined with the earlier hypotheses to lead to the following result: If the expected value of home wages in the population exceeds the expected value of destination wages minus the moving costs in the population ($\mu_0 + \mu_c - \mu_1 > 0$), then an increased access to information about the destination leads to a decreased migration rate and an increased return to migration among migrants. The model also implies that if the wage inequality at the destination increases under this assumption, the return to migration decreases as a result.

# 3 Empirical Strategy

The empirical strategy employed in this paper involves investigating the effect that greater access to information has on migration outcomes by estimating the reduced form of the behavioral model discussed above. A critical part of this strategy is to distinguish between individuals facing different costs of information and thus different levels of access to information regarding destination labor markets. One of the channels through which people obtain information about other regional labor markets is their interaction with their friends and neighbors. For instance, one can learn about job prospects in another state by talking to her friends and neighbors who have already migrated to that state or migrated from that state.

Social interaction as an information-enhancing factor in the migration decision is implicit in several other studies. For instance, Carrington et al. present a 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 develop a model which specifies the role of social interaction in the migration decision (2004). Their work is based on the assumption that one's family and friend network at home might discourage her from migrating. They find multiple equilibria and use the existence of multiple equilibria to explain why different groups have persistently exhibited different migration rates (eg. White vs. African-American, U.S. vs. Europe). In addition, previous literature on migration has pointed to the role of network externalities in migration, referring to the fact that each migrant lowers the cost of migration for family and friends at home (Massey et al., 1993). Social interaction is implicit in these studies because one of the causes of network externalities in this setting is the informal information channels that carry information from people who have already migrated to those who are at the origin. Network externalities have been used as a reason behind ethnic clustering of immigrants in host countries across Europe and in the U.S.

If interaction with friends, family and neighbors who have migrated to or from one's home state is an informal medium for information about other labor markets outside her state, then residents of states that experience a high level of outmigration or inmigration are more likely to learn about other labor markets. Information gathering through social interaction implies that people who live in states with high migration rates are on average better informed about destination labor markets since they have a higher likelihood of interacting with people who have moved to or from a different state. Based on this reasoning, I use the gross migration rate in one's home state, which is based on the sum of inmigration and outmigration in the state, as a proxy for her level of information about destination labor markets. I assume that people who live in states with high gross migration rates have easier access to information about different labor markets outside their states and thus are able to collect information at a lower cost.

A potential problem with this strategy is that the gross migration rate is likely to be endogenous to the residents' migration decisions. A state's gross migration rate is partially determined by the state's economic conditions, which can also affect the residents' migration decisions. To address this potential endogeneity problem, I use the median age in the state as an instrument for the state's gross migration rate in the estimation. I believe that the median age in the state is a suitable instrument as it is likely to be related to the endogenous variable, the state's gross migration rate, and unrelated to an individual's migration decision. The median age in the state is likely to be a factor in the state's gross migration rate because previous studies have documented age to be a significant determinant of migration and one that is negatively related to the probability of migration (Greenwood, 1985). Therefore, one can expect states with a younger population to have more outmigration, which directly impacts the gross migration rate in the state. Furthermore, the median age in the state is not likely to affect a resident's decision to migrate, conditional on the age of the resident. Although the individual's own age is a factor that affects her migration propensity, the median age in her state of origin should have no influence on her migration decision when conditional on her own age.

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

$$y_0 = \beta_0 X + \alpha_0 g + \upsilon_0 \tag{18}$$

$$y_1 = \beta_1 X + \alpha_1 g + \omega \tag{19}$$

$$y^* = \delta Z + \eta \tag{20}$$

$$M(Z) = 1[y^* \geq 0] \tag{21}$$

$$y = My_1 + (1 - M)y_0 \tag{22}$$

The error terms, $\upsilon_0$, $\omega$, and $\eta$ 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:

$$\begin{bmatrix} \sigma_0^2 & \sigma_{0\omega} & \sigma_{0\eta} \\ \sigma_{0\omega} & \sigma_\omega^2 & \sigma_{\omega\eta} \\ \sigma_{0\eta} & \sigma_{\omega\eta} & \sigma_\eta^2 \end{bmatrix} \tag{23}$$

$y$ is the observed wages in the sample, with $y_1$ denoting the wages of migrants and $y_0$ denoting the wages of stayers. The latent variable, $y^*$, determines the regime, i.e. the individual's migration status $(M)$. This specification assumes that the population means of earnings and costs $(\mu_0, \mu_1, \mu_c)$, which capture the deterministic components of earnings and cost in the theoretical model, are linear functions of observable characterictics. The vector of observables, $X$, includes variables that affect one's earnings, and the vector $Z$ includes explanatory variables that affect earnings as well as those that determine moving costs. Both $X$ and $Z$ contain exogenous explanatory variables. $g$ is an endogenous variable, which is correlated with the error terms in the earnings equations, $\upsilon_0$ and $\omega$.

The given specification describes a switching regression model where one of the regressors in the outcome equation (earnings equation) is endogenous. The estimation technique should take into account two issues: 1) 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 should allow for the condition that $Cov(\upsilon_0, \eta) \neq 0$ and/or $Cov(\omega, \eta) \neq 0$); 2) the endogeneity of $g$ in the earnings equation (i.e. $Cov(g, \upsilon_0) \neq 0$ and/or $Cov(g, \omega) \neq 0$). I estimate the model using a modified version of the two-step method developed by Heckman (1974, 1976, 1978) and Lee (1978, 1979). In the first stage, the migration equation (Equation 21) 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 22) as a regressor, and the earning equation is estimated by two stage least squares using instrumental variables for $g$. Identification requires that the instruments include at least one exogenous variable that is not included in $X$ (Wooldrigde, 2002).

# 4 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 migrated to a different state within the past year, and the second variable, *mig5*, indicates whether the individual migrated to a different state within the past five years. Although the respondent's migration activity within the past year is asked in every year of the CPS with a few exceptions, her migration activity over the past five years is not available in every year. Currently, the five-year migration indicator is available for 1980, 1985, 1995 and 2005.

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 as 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. Therefore, I assume that the sample of migrants defined by *mig1* consists of a greater proportion of individuals who will eventually return to their home states. In the empirical analysis, I use both state-to-state migration indicators. 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* can enable 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. Both household earnings and wages have been used by previous studies investigating the return to migration. For instance, while Ham et al. focus on wage growth due to migration, Tunali (2000) studies the change in household earnings resulting from 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.

State migration 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). These calculations are based on the number of people who reported having moved across states between 1995 and 2000. The median age in each state also comes from the U.S. Census Bureau's tabulations on the U.S. Census 2000 (Meyer, 2001). The state-level data on migration rates and median age are linked to the individual-level data by the non-migrant's state of residence and the migrant's state of origin.

The explanatory variables used in the analysis are listed in Table 1. The number of children in the household and whether the individual owns her dwelling are used as factors affecting migration cost and thus are excluded from the wage equation. The sample is limited to civilians aged 15 or older who are in the labor force. It contains 97,864 observations.

According to the statistics reported in Table 1, 2.6 percent of the sample reported living in a different state than they did a year ago, and 8.1 percent reported living in a different state than they did five years ago. Table 1 also lists the average characteristics of movers and stayers where migration is defined by the individuals" movement across states within the past year. These descriptive statistics reveal that movers 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. Both of these findings suggest that stayers tend to have higher moving costs generated by selling home, changing children's schools and finding work for the spouse at the destination. Movers and stayers are also quite different with respect to their employment status. Movers are more than twice as likely to be unemployed during the week of the interview than stayers, and they are less likely to be self-employed after the move compared to stayers.

# 5 Results

## 5.1 The Probability of Migration

Table 2 presents the maximum likelihood estimation results of the migration equation. Specifications (I) and (II) estimates come from the estimation of an ordinary probit regression with and without the gross migration rate in the state of origin as an explanatory variable. In specification (III), the gross migration rate is instrumented using the median age in the state of origin. 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 marginal effects reported in column 1 indicate that the probability of migration across states is positively affected by age and negatively affected by the number of years in the labor market. Men are 0.26 percent more likely to move than women while race and marital status do not significantly affect the decision to migrate. Number of children and owning a house have a negative effect on one's likelihood to migrate. This result is consistent with the notion that the number of children and owning a house lead to higher moving costs and thus hinder migration. People who report being unemployed are 1.6 percent 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.

The marginal effects reported in column 2 refer to the estimation of specification (I) probit using *mig5* as the dependent variable. As stated above, one of the key differences between *mig1* and *mig5* is that due to its wider span, *mig5* indicator points to migrant sample which consists of a smaller proportion of individuals who would return to their home state after the migration. *Mig1*, on the other hand, yields a migrant sample consisting of a greater proportion of people who may become return migrants in the future. The marginal effects, found using *mig1* as the dependent variable, retain their signs when *mig5* is used as a dependent variable. However, the magnitudes of the marginal effects become greater for most variables when migration status is determined by migration activity over the past five years. 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, and these effects seem to be diluted when the return migrants make up a greater portion of the migrant sample. Although the line of inquiry regarding return migration is not pursued in this paper, these results provide indirect support for other research showing that return migrants are systematically different from permanent migrants.

Specification (II) probit results reveal that gross migration rates in the state of origin are positively related to the probability of migration. However, a positive relationship between a state's gross migration rate and an individual's decision to migrate may be spurious in nature if the state migration rate is affected by local economic conditions, which are also likely to influence the resident's migration decision. In order to address the potential endogeneity of the gross migration rate, I use median age in the state as an instrument for the state's gross migration rate (Specification (III)). The results of the IV probit regression presented in columns (5) and (6) of Table 2 show that the gross migration rate in the home state has a positive impact on an individual's likelihood to migrate to another state. According to these estimates, increasing the gross migration rate in the home state by 10 percent increases a resident's probability of migrating out of the state by 2.7 percent. The magnitude of this effect is higher when *mig5* is used as the migration indicator. In that case, increasing the gross migration rate in the home state by 10 percent leads to a 5.8 percent increase in the a resident's propensity to move out of the state. These results suggest that greater access to information about the destination labor market positively affects one's probability of migrating to another state. The effect is stronger in magnitude when the migrant sample includes a smaller share of return migrants.

## 5.2 The Return to Migration

The calculation of the return to migration requires the estimation of the wage equation for movers and stayers. The wage equation is estimated as the second step in a two-step procedure, which is designed to take into account the self-selection of movers. The parameter estimates of specification (I) in Table 2 are used in calculating the inverse mill's ratio for each individual in the sample, and then the estimated inverse mill's ratio is added to the wage equation as an exogenous regressor. In Table 3, I present the parameter estimates of the wage equation using OLS and 2SLS. In the 2SLS estimation, the median age in state is used as an instrument for the gross migration rate. The econometric model is estimated using two separate indicators for state-to-state migration: *mig1* and *mig5*.

The OLS results with *mig1* as the migration indicator reveal that age, college education, being male, white and married positively affect wages of both movers and stayers. Managers have higher wages among migrants, and service and production workers have lower wages among stayers. The convexity 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 effect on the wages of stayers while its effect is insignificant on the wages of migrants. The OLS estimates also show that the gross migration rate in the state of origin does not have a statistically significant effect on the wages of migrants while it has a significantly negative effect on the wages of stayers.

When the gross migration rate is instrumented using median age in the state, the coefficient of the gross migration rate changes in magnitude for both migrants and non-migrants. In particular, the coefficient becomes remains statistically insignificant for the mover subsample. However, it becomes more negative for the stayer subsample. According to the 2SLS estimates, non-migrants living in states with high gross migration rates earn less compared to non-migrants living in states with low gross migration rates.

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.

The wage equations are also estimated using *mig5* as the state-to-state migration indicator. The last four columns of Table 3 show the results of these estimations. Most of the coefficients retain their sign and significance when *mig5* is used as the migration indicator. One interesting finding is that the gross migration rate does not have a statistically significant effect on the migrant wages when *mig5* is used in defining migration. This finding is not surprising when one considers the difference between the migrant samples defined by *mig1* and *mig5*. *Mig5* is likely to generate a migrants sample that consists of a smaller proportion of people who may become return migrants. More information about destination labor markets is expected to have the biggest impact on the wages of migrants who may become return migrants since these are the people who potentially made wrong migration decisions by overpredicting or underpredicting their post-migration wages. Additional information would enable such individuals to correct their migration decisions. The subsample of migrants defined by *mig5* is likely to consist of a relatively small proportion of people who potentially made bad migration decisions. Therefore, it is plausible to expect information about other labor markets to have a small influence on the wages of these migrants. The gross migration rate continues to have a negative and significant effect on the wages of stayers when *mig5* is used as the migration indicator.

Next, I calculate the *return to migration* by taking the difference between the wages earned by migrants and the wages that they would have earned at origin had they chosen not to migrate. The latter is calculated based on the 2SLS estimates for the migrant subsample presented in Table 3. 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. When migration is defined by movement across states within the past year, the average return to migration among migrants, measured as the difference in log wages, is 0.074, and the median return among migrants is 0.033. Approximately 44 percent of migrants realize a negative return to migration. When migration is defined by movement across states within the past five years, the average return to migration among migrants, measured is 0.044. Although the average return calculated using *mig5* is smaller than the one calculated using *mig1*, the median return among migrants is the same in both cases (0.033). As expected, when migration is defined by *mig5*, a smaller proportion of migrants, namely 27 percent, experience a negative return to their migration decisions.

The theoretical model presented in this paper has several implications regarding the changes in the return to migration brought about by changes in the availability of information and by changes in wage inequality at the destination. In order to empirically test these implications, I compare the average return to migration across different groups of migrants. First, I present the average return to migration by educational attainment (Table 4). Standard errors are calculated using bootstrap with 1000 repetitions. Results using *mig1* as the migration indicator shows that the average return to migration is highest for highschool dropouts. Highschool graduates experience higher returns to migration than college graduates and lower returns to migration than high school dropouts. Individuals with some college credits gain the lowest wage growth as a result of migration compared to other educational groups. With the exception of the individuals in the 'some college' category, the return to migration seems to decrease with education. 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 people with low education face higher migration costs compared to higher educated people, then they would require higher return to migration to justify migrating and incurring the associated high migration costs. In such a case, we would observe a higher average return to migration among migrants with low education.

The observed higher return to migration among lower educated migrants 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 highschool 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.

When migration is defined as changing states within the last five years, the average return to migration in the entire migrant subsample is smaller relative to the case where *mig1* is used as the migration indicator. The sample of migrants defined by *mig5* consists of a smaller portion of individuals who would eventually return back to their home states. Therefore, a comparison of the average return to migration under the *mig1* and *mig5* columns may suggest that return migrants tend to have higher wage growth due to migration than those whose moves are permanent. Based on this result, one can infer that the return migration need not be driven by lower than expected wages at the destination but perhaps by other factors such as those related to one's adjustment to her new surroundings.

I also calculate the average return to migration among various migrant categories based on the gross migration and outmigration rates in the migrants' states of origin. The cut-off migration rates used in the definition of these categories are the gross migration and outmigration rates which correspond to the 25th percentile, the median and the 75th percentile observation in the migrant subsample. The findings presented in Table 5 suggest that the return to migration increases with the migration rate in home state. In fact, individuals from states with gross migration rates less than 13 percent and outmigration rates less than 7.17 percent on average experience a negative return to migration. Since residents of states with high migration rates are likely to have easier access to information about other labor markets, these findings suggest that individuals with better information about other labor markets outside their home states experience higher returns to their migration decisions. However, different average return to migration may arise for states with different migration rates due to variation in resident characteristics across different states. For instance, if states with high gross migration rates also have a high percentage of residents with less than a high school degree, the high return to migration observed in these states may be due to the substantial proportion of high school dropouts, who based on the Table 4 results tend to reap high returns to their migration decisions.

In order to better tease out the relationship between gross migration rates and the return to migration, I perform a comparative static exercise in which I increase the gross migration rate for all migrants in the sample by 10 percent. Then, I calculate counterfactual post-migration wages for migrants based on the parameters estimates of the wage equation. I use the counterfactual migrant wages to calculate the return to migration and compare these returns to the ones presented in Table 5. Any change in the return to migration can then be attributed to the 10 percent increase in the gross migration rate because the gross migration rate is the only variable that changes in this exercise. Furthermore, since the IV regression estimates of the wage equation are used in calculating the post-migration wages under the higher gross migration rate, the resulting change in the return to migration is due to the gross migration rate and not due to any economic factors that are jointly related to the state migration rate and the residents' migration decisions. A limitation of this comparative static exercise, however, is that it does not allow the composition of migrants to change. While an increase in the gross migration rate can potentially change the composition of the migrant sample, the sample of migrants are assumed to remain fixed in this exercise after the increase in the state migration rate.

Table 6 shows that when the gross migration rate in the migrants' home state rises by 10 percent, the average return to migration also rises in the entire migrant subsample as well as in all migrant subsamples defined by migration rate categories. This provides further support for the hypothesis that increased information about destination labor markets leads to higher returns to migration as migrants are able to secure higher post-migration wages when they have access to more information about job opportunities in other states.

Finally, I investigate the relationship between the 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 subsamples based on 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, the median and the 75th percentile observation in the migrant subsample.

Table 7 presents the average return to migration experienced by migrants stratified by income inequality in their destination states. In general, the results suggest a negative relationship between income inequality at the destination state and return to migration. Migrants who have migrated to states with high income inequality seem to experience a smaller wage growth due to migration. According to the theoretical model, high income inequality at the destination leads to greater uncertainty faced by individuals regarding their post-migration wages. Therefore, the results in Table 7 suggest that migrants, who face greater uncertainty about their post-migration wages, experience lower wage growth due to migration.

# 6 Conclusion

In this paper, I present a model in which individuals make migration decisions under incomplete information about the destination labor market. The model is based on the assumption that information about destination labor market conditions is an important factor in the migration decision, and one of the channels that carry such information to people is the network of friends and family who have migrated previously. The model's implications reveal that increased information about destination labor markets can have both positive and negative effects 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.

The model's implications regarding the effect of information on the probability of migration critically depend on the population mean of the home and destination wages as well as the moving costs. If the population mean of wages at origin exceeds the population mean of the wages at destination, increased information regarding destination labor market conditions leads to lower probability of migration. Otherwise, probability of migration increases with increased access to information about other labor markets. Information about the destination affects the return to migration through a composition and a scale effect. The composition effect reflects the impact of information on the return to migration through its effect on the composition of the migrant sample, conditional on the rate of migration. The scale effect describes the effect of information on the return to migration through its impact on the size of the migrant sample. While the composition effect is always positive, the scale effect can be both 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. I use gross migration rate in one's home state as a proxy for the level of information available to the individual regarding destination labor markets. This choice is based on the reasoning that people learn about other labor markets through their friends and neighbors who have migrated to or from other locations. Therefore, people who live in states with high gross migration rates have greater access to information about labor market conditions in other states since residents of such states are more likely to come into contact with people who have moved to or from other states. The gross migration rate, however, is likely to be endogenous in both the migration and wage equations since it may be correlated with unobservables that can affect both the individual's migration decision and her wage. The median age in the state of origin is used as an instrument for the gross migration rate in the analysis.

The estimation results indicate that increased access to information about destination labor market conditions increases one's probability of migration to another state. Information has a statistically insignificant effect on the wages of migrants and a significantly negative effect on the wages of stayers. Furthermore, the findings suggest that more information about other labor markets leads to an increase in the return to migration among migrants. The return to migration also varies with educational attainment. According to the empirical results, migrants with lower education experience higher return to migration, measured by wage growth due to migration. In addition, the average return to migration is lower among people who migrate to states with high income inequality, suggesting that higher uncertainty about destination wages leads to lower return to migration among migrants.

This paper emphasizes the role of information in migration outcomes. The theoretical model provides a framework in which to study this concept. The empirical analysis provided in this paper suggests that the information structure available to individuals at the time of the 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 available to different groups can explain the variation in migration outcomes observed across migrants with different demographic and socioeconomic characteristics.

## 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.

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

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

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

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

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. "Human Migration: Theory, Models, and Empirical Studies," *Journal of Regional Science*, (1985), 25: 521-544.

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.

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.

Massey, D., J. Arango, G. Hugo, A. Kouaouci, A. Pellegriono and J. Taylor. "Theories of International Migration: A Review and Appraisal," *Population and Development Review*, (1993), 19: 431-466.

Meyer, Julie. "Age: 2000," Census 2000 Brief, (Oct. 2001).

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.

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.

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

Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA, (2002).

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

$$P = \Pr\left(\upsilon_1 + \overline{\varepsilon} - \upsilon_0 - \upsilon_c > \mu_0 + \mu_c - \mu_1\right) \tag{24}$$

If we let $\eta = \upsilon_1 + \overline{\varepsilon} - \upsilon_0 - \upsilon_c$, then

$$P = \Pr(\eta > \mu_0 + \mu_c - \mu_1) = 1 - \Phi(z) \tag{25}$$

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

$$\frac{\partial P}{\partial n} = -\phi(z) \cdot \frac{\partial z}{\partial \sigma_\eta} \cdot \frac{\partial \sigma_\eta}{\partial n} \tag{26}$$

First, I will simplify the expression for $\sigma_\eta$. By definition, $\sigma_\eta$ can be stated as

$$\sigma_\eta = \sqrt{Var(\upsilon_1 + \overline{\varepsilon} - \upsilon_0 - \upsilon_c)} \tag{27}$$

$$\begin{aligned} &Var(\upsilon_1 + \overline{\varepsilon} - \upsilon_0 - \upsilon_c) = \\ &\qquad Var(\upsilon_1) + Var(\overline{\varepsilon}) + Cov(\upsilon_1, \overline{\varepsilon}) - Cov(\upsilon_1, \upsilon_0) - Cov(\upsilon_1, \upsilon_c) \\ &\qquad\qquad - Cov(\upsilon_0, \overline{\varepsilon}) - Cov(\upsilon_c, \overline{\varepsilon}) + Cov(\upsilon_0, \upsilon_c) + Var(\upsilon_0) + Var(\upsilon_c) \end{aligned}$$

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

$$\begin{aligned} &Var(\upsilon_1 + \overline{\varepsilon} - \upsilon_0 - \upsilon_c) = \\ &\quad = Var(\upsilon_1) + Var(\overline{\varepsilon}) - Cov(\upsilon_1, \upsilon_0) - Cov(\upsilon_1, \upsilon_c) + Cov(\upsilon_0, \upsilon_c) + Var(\upsilon_0) + Var(\upsilon_c) \\ &\qquad = \sigma_1^2 + \sigma_{\overline{\varepsilon}}^2 + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c} = \sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c} \end{aligned}$$

Then, $\sigma_\eta$ is given by

$$\sigma_\eta = \left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c}\right)^{0.5} \tag{28}$$

As a result, the third factor in the expression of $\frac{\partial P}{\partial n}$ can be written as

$$\frac{\partial \sigma_\eta}{\partial n} = 0.5\left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c}\right)^{-0.5} \cdot \left(-\frac{\sigma_\varepsilon^2}{n^2}\right) \tag{29}$$

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

$$\frac{\partial z}{\partial \sigma_\eta} = \left(-\frac{\mu_0 + \mu_c - \mu_1}{\sigma_\eta^2}\right) \tag{30}$$

Finally, the derivative of $n$ with respect to $P$ is given by

$$\begin{aligned}\frac{\partial P}{\partial n} &= -\phi(z) \cdot \frac{\partial z}{\partial \sigma_\eta} \cdot \frac{\partial \sigma_\eta}{\partial n} \\ &= \phi(z) \cdot \left(-\frac{\mu_0 + \mu_c - \mu_1}{\sigma_\eta^2}\right) \cdot 0.5\left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c}\right)^{-0.5} \cdot \left(-\frac{\sigma_\varepsilon^2}{n^2}\right)\end{aligned} \tag{31}$$

Since $\phi(z) > 0$, $\sigma_\eta^2 > 0$, $\left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c}\right)^{-0.5} > 0$, and $\frac{\sigma_\varepsilon^2}{n^2} > 0$,

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

The derivative of $P$ with respect to $\sigma_\varepsilon$ is

$$\begin{aligned}\frac{\partial P}{\partial \sigma_\varepsilon} &= -\phi(z) \cdot \frac{\partial z}{\partial \sigma_\eta} \cdot \frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} \\ &= -\phi(z) \cdot \left(-\frac{\mu_0 + \mu_c - \mu_1}{\sigma_\eta^2}\right) \cdot 0.5\left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c}\right)^{-0.5} \cdot \left(\frac{2\sigma_\varepsilon}{n}\right)\end{aligned} \tag{33}$$

Since $\phi(z) > 0$, $\sigma_\eta^2 > 0$, $\left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c}\right)^{-0.5} > 0$, and $\frac{2\sigma_\varepsilon}{n} > 0$,

$$\begin{aligned}\frac{\partial P}{\partial \sigma_\varepsilon} &> 0 \;\; if \;\; \mu_0 + \mu_c - \mu_1 > 0 \\ \frac{\partial P}{\partial \sigma_\varepsilon} &< 0 \;\; if \;\; \mu_0 + \mu_c - \mu_1 < 0\end{aligned} \tag{34}$$

## A2. The Return to Migration

The return to migration, $R$, can be expressed as

$$R = E(y_1|M = 1) - E(y_0|M = 1) \tag{35}$$

The first term in this expression is the expected value of destination earning conditional on migration, and the second term is the expected value of earnings at origin if the migrants had stayed at origin, conditional on migration. These two terms can be expressed as follows:

$$E(y_1|M = 1) = \mu_1 + E(b|\eta > \mu_0 + \mu_c - \mu_1) = \mu_1 + \rho_{\omega\eta}\sigma_\omega\lambda_1 \tag{36}$$

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

$$E(y_0|M = 1) = \mu_0 + E(\upsilon_0|\eta > \mu_0 + \mu_c - \mu_1) = \mu_0 + \rho_{0\eta}\sigma_0\lambda_1 \tag{37}$$

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

$$\begin{aligned} R &= E(y_1|M = 1) - E(y_0|M = 1) \\ &= \mu_1 - \mu_0 + \left(\rho_{\omega\eta}\sigma\omega - \rho_{0\eta}\sigma_0\right)\lambda_1 \end{aligned} \tag{38}$$

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

$$\rho_{0\eta} = Corr(\upsilon_0, \eta) = \frac{Cov(\upsilon_0, \eta)}{\sigma_0\sigma_\eta} \tag{39}$$

$$\begin{aligned} Cov(\upsilon_0, \eta) &= Cov(\upsilon_0, \upsilon_1 + \overline{\varepsilon} - \upsilon_0 - \upsilon_c) \\ &= Cov(\upsilon_0, \upsilon_1) + Cov(\upsilon_0, \overline{\varepsilon}) - \sigma_0^2 - Cov(\upsilon_0, \upsilon_c) \end{aligned} \tag{40}$$

Since $Cov(\upsilon_0, \overline{\varepsilon})$,

$$Cov(\upsilon_0, \eta) = \sigma_{01} - \sigma_0^2 - \sigma_{0c} \tag{41}$$

$$\rho_{0\eta}\sigma_0\lambda_1 = \frac{\sigma_{01} - \sigma_0^2 - \sigma_{0c}}{\sigma_0\sigma_\eta}\sigma_0\lambda_1 = \frac{\lambda_1}{\sigma_\eta}\left(\sigma_{01} - \sigma_0^2 - \sigma_{0c}\right) \tag{42}$$

$$\rho_{\omega\eta} = Corr(\omega, \eta) = Corr(\upsilon_1 + \varepsilon, \upsilon_1 + \overline{\varepsilon} - \upsilon_0 - \upsilon_c) = \frac{Cov(\upsilon_1 + \varepsilon, \upsilon_1 + \overline{\varepsilon} - \upsilon_0 - \upsilon_c)}{\sigma_\omega\sigma_\eta} \tag{43}$$

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

Since $Cov(\upsilon_1,\bar{\varepsilon}) = Cov(\varepsilon,\bar{\varepsilon}) = 0$,

$$Cov(\upsilon_1+\varepsilon, \upsilon_1+\bar{\varepsilon}-\upsilon_0-\upsilon_c) = \sigma_1^2 - \sigma_{10} - \sigma_{1c} + \sigma_{1\varepsilon} - \sigma_{0\varepsilon} - \sigma_{\varepsilon c} \tag{44}$$

Then,

$$\rho_{\omega\eta}\sigma_\omega\lambda_1 = \frac{\sigma_1^2 - \sigma_{10} - \sigma_{1c} + \sigma_{1\varepsilon} - \sigma_{0\varepsilon} - \sigma_{\varepsilon c}}{\sigma_\omega\sigma_\eta}\sigma_\omega\lambda_1 = \frac{\lambda_1}{\sigma_\eta}\left(\sigma_1^2 - \sigma_{10} - \sigma_{1c}\right) \tag{45}$$

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

$$R = \mu_1 - \mu_0 + \frac{\lambda_1}{\sigma_\eta}Var(\upsilon_1 - \upsilon_0) \tag{46}$$

Given Equation 46, the derivative of $R$ with respect to $n$ is given by

$$\frac{\partial R}{\partial n} = \left[\frac{1}{\sigma_\eta}\cdot\frac{\partial\lambda_1}{\partial n} - \frac{\lambda_1}{\sigma_\eta^2}\cdot\frac{\partial\sigma_\eta}{\partial n}\right]Var(\upsilon_1 - \upsilon_0) \tag{47}$$

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

$$\frac{\partial\lambda_1}{\partial n} = \frac{\partial\lambda_1}{\partial z}\cdot\frac{\partial z}{\partial\sigma_\eta}\cdot\frac{\partial\sigma_\eta}{\partial n} \tag{48}$$

The three factors in this equation have the following signs:

$$\frac{\partial\lambda_1}{\partial z} > 0 \tag{49}$$

As shown above,

$$\begin{aligned} \frac{\partial z}{\partial\sigma_\eta} &> 0 \;\; if \;\; \mu_0 + \mu_c - \mu_1 < 0 \\ \frac{\partial z}{\partial\sigma_\eta} &< 0 \;\; if \;\; \mu_0 + \mu_c - \mu_1 > 0 \end{aligned} \tag{50}$$

and

$$\frac{\partial\sigma_\eta}{\partial n} = 0.5\left(\sigma_1^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_0^2 + \sigma_c^2 - \sigma_{01} - \sigma_{1c} + \sigma_{0c}\right)^{-0.5}\cdot\left(-\frac{\sigma_\varepsilon^2}{n^2}\right) < 0 \tag{51}$$

Therefore,

$$\begin{aligned}\frac{\partial \lambda_1}{\partial n} &> 0 \quad if \quad \mu_0+\mu_c-\mu_1>0\\ \frac{\partial \lambda_1}{\partial n} &< 0 \quad if \quad \mu_0+\mu_c-\mu_1<0\end{aligned} \tag{52}$$

Then,

$$\begin{aligned}\frac{\partial R}{\partial n} &< 0 \quad if \quad \mu_0+\mu_c-\mu_1<0 \quad and \quad \frac{1}{\sigma_\eta}\cdot\frac{\partial \lambda_1}{\partial n}>\frac{\lambda_1}{\sigma_\eta^2}\cdot\frac{\partial \sigma_\eta}{\partial n}\\ \frac{\partial R}{\partial n} &< 0 \quad otherwise\end{aligned} \tag{53}$$

Equation 46 can also be used to calculate the derivative of $R$ with respect to $\sigma_\varepsilon$.

$$\frac{\partial R}{\partial \sigma_\varepsilon}=\left[\frac{1}{\sigma_\eta}\cdot\frac{\partial \lambda_1}{\partial \sigma_\varepsilon}-\frac{\lambda_1}{\sigma_\eta^2}\cdot\frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon}\right]Var(\upsilon_1-\upsilon_0) \tag{54}$$

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,

$$\frac{\partial \lambda_1}{\partial \sigma_\varepsilon}=\frac{\partial \lambda_1}{\partial z}\cdot\frac{\partial z}{\partial \sigma_\eta}\cdot\frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon} \tag{55}$$

$$\frac{\partial \lambda_1}{\partial z}>0 \tag{56}$$

$$\frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon}>0 \tag{57}$$

$$\begin{aligned}\frac{\partial z}{\partial \sigma_\eta} &> 0 \quad if \quad \mu_0+\mu_c-\mu_1<0\\ \frac{\partial z}{\partial \sigma_\eta} &< 0 \quad if \quad \mu_0+\mu_c-\mu_1>0\end{aligned} \tag{58}$$

Therefore,

$$\begin{aligned}\frac{\partial R}{\partial \sigma_\varepsilon} &> 0 \quad if \quad \mu_0+\mu_c-\mu_1<0 \quad and \quad \frac{1}{\sigma_\eta}\cdot\frac{\partial \lambda_1}{\partial \sigma_\varepsilon}>\frac{\lambda_1}{\sigma_\eta^2}\cdot\frac{\partial \sigma_\eta}{\partial \sigma_\varepsilon}\\ \frac{\partial R}{\partial \sigma_\varepsilon} &< 0 \quad otherwise\end{aligned} \tag{59}$$

## 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

$$E(y_1^e - y_1) = E(\mu_1 + \upsilon_1 + \overline{\varepsilon} - \mu_1 - \upsilon_1 - \varepsilon) = \mu_1 - \mu_1 + E(\upsilon_1 + \overline{\varepsilon} - \upsilon_1 - \varepsilon) \tag{60}$$

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

$$E(y_1^e - y_1) = 0 \tag{61}$$

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:

$$E(y_1^e|M = 1) = \mu_1 + E(\upsilon_1 + \overline{\varepsilon}|\eta > \mu_0 + \mu_c - \mu_1 \tag{62}$$

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

$$E(y_1^e|M = 1) = \mu_1 + \rho_{a\eta}\sigma_a\lambda_1 \tag{63}$$

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

$$\rho_{a\eta} = \frac{Cov(\upsilon_1 + \overline{\varepsilon}, \upsilon_1 + \overline{\varepsilon} - \upsilon_0 - \upsilon_c)}{\sigma_a\sigma_\eta} \tag{64}$$

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

Since $Cov(\upsilon_1, \overline{\varepsilon}) = Cov(\upsilon_0, \overline{\varepsilon}) = Cov(\upsilon_c, \overline{\varepsilon}) = 0$,

$$Cov(\upsilon_1 + \overline{\varepsilon}, \upsilon_1 + \overline{\varepsilon} - \upsilon_0) = \sigma_1^2 - \sigma_{01} - \sigma_{1c} + \frac{\sigma_\varepsilon^2}{n} \tag{65}$$

Then,

$$E(y_1^e|M = 1) = \mu_1 + \frac{\sigma_1^2 - \sigma_{01} - \sigma_{1c} + \frac{\sigma_\varepsilon^2}{n}}{\sigma_\eta}\lambda_1 \tag{66}$$

and the prediction error among migrants can be stated as

$$E(y_1^e|M=1) - E(y_1|M=1) = \\ \mu_1 + \frac{\lambda_1}{\sigma_\eta}\left(\sigma_1^2 - \sigma_{01} - \sigma_{1c} + \frac{\sigma_\varepsilon^2}{n}\right) - \mu_1 - \frac{\lambda_1}{\sigma_\eta}\left(\sigma_1^2 - \sigma_{01} - \sigma_{1c}\right) = \frac{\lambda_1}{\sigma_\eta}\left(\frac{\sigma_\varepsilon^2}{n}\right) > 0$$

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

**Table 1: Variable Definitions and Descriptive Statistics**

| Variable Name | Variable Definition | Mean (Standard Deviation) | | | |
|---|---|---|---|---|---|
| | | Whole Sample | Migrants | Stayers | Difference |
| Mig1 | =1 if R moved to another state within the past year | 0.026 | | | |
| | | (0.001) | | | |
| Mig5 | =1 if R moved to another state within past five years | 0.081 | | | |
| | | (0.001) | | | |
| Log(wage) | Log of hourly wage | 2.694 | 2.645 | 2.695 | 0.050 |
| | | (0.003) | (0.020) | (0.003) | (0.020) |
| Age | Age in years | 40.597 | 34.739 | 40.753 | 6.014 |
| | | (0.053) | (0.303) | (0.053) | (0.308) |
| Experience | Age - years of schooling - 6 | 23.244 | 17.159 | 23.405 | 6.246 |
| | | (0.052) | (0.309) | (0.053) | (0.314) |
| High school | =1 if highest degree earned is high school diploma | 0.588 | 0.530 | 0.590 | 0.060 |
| | | (0.002) | (0.013) | (0.002) | (0.013) |
| College | =1 if highest degree earned is bachelor's or higher | 0.295 | 0.353 | 0.293 | -0.059 |
| | | (0.002) | (0.013) | (0.002) | (0.013) |
| Male | =1 if male | 0.534 | 0.561 | 0.533 | -0.027 |
| | | (0.002) | (0.013) | (0.002) | (0.013) |
| White | =1 if white | 0.823 | 0.786 | 0.824 | 0.038 |
| | | (0.001) | (0.011) | (0.001) | (0.011) |
| Married | =1 if married | 0.595 | 0.456 | 0.599 | 0.143 |
| | | (0.002) | (0.013) | (0.002) | (0.013) |
| Number of children | Number of children in the household | 0.838 | 0.631 | 0.843 | 0.212 |
| | | (0.004) | (0.025) | (0.004) | (0.025) |

| | | | | | |
|---|---|---|---|---|---|
| Own house | =1 if household owns home | 0.730 | 0.376 | 0.740 | 0.363 |
| | | (0.002) | (0.013) | (0.002) | (0.013) |
| Center city | =1 if respondent lives in central city in metro area | 0.836 | 0.856 | 0.836 | -0.021 |
| | | (0.001) | (0.009) | (0.001) | (0.009) |
| Self-employed | =1 if respondent is self-employed | 0.099 | 0.069 | 0.100 | 0.030 |
| | | (0.001) | (0.007) | (0.001) | (0.008) |
| Unemployed | =1 if employment status is unemployed | 0.041 | 0.083 | 0.040 | -0.043 |
| | | (0.001) | (0.007) | (0.001) | (0.007) |
| Manager | =1 if R works in management occupation | 0.342 | 0.338 | 0.342 | 0.003 |
| | | (0.002) | (0.012) | (0.002) | (0.012) |
| Service | =1 if R works in a service occupation | 0.417 | 0.428 | 0.417 | -0.012 |
| | | (0.002) | (0.013) | (0.002) | (0.013) |
| Production | =1 if R works in production occupation | 0.132 | 0.124 | 0.132 | 0.009 |
| | | (0.001) | (0.008) | (0.001) | (0.008) |
| Northeast | =1 if respondent lives in northeast US | 0.188 | 0.150 | 0.189 | 0.039 |
| | | (0.001) | (0.009) | (0.002) | (0.009) |
| Midwest | =1 if respondent lives in midwest US | 0.232 | 0.189 | 0.233 | 0.045 |
| | | (0.002) | (0.010) | (0.002) | (0.010) |
| South | =1 if respondent lives in southern US | 0.352 | 0.406 | 0.351 | -0.055 |
| | | (0.002) | (0.013) | (0.002) | (0.013) |
| *Sample Size* | | 97,864 | 2,374 | 95,490 | |

Notes:

1. Standard errors of means are in parentheses.
2. Sampling weights are used in the calculation of the statistics presented in this table.
3. Mig1 migration indicator is used in the definition of migration; therefore, migrants moved to a different state within the past year.

**Table 2: Estimates of the Migration Equation**

| | (I) | | (II) | | (III)[3] | |
|---|---|---|---|---|---|---|
| | Mig1 | Mig5 | Mig1 | Mig5 | Mig1 | Mig5 |
| Age | 0.0044*** | 0.0190*** | 0.0045*** | 0.0190*** | 0.0047*** | 0.0191*** |
| | (0.0009) | (0.0018) | (0.0009) | (0.0018) | (0.0010) | (0.0018) |
| Age squared | 0.0000*** | -0.0001*** | 0.0000*** | -0.0002*** | 0.0000*** | -0.0002*** |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| Experience | -0.0035*** | -0.0148*** | -0.0035*** | -0.0146*** | -0.0036*** | -0.0146*** |
| | (0.0007) | (0.0014) | (0.0007) | (0.0014) | (0.0008) | (0.0014) |
| Experience squared | 0.0000*** | 0.0001*** | 0.0000*** | 0.0001*** | 0.0000*** | 0.0001*** |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| High school | -0.0030 | -0.0175*** | -0.0036 | -0.0186*** | -0.0046 | -0.0196*** |
| | (0.0029) | (0.0055) | (0.0028) | (0.0054) | (0.0031) | (0.0055) |
| College | 0.0009 | -0.0091 | 0.0002 | -0.0099 | 0.0000 | -0.0100 |
| | (0.0048) | (0.0083) | (0.0046) | (0.0082) | (0.0050) | (0.0082) |
| Male | 0.0026** | 0.0089*** | 0.0026** | 0.0089*** | 0.0029** | 0.0091*** |
| | (0.0011) | (0.0021) | (0.0011) | (0.0021) | (0.0012) | (0.0021) |
| White | 0.0010 | -0.0021 | 0.0010 | -0.0019 | 0.0009 | -0.0021 |
| | (0.0013) | (0.0026) | (0.0013) | (0.0026) | (0.0014) | (0.0026) |
| Married | 0.0003 | 0.0070*** | 0.0004 | 0.0069*** | 0.0004 | 0.0069*** |
| | (0.0013) | (0.0024) | (0.0013) | (0.0024) | (0.0014) | (0.0024) |
| Number of children | -0.0018*** | -0.0062*** | -0.0017*** | -0.0058*** | -0.0016*** | -0.0054*** |
| | (0.0006) | (0.0010) | (0.0005) | (0.0010) | (0.0006) | (0.0010) |
| Own house | -0.0392*** | -0.0926*** | -0.0392*** | -0.0934*** | -0.0429*** | -0.0957*** |
| | (0.0020) | (0.0032) | (0.0020) | (0.0032) | (0.0024) | (0.0032) |
| Center city | -0.0009 | -0.0076*** | 0.0001 | -0.0039 | 0.0028* | 0.0004 |
| | (0.0015) | (0.0028) | (0.0015) | (0.0027) | (0.0015) | (0.0028) |
| Self-employed | -0.0005 | -0.0031 | -0.0005 | -0.0032 | -0.0008 | -0.0036 |
| | (0.0022) | (0.0035) | (0.0021) | (0.0034) | (0.0022) | (0.0034) |
| Unemployed | 0.0161*** | 0.0296*** | 0.0161*** | 0.0291*** | 0.0171*** | 0.0292*** |
| | (0.0034) | (0.0061) | (0.0034) | (0.0059) | (0.0036) | (0.0059) |
| Manager | 0.0001 | 0.0070* | 0.0004 | 0.0076* | 0.0009 | 0.0082* |
| | (0.0022) | (0.0041) | (0.0021) | (0.0040) | (0.0023) | (0.0041) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Service | 0.0006 | 0.0045 | 0.0009 | 0.0048 | 0.0010 | 0.0050 |
| | (0.0020) | (0.0037) | (0.0019) | (0.0037) | (0.0021) | (0.0037) |
| Production | 0.0005 | -0.0027 | 0.0009 | -0.0014 | 0.0018 | -0.0002 |
| | (0.0022) | (0.0041) | (0.0022) | (0.0041) | (0.0025) | (0.0042) |
| Northeast | -0.0053*** | -0.0192*** | -0.0021 | -0.0078** | 0.0082*** | 0.0071* |
| | (0.0014) | (0.0026) | (0.0016) | (0.0030) | (0.0029) | (0.0041) |
| Midwest | -0.0025* | -0.0128*** | 0.0012 | -0.0002 | 0.0133** | 0.0165*** |
| | (0.0014) | (0.0026) | (0.0017) | (0.0030) | (0.0034) | (0.0045) |
| South | 0.0035** | 0.0114*** | 0.0041*** | 0.0139*** | 0.0052*** | 0.0156*** |
| | (0.0014) | (0.0027) | (0.0014) | (0.0027) | (0.0016) | (0.0027) |
| Gross migration rate | | | 0.0007*** | 0.0026*** | 0.0027*** | 0.0058*** |
| | | | (0.0001) | (0.0002) | (0.0004) | (0.0007) |

Notes:

1. Mig1 and Mig5 refer to the dependant variable used in the regression.

2. Robust standard errors are given in parentheses.

3. Probit regressions are weighted.

4. In specification (III), gross migration rate is instrumented with the median age in state.

5. Marginal effects are reported.

6. (*) Significant at 10% level, (**) Significant 5% level, (***) Significant 1% level.

**Table 3: Estimates of the Wage Equation**

| | Migration Indicator is *Mig1* | | | | Migration Indicator is *Mig5* | | | |
|---|---|---|---|---|---|---|---|---|
| | **OLS** | | **2SLS** | | **OLS** | | **2SLS** | |
| | **Movers** | **Stayers** | **Movers** | **Stayers** | **Movers** | **Stayers** | **Movers** | **Stayers** |
| Age | 0.0652** | 0.1107*** | 0.0698** | 0.1105*** | 0.1008*** | 0.1150*** | 0.0981*** | 0.1141*** |
| | (0.0272) | (0.0044) | (0.0316) | (0.0044) | (0.0152) | (0.0047) | (0.0206) | (0.0047) |
| Age squared | -0.0001 | -0.0004*** | -0.0001 | -0.0004*** | -0.0005*** | -0.0005*** | -0.0004** | -0.0004*** |
| | (0.0003) | (0.0000) | (0.0003) | (0.0000) | (0.0002) | (0.0000) | (0.0002) | (0.0000) |
| Experience | -0.0299 | -0.0657*** | -0.0343 | -0.0661*** | -0.0584*** | -0.0677*** | -0.0562*** | -0.0678*** |
| | (0.0213) | (0.0033) | (0.0261) | (0.0033) | (0.0112) | (0.0035) | (0.0156) | (0.0035) |
| Experience squared | -0.0003 | -0.0001* | -0.0003 | -0.0001** | 0.0000 | -0.0001* | 0.0000 | -0.0001** |
| | (0.0003) | (0.0000) | (0.0003) | (0.0000) | (0.0002) | (0.0000) | (0.0002) | (0.0000) |
| High school | 0.0324 | 0.0090 | 0.0092 | 0.0100 | 0.0396 | 0.0132 | 0.0464 | 0.0115*** |
| | (0.0925) | (0.0125) | (0.1216) | (0.0126) | (0.0458) | (0.0130) | (0.0560) | (0.0132) |
| College | 0.2865* | 0.1464*** | 0.2522 | 0.1367*** | 0.2468*** | 0.1607*** | 0.2554*** | 0.1409*** |
| | (0.1492) | (0.0209) | (0.1886) | (0.0212) | (0.0737) | (0.0220) | (0.0844) | (0.0227) |
| Male | 0.2981*** | 0.2591*** | 0.2978*** | 0.2573*** | 0.2669*** | 0.2641*** | 0.2689*** | 0.2608*** |
| | (0.0372) | (0.0053) | (0.0374) | (0.0054) | (0.0193) | (0.0056) | (0.0218) | (0.0057) |
| White | 0.0663* | 0.0363*** | 0.0701* | 0.0353*** | 0.0119 | 0.0298*** | 0.0111 | 0.0300*** |
| | (0.0373) | (0.0061) | (0.0386) | (0.0062) | (0.0203) | (0.0065) | (0.0207) | (0.0065) |
| Married | 0.1089*** | 0.0927*** | 0.1117*** | 0.0972*** | 0.1393*** | 0.1001*** | 0.1396*** | 0.1049*** |
| | (0.0360) | (0.0056) | (0.0372) | (0.0057) | (0.0188) | (0.0058) | (0.0189) | (0.0059) |
| Manager | 0.2293*** | 0.1947*** | 0.2402*** | 0.1895*** | 0.2305*** | 0.2021*** | 0.2269*** | 0.1946*** |
| | (0.0734) | (0.0096) | (0.0875) | (0.0098) | (0.0343) | (0.0101) | (0.0396) | (0.0103) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Service | -0.0059 | -0.0895*** | 0.0030 | -0.0939*** | -0.0614* | -0.0857*** | -0.0634** | -0.0910*** |
| | (0.0640) | (0.0085) | (0.0768) | (0.0087) | (0.0300) | (0.0090) | (0.0321) | (0.0091) |
| Production | 0.0462 | -0.0678*** | 0.0545 | -0.0793*** | -0.0837** | -0.0721*** | -0.0908* | -0.0821*** |
| | (0.0772) | (0.0093) | (0.0828) | (0.0097) | (0.0356) | (0.0098) | (0.0507) | (0.0101) |
| Gross migration rate | 0.0031 | -0.0013*** | 0.0113 | -0.0166*** | 0.0010 | -0.0011*** | -0.0087 | -0.0173*** |
| | (0.0022) | (0.0004) | (0.0290) | (0.0036) | (0.0012) | (0.0004) | (0.0501) | (0.0035) |
| Constant | 0.3905 | -0.3003*** | 0.1674 | 0.0291 | -0.0957 | -0.3035*** | 0.1493 | 0.0747 |
| | (0.4051) | (0.0639) | (0.8910) | (0.1006) | (0.2329) | (0.0672) | (1.2805) | (0.1059) |
| Inv. Mill's Ratio | 0.1198* | -0.2342*** | 0.1272* | -0.2085*** | 0.2565*** | -0.2545*** | 0.2419*** | -0.2081*** |
| | (0.0651) | (0.0100) | (0.0697) | (0.0119) | (0.0374) | (0.0117) | (0.0841) | (0.0156) |

Notes:

1. Standard errors are given in parentheses.

2. In the 2SLS estimation, gross migration is instrumented by the median age in the state.

3. Both regressions are weighted.

4. (*) Significant at 10% level, (**) Significant 5% level, (***) Significant 1% level.

**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 | 0.074 | 0.0045 | 0.0445 | 0.0007 |
| *Education Categories* | | | | |
| Less than highschool | 0.1343 | 0.0127 | 0.0610 | 0.0025 |
| Highschool | 0.0877 | 0.0089 | 0.0436 | 0.0015 |
| Some college | 0.0404 | 0.0079 | 0.0241 | 0.0014 |
| College | 0.0695 | 0.0075 | 0.0550 | 0.0012 |

Notes:

1. Sample includes migrants.
2. Return to migration is calculated by each migrant's post-migration wage minus his/her estimated counterfactual pre-migration wage. The average of the individual returns are presented in the table.
3. Standard errors are calculated from 1000 bootstrap repetitions.

**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** | **Std. Error** | **Mean** | Std. Error |
| *By Gross Migration Rate in the Home State* | | | | |
| Gross migration rate less than or equal to 13 | -0.120 | 0.003 | -0.016 | 0.001 |
| Gross migration rate between 13 and 22.6 | 0.021 | 0.003 | 0.029 | 0.001 |
| Gross migration rate greater than or equal to 22.6 | 0.310 | 0.008 | 0.114 | 0.001 |
| | | | | |
| *By Outmigration Rate in the Home State* | | | | |
| Outmigration rate less than or equal to 7.17 | -0.101 | 0.004 | -0.010 | 0.001 |
| Outmigration rate between 7.17 and 10.69 | 0.013 | 0.003 | 0.026 | 0.001 |
| Outmigration rate greater than or equal to 10.69 | 0.339 | 0.009 | 0.126 | 0.001 |

Notes:

1. Sample includes migrants.
2. Return to migration is calculated by each migrant's post-migration wage minus his/her estimated counterfactual pre-migration wage. The average of the individual returns are presented in the table.
3. Standard errors are calculated from 1000 bootstrap repetitions.
4. State migration rates used in the analysis are obtained from the calculations done by U.S. Census Bureau using the U.S. Census 2000 (Franklin, 2003).

**Table 6: Effect of an Increase in the Gross Migration Rate on Average Return to Migration**

| | Average Return to Migration Among Migrants | Average Return to Migration when Gross Migration Rate Increases by 10% | Change in Average Return to Migration |
|---|---|---|---|
| | **(I)** | **(II)** | **(II)-(I)** |
| *Entire Migrant Subsample* | 0.074 | 0.128 | 0.054 |
| | (0.004) | (0.005) | |
| *By Gross Migration Rate in the Home State* | | | |
| Gross migration rate less than or equal to 13 | -0.120 | -0.086 | 0.034 |
| | (0.003) | (0.003) | |
| Gross migration rate between 13 and 22.6 | 0.021 | 0.069 | 0.049 |
| | (0.003) | (0.003) | |
| Gross migration rate greater than or equal to 22.6 | 0.310 | 0.387 | 0.078 |
| | (0.008) | (0.008) | |

Notes:

1. Sample includes migrants.
2. Return to migration is calculated by each migrant's post-migration wage minus his/her estimated counterfactual pre-migration wage. The average of the individual returns are presented in the table.
3. Standard errors are calculated from 1000 bootstrap repetitions and are given in parentheses.
4. State migration rates used in the analysis are obtained from the calculations done by U.S. Census Bureau using the U.S. Census 2000 (Franklin, 2003).

**Table 7: Average Return to Migration by Income Inequality in the Destination State**

| | Migrants Moved Within Last Year | | Migrants Moved Within Last Five Years | |
|---|---|---|---|---|
| | **Mean** | **Std. Error** | **Mean** | **Std. Error** |
| *By 80:20 Income Ratio in the Destination State* | | | | |
| 80:20 Income Ratio less than or equal to 6.4 | 0.084 | 0.007 | 0.048 | 0.001 |
| 80:20 Income Ratio between 6.4 and 7.6 | 0.067 | 0.006 | 0.043 | 0.001 |
| 80:20 Income Ratio greater than or equal to 7.6 | 0.080 | 0.012 | 0.039 | 0.002 |
| *By 95:20 Income Ratio in the Destination State* | | | | |
| 95:20 Income Ratio less than or equal to 10.3 | 0.082 | 0.007 | 0.047 | 0.001 |
| 95:20 Income Ratio between 10.3 and 13 | 0.072 | 0.007 | 0.047 | 0.001 |
| 95:20 Income Ratio greater than or equal to 13 | 0.067 | 0.010 | 0.034 | 0.002 |

Notes:

1. Sample includes migrants.
2. Return to migration is calculated by each migrant's post-migration wage minus his/her estimated counterfactual pre-migration wage. The average of the individual returns are presented in the table.
3. Standard errors are calculated from 1000 bootstrap repetitions.
4. Income ratios used in the analysis are calculated by Bernstein et al. using 2001-2003 Current Population Survey (Bernstein, 2006).