Labor Migration and Wage Growth in Malaysia

Rebecca Lessem[*]

October 4, 2011

Abstract

I estimate a discrete choice dynamic programming model to calculate how wage differentials affected internal migration decisions in Malaysia between 1978 and 1988. In the model, individuals pick a location at each point in time, thereby allowing for repeat and return migration. I calculate total income in a location as wages plus in-kind payments. I find evidence that wages motivate migration decisions; however, I do not find evidence that in-kind payments play a role. As a person's wage decreases, his likelihood of migration increases. People move from low to high wage locations, and people with a low wage draw in their current location are more likely to move. People prefer to live in their home location. In Malaysia at this time, there were significant urban-rural and regional earnings disparities. If people move to take advantage of higher earnings, there will be substantial wage growth through migration. I find that migration increases earnings over the course of a lifetime by about 10%.

1. Introduction

This paper studies the determinants of labor migration in Malaysia between 1978 and 1988 to understand how migration contributes to wage growth. Malaysia at this time is an interesting case for several reasons. In this time period, Malaysia's economy was growing fairly rapidly, as GDP growth averaged 7% a year in the 1980's.[1] Due to this growth and to government intervention, new urban areas were growing rapidly. There were large urban-rural and regional earnings disparities, creating incentives to move.

To understand economic behavior in Malaysia at this time, it is important to account for the ethnic differences in the country . Approximately 60% of the country is ethnically Malay (called *bumiputra*), while the remainder is predominantly Chinese and Indian. There historically were large economic disparities between the groups, which still remain but have lessened somewhat. Most notably, ethnic Malays, who make up a majority of the population and live predominantly in rural areas, were significantly poorer than the other ethnic groups.

The New Economic Policy (NEP), implemented starting in 1971, was a series of policies aiming to eliminate these ethnically-driven economic disparities. Since Malays were concentrated in rural areas, policies to increase the wealth of the *bumiputra* affected rural-urban disparities and therefore migration. The government aimed to increase *bumiputra* income. One way of doing this was to increase productivity in rural areas. This would decrease rural to urban migration by increasing rural incomes. The government also encouraged rural to urban migration so that the *bumiputra* could take advantage of the higher wages in cities.

Data from the Malaysia Family Life Survey (MFLS), conducted in 1988, shows the frequency of migration in this time period. As of 1988, around 65% of this sample had moved to a new location within Malaysia at least once in their lifetime. Much of this migration was from rural to urban locations. At birth, 1.5% of the sample lived in a city, whereas at the time of the survey, at which point the average age of respondents is close to 40, around 5.5% of the sample lived in a city. Accompanying these high migration rates are high rates of repeat and return migration, showing that migration decisions are not permanent. The average migrant moved 2.4 times from 1978 to 1988. To explain this behavior, we need a model that allows for multiple decisions over the course of a lifetime.

In this paper, I estimate a model of internal migration in Malaysia. In the model, each person picks a location at each point of time. Each individual knows the income distribution in all locations but does not know his exact earnings in a location before moving there. Total income includes monetary wages plus in-kind payments, specifically in food and housing. In-kind payments can be an important component of an

---

[1]Calculated using data from the World Bank's World Development Indicators.

individual's total income, as if they do not have to pay for food or housing their expenses drop significantly. When moving to a new location, a person does not know if he will receive in-kind payments there; the probability of this depends on a person's characteristics. There is also a probability of unemployment in each location, which also depends on his characteristics. There is a bias in preferences so that a person's utility increases if he is living at his home location. This is used to account for the high observed rates of return migration. Utility also depends on the ethnic composition of a location, allowing for the possibility that people prefer to live in places where a greater share of the population is of the same race. There is a cost of moving between locations, which includes a fixed cost and also depends on the distance between locations. I also allow the moving cost to vary with race. This accounts for different rates of migration in the different ethnic groups, which could potentially be due to government policies at the time.

I use data from the MFLS to estimate this model. In a first step, wage regressions are used to calculate the wage distribution for each individual at each location at each point in time. I use a probit model to calculate a person's likelihood of being unemployed and of receiving in-kind payments in new locations. I then estimate the utility and cost parameters of the model using maximum likelihood.

I find that both wages affect migration decisions. Wages affect migration in two ways. As a person's wage increases, his likelihood of migration decreases. In addition, people move from low to high wage locations. However, I do not find evidence that in-kind payments affect migration decisions. People prefer to live at their home location.

The paper is organized as follows. Section 2 reviews the literature. Section 3 develops a model of migration. Section 4 explains the data in the MFLS and provides some descriptive statistics. Section 5 describes the empirical analysis, and the results from the structural estimation are reported in section 6. Section 7 concludes.

2. Literature Review

Most of the past literature on migration analyzes migration as a one-time decision, and therefore cannot explain the fact that many people move multiple times. Pessino (1991) studies migration in a dynamic setting and therefore can explain repeat and return migration. She develops a migration model with uncertainty about an individual's wages, where people have beliefs on the value of the mean wage. When a person migrates, he observes an a wage offer, updates his beliefs on the mean wage, and then decides to stay, return home, or go to a new location. Therefore, the model can predict repeat and return migration. Pessino tested the model with data from Peru from 1985-1986, and found that the predictions of the human capital model apply only to

primary movers.

Kennan and Walker (2011) estimate a model where agents move based on differences in expected income. Individuals decide where to live in each period; therefore, people can make multiple moves. They tested their model using NLSY data on white males with a high school education and found that state-to-state migration was significantly affected by expected income. People will move to a new location when earning a low wage in their current location. In this paper, I use a similar structure to understand migration behavior in Malaysia. I expect different migration behavior due to the fact that Malaysia is a much poorer country. I expect moving costs to play a larger role in Malaysia because it will be harder for people to finance a move due to higher poverty and a less-developed credit market.

Gemici (2011) compares the migration behavior of married couples and singles. She develops a dynamic model of household migration with bargaining between the husband and wife. Individuals face uncertainty about future earnings and about the likelihood of divorce. The model is tested using PSID data. She finds that migration of married couples occurs much more in response to the earnings of men than to the earnings of women, as women have lower wage offers, a lower arrival rate of offers, and a lower variance in offers. Gemici (2011) shows that migration incentives are different for men and women, and that men are more likely to respond to wage differentials. For this reason, in this paper I focus only on the migration behavior of men.

3. Model

The model in this paper follows the framework of Kennan and Walker (2011). Each person picks a location at each point in time. Agents know the income distribution in different locations, but they do not know their wage in a location before moving there. There is some probability of unemployment. The utility function allows for a home bias in utility. In addition, a person's preferences over locations depend on the ethnic composition of each location. An individual must pay a moving cost if he wishes to move to a new location. Each period, individuals receive payoff shocks to living in each location; these represent the non-economic value of living in each location. The preference shocks are only observed by the individual, but are known to be drawn independently from an identical distribution in each period and location.

At the beginning of each period, a person observes his payoff shocks over all locations for that period, and then decides whether to stay in his current location or to move to a new location. If he moves to a new location, he then learns the value of his income in that location and lives there for one period. Then, at the start of the next period the same process repeats. I assume a finite horizon, as a person makes location

decisions for each year until he retires.

3.1 Structure of Model

Assume that a person's utility depends on his total earnings (Y), his current location (ℓ), and his characteristics X, so the utility function can be written as $u(Y, \ell, X)$. Earnings include wage payments plus in-kind payments, which are common, particularly in this setting. The vector X contains characteristics of an individual- most importantly, his home location H and his race r. I assume that a person prefers to live at his home location, so his utility will increase if $\ell = H$. I assume that people prefer to live in a location where the fraction of the population of the same race is greater.

Denote η_t as a person's payoff shocks at time t, where $\eta_t = \{\eta_t^j\}_{j=1}^J$. I assume that these are independently drawn from an extreme value Type I distribution.

At the start of period t, a person is living in location ℓ_{t-1}. He knows his earnings $Y_{\ell_{t-1}}$ in this location. He chooses location ℓ_t for period t. He realizes his payoff shocks (η_t) at the beginning of the period, and then decides where to live in this period. Denote $V_t(Y_{\ell_{t-1}}, \ell_{t-1}, \eta_t, X_t)$ as the value function at time t. It depends on his income in his current location $Y_{\ell_{t-1}}$, his current location ℓ_{t-1}, his payoff shocks η_t, and his characteristics X_t. Denote β as the discount factor and denote $E_t V_{t+1}(Y_j, j, \eta_{t+1}, X_t)$ as the expected value of living in location j in period $t+1$. Then

$$V_t(Y_{\ell_{t-1}}, \ell_{t-1}, \eta_t, X_t) = \max_j v(j, Y_{\ell_{t-1}}, \ell_{t-1}, X_t) + \eta_t^j. \tag{1}$$

The value function in equation (1) indicates that a person picks the location that has the highest value, where the value of living in a location has a deterministic and a random component. The deterministic component is given by:

$$
v(j, Y_{\ell_{t-1}}, \ell_{t-1}, X_t)
$$
$$
= \begin{cases} u(Y_j, j, X_t) + \beta E_t V_{t+1}(Y_j, j, \eta_{t+1}, X_t) & \text{for } j = \ell_{t-1} \\ E u(Y_j, j, X_t) - c(\ell_{t-1}, j, X_t) + \beta E_t V_{t+1}(Y_j, j, \eta_{t+1}, X_t) & \text{for } j \neq \ell_{t-1} \end{cases}
$$

If a person does not move, he knows his utility for the current period because his earnings are known. If he moves, he only knows his expected utility because he only has an expectation over his income in the new location. If he moves, he also must pay a moving cost, where the cost of moving from location x to location y is denoted as $c(x, y, X)$. The cost of moving depends on an individual's characteristics (X). This is because I allow ethnicity and education to affect the fixed cost of moving. An individual only knows the expected value of living in each location in the future periods, as the payoff shocks and future wages in new locations are not known. Therefore to

calculate the expected values in future periods, I integrate over both the income and payoff shock distributions.

Following McFadden (1973) and Rust (1987), the agent can solve for the expected value when the uncertainty is over future realizations of the preference shocks (η). If θ_j is known and η_j is distributed with a Type I extreme value distribution, then

$$\mathrm{E} \max_j \left(\theta_j + \eta_j \right) = \log \left(\sum_j \exp(\theta_j) \right) + \gamma, \tag{2}$$

where γ is Euler's constant ($\gamma = 0.58$).

In this case, the agent needs to take an expectation over future values of the payoff shocks and wages, which I assume are drawn from independent distributions. First calculate the expected value of living in location ℓ, conditional on earnings there Y_ℓ. I do this using equation (2).

$$\mathrm{E}_t \left[V_{t+1}(Y_\ell, \ell, \eta_{t+1}, X_t | Y_\ell) \right] = \log \left(\sum_j \exp \left(v(j, Y_\ell, \ell, X_t) \right) \right) + \gamma. \tag{3}$$

Then the expected value of living in location ℓ is the integral over the wage distribution of equation (3):

$$\mathrm{E}_t \left[V_{t+1}(Y_\ell, \ell, \eta_{t+1}, X_t) \right] = \int \left(\log \left(\sum_j \exp \left(v(j, Y_\ell, \ell, X_t) \right) \right) + \gamma \right) f_\ell(Y_\ell) dY_\ell, \tag{4}$$

where $f_\ell(Y_\ell)$ is the wage distribution in location ℓ. Then I can solve for the expected values using backwards induction, assuming a finite horizon. This allows the agent, once he sees his payoff shocks in a period, to make a migration decision.

If the values of the payoff shocks were observed, I could determine where a person will live in each period. However, since the values of the payoff shocks are unknown, I instead calculate the probability that a person will move to a given location in each period.

3.2 Migration Probabilities

I use properties of the extreme value distribution to calculate migration probabilities in each period. Suppose a person is choosing j as to maximize $(\theta_j + \eta_j)$, where θ_j is known and η_j is distributed with an extreme value Type I distribution. Then the probability that a person chooses choice q (denoted by $\mathrm{P}(q)$) is given by

$$\mathrm{P}(q) = \frac{\exp(\theta_q)}{\sum_j \exp(\theta_j)}. \tag{5}$$

Define $P_t(\ell_t|\ell_{t-1}, Y_{\ell_{t-1}}, X_t)$ as the probability that a person chooses location ℓ_t in period 1, when living in ℓ_{t-1} and earning $Y_{\ell_{t-1}}$ at the start of the period. Using equation (5),

$$P_t(\ell_t|\ell_{t-1}, Y_{\ell_{t-1}}, X_t) = \frac{\exp(v(\ell_t, Y_{\ell_{t-1}}, \ell_{t-1}, X_t))}{\sum_j \exp(v(j, Y_{\ell_{t-1}}, \ell_{t-1}, , X_t))} \tag{6}$$

Given a person's location and wage history, I can calculate the probability that he chooses each location in each period. This will be used to develop the likelihood function in section 5.6.

4. Data and Preliminary Evidence

I use data from the Malaysia Family Life Survey (MFLS) to estimate the model explained above. The MFLS was a panel survey conducted by RAND in 1976 and 1988. In this paper, I just use the data from the 1988 round. The survey is at the individual level. The focus was on family structure, fertility, economic status, education and training, transfers, and migration. The respondents were selected from 52 communities selected to be representative of West Malaysia[2].

Survey respondents were asked to list each place they had lived in the past and when they lived at each of those places. They also reported their job and wage histories. I used this information to construct a database with a person's wage and location in each year.

I study migration over a 10 year period, from 1978 to 1988. A location is defined as a state and the urban or rural classification of the village, town, or city in which a person lives.[3] Because the dataset is only representative of West Malaysia, I only consider location choices in West Malaysia. It does not seem that East to West migration is common, which can be due to cultural differences and also government restrictions on immigration.[4] There are 12 states in West Malaysia, so there are 24 possible location choices. The home location is the place a person lived at age 15.

As explained earlier, Gemici (2011) shows that migration incentives are different for men and woman. She finds that men are much more likely to move in response to

---

[2]Malaysia is geographically split into West and East Malaysia. Currently 80% of the population lives in West Malaysia.

[3]One problem with this is that a person could move between cities in the same state. In this analysis, this would not be counted as migration. However, the data does not allow me to study such migration as it only reports a person's state and whether he lives in a kampung, estate, land scheme, new village, small town, large town, or city. I define an urban area as a large town or a city and a rural area as everything else.

[4]In the data, only individuals living in West Malaysia in 1988 were surveyed. Only a small percentage of them had lived in East Malaysia at some point in the past. These observations were dropped.

wages, and women are more likely to move for non-economic reasons (for example, a woman may move due to a job offer for her husband). For this reason, in this paper I only study the migration behavior of men. To study behavior with race, I only include respondents from the three main ethnic groups (Malay, Chinese, Indian).

Table 1 shows summary statistics on the sample, divided by race. The majority of the sample (around 70%) is ethnically Malay, but there is a large percentage of Chinese and Indians. I show the percent of each group that has completed each level of education. The average age in the sample is around 35.

Table 2 compares the migration behavior of different ethnic groups. The first column shows the percent of the men in each group that move between 1978 and 1988. The second column shows the average number of moves made by migrants in each group (excluding those who do not move and therefore have zero moves). About 38% of the sample migrated at least once over the time period. Malays are the most likely to have moved. Many of the migrants move more than once, as the average number of moves by migrants is 2.4.

To more clearly see what factors affect migration, I ran a probit regression estimating the likelihood that a person migrates at least once between 1978 and 1988.[5] I allow the migration probability to depend on a person's wage relative to his expected wage at the start of the period. [6] The intuition is that people with a lower wage draw in their initial location should be more likely to migrate to get a new wage draw. I control for whether or not a person is located at his home location in 1978. If people prefer to live at their home location, a person will be less likely to move if he is at home. I control for the respondent's age as we expect the likelihood of migration to decline with age. This is because younger people have more time to enjoy the benefits of higher wages. I also control for a person's race and education level. I control for whether or not a person is living in an urban area in 1978. I also include state fixed effects.

The results of the regression are shown in Table 3. I find no evidence that wages affect migration decisions using this specification. People who are at their home location in 1978 are less likely to migrate, showing a preference for remaining at home. More educated people are more likely to migrate. This could be because people with more education have more access information about job opportunities in other locations. Malays are more likely to migrate than Indians, but there is no difference between the migration rates of Chinese and Indians. This suggests that there is some difference unique to Malays, indicating that the government policy may have played a role. Older people are less likely to migrate, as expected because they have more time periods to benefit from higher wages. People who live in urban areas are more likely to migrate. This indicates that urban to urban migration may be more common

---

[5]In this regression, a migration is defined the same as it will be in the structural estimation. A migration is when a person moves from a rural to urban area in the same state or to a new state.

[6]The expected wages are shown in section 5.1.

8

than rural-urban migration.

I use this reduced form specification to calculate how ethnicity and whether or not a person is at his home location affect migration probabilities. I do this for a Malay male who is 35 years old (the average age in the sample), attended lower secondary school, and is earning exactly his expected wage. People living at their home location are less likely to migrate. The migration probability increases by about 8 to 16 percentage points when a person is not living at his home location.[7] This shows that living away from home has a large effect on the likelihood of migration.

I performed the same exercise to see how migration probabilities differ by ethnicity. I find that the migration probability for a Malay is about 2 to 7 percentage points higher than that of a non-Malay, holding all other factors constant.[8] This shows that the differences in migration rates across ethnic groups is significant.

5. Estimation

I estimate the model in a two-step process. First, I calculate the income distribution, which consists of wages, in-kind payments, and the probability of unemployment. Then, taking the income distribution as given, I use maximum likelihood to estimate the utility and cost parameters.

5.1 Wages

In order to calculate a person's expected wage at each time and location, I calculate the wage distribution. Table 4 shows some summary statistics on 1988 wages, divided by education and race. Table 4 shows monthly earnings, in Malaysian Ringgits. As education increases, earnings increase. Furthermore, Chinese earn more than ethnic Malays or Indians, except for the group with the most education, where the earnings of Malays and Chinese are approximately equal.

Figure 1 shows how wages vary across locations. There is significant variation across states. These wage differentials provide an incentive for migration.

To estimate the model, I need to know the wage distribution for each person and location. This is necessary to calculate a person's expected utility in new locations. I

---

[7]This is for a Malay who attended lower secondary school who earns the mean wage in a location for a person with his characteristics. The increase in probability depends on his initial location (urban or rural and the state). In the extreme cases, the migration probability increases by 8 percentage points for a person in a rural location in the state of Pahang. The migration probability increases by 16 percentage points for a person in a rural location in W. Persekutuan.

[8]In the extreme cases, the migration probability increases by 2 percentage points for a person living in his home location in an rural area of Pahang. The migration probability increase by about 7 percentage points for a person who is living away from his home location in a urban location in Pulau Pinang.

write the wage for person i living in location j at time t as follows:

$$w_{ijt} = X_{it}\Omega + \mu_j i + \Gamma_t + \epsilon_{ijt} \tag{7}$$

The vector X_{it} represents the characteristics of person i at time t, such as his age, education, race, and work experience.[9][10] The term μ_j is a location fixed effect. The expected wages in each location vary because of the state fixed effects. I control for time fixed effects with Γ_t. The term ϵ_{ijt} is a match parameter, which is drawn for each person and location pair. A person only realizes his value of the match parameter in a given location if he moves there and gets a wage draw. Migration decisions are affected by μ_j and realized values of ϵ_{ijt}, as the other terms do not vary across locations. A person knows the value of the state fixed effects in each location so he should be more likely to move to places with higher wages. However, he does not know the value of the match parameter in a location unless he is currently living there. This creates a degree of uncertainty for an individual when moving.

I estimated equation (7), and the results are shown in Table 6.[11] Wages increase with experience and education. People living in urban areas earn more. This is one component of how wages can be affected by migration. In addition, the state fixed effects (not shown) vary significantly, showing that people can earn higher wages by moving to cities or to new states. The values of the state fixed effects vary, ranging from close to -49 in Melaka to around 46 in Kelantan. Table 5 shows that a Malay who attends lower secondary school has a wage of approximately 256. Therefore, the state fixed effects can significantly affect wage levels, as the difference between the highest and lowest state fixed effects is equal to around 37% of that wage. Chinese people earn the most, followed by Indians and then Malays.

I used the results of this regression to impute an expected wage for each person in each location and year. To estimate the model, I also need to calculate the distribution of the match parameter. The match parameter is the difference between a person's expected wage and his actual wage, which is the error term in the regression. I use the error terms to approximate the distribution of the match parameter, which I assume to be distributed normally. For computational simplicity, I assume that the error term can take on one of 3 discrete values, which are calculated using a discrete approximation of the normal distribution.

---

[9]There is no exact information in the data on work experience, so I just use age minus 17, assuming people start work at age 17 and are always employed.

[10]I also specified this regression to allow for experience in a given location to affect wages by controlling for the time living in each location. The coefficient on this variable was small and insignificant and is therefore not included in the analysis.

[11]Because there were too few observations to identify location and time fixed effects, I combined the years 1978 and 1979. Thus there are year fixed effects for every year from 1980 and 1988 and a fixed effect for 1978 and 1979 combined.

5.2 In-Kind Payments

Total earnings also depend on in-kind payments. Data is available on three types of in-kind payments: meals, housing, and food for their own use. Table 5 shows the percent of the population that receives in-kind payments in rural and urban areas of each state, respectively. These numbers show that in-kind payments are common and therefore should be included as part of the value of employment.

To calculate expected income, and therefore expected utility, in a location, I need to know the probability of receiving each kind of in-kind payment. I estimate the probability of receiving each type of in-kind payment using a probit model. I let the probability of receiving each kind of payment depend on location, race, education, sex, and year. The results of these regressions are reported in Table 7.

The results show variations across locations, race, and education. [12] The in-kind payments provide another source of variation across ethnic groups. This could allow for different migration behavior between people of different races. Overall, expectations over and current realizations of in-kind payments can affect migration behavior. If in-kind payments are valuable, a person may move to a place where there is a high probability of receiving in-kind payments. It seems that in-kind payments are more likely in rural areas, which could compensate for some of the rural-urban wage gap.

When estimating the model, I assume that a person only knows the probability that he will receive each type of in-kind payment in new locations. I use the results in this section to calculate these probabilities.

5.3 Unemployment

When a person moves to a location, there is a chance that he will be unemployed there, as I do not allow for individuals to receive wage offers before moving to a new place. Therefore, the probability of unemployment affects expected income, expected utility, and migration decisions. Unemployment rates vary across states, which should also affect migration decisions. Figure 2 shows the percent of the sample living in each location that is unemployed in each period.

I allow the probability of unemployment to depend on a person's race, gender, whether or not they are in a rural or urban location, whether or not they are a migrant, and education.[13] I expect higher unemployment for migrants. I also included state

---

[12]I also tested whether wages had an effect on in-kind payment realizations. Wages had a positive and significant effect n all 3 types of in-kind payments, but the magnitude was so small that it only causes small shifts in the imputed probability of receiving each type of in-kind payment. In addition, dropping the wage term did not have a noticeable difference on the other estimates in these regressions. For these reasons, for simplicity I dropped the wage from these regressions.

[13]Migrants are defined as those who did not live in that location in the previous period.

and time fixed effects. I estimate the probability of unemployment using a probit specification.

The results of this regression are shown in Table 8. Chinese are more likely to be unemployed than Indians or Malays. Education affects unemployment rates, although in no clear pattern. As expected, migrants are more likely to be unemployed. This again can affect the value of moving to new locations. A place with high unemployment (captured by the location fixed effects) will be less attractive even if there are high wages.

When estimating the model, I assume that a person only knows the probability that he will be unemployed in new locations. I use the results in this section to calculate these probabilities.

5.4 Utility Function

Now that I have identified the wage distribution and the probability of receiving each type of in-kind payment, I can define total income. Total income is a combination of wage payments and in-kind payments:

$$Y_{ijt} = w_{ijt} + \lambda_1 ik_1 + \lambda_2 ik_2 + \lambda_3 ik_3 , \tag{8}$$

where ik_1, ik_2, and ik_3 are dummy variables that equal 1 if a person gets that type of in-kind payment and 0 otherwise. The parameters λ_1, λ_2, and λ_3 are parameters to transform the value of the in-kind payments into dollars, and they will be estimated in the model. For each in-kind payment k, the earnings increase by λ_k if $ik_k = 1$.

In new locations, a person will not know his income and will only know his expected income. This is defined as

$$EY_{i\ell t} = (1 - \chi_u(\ell, X)) \Big(Ew_\ell + \chi_1(\ell, X)\lambda_1 + \chi_2(\ell, X)\lambda_2 + \chi_3(\ell, X)\lambda_3 \Big) \tag{9}$$

There is some probability of unemployment. If a person is employed, he knows his expected wage. Then, there is also some probability that he receives each type of in-kind payment.

I assume a linear utility function. For a person with home location H, living in location ℓ where he earns Y_ℓ, I write the utility function as

$$u(Y_\ell, \ell, X) = \alpha_0 Y_\ell + \alpha_H \mathcal{I}(\ell = H) + \alpha_r r_{\ell q} \mathcal{I}(\text{race} = q) \tag{10}$$

In equation (10), α_0, α_H, and α_r are parameters, and $\mathcal{I}(\ell = H)$ and $\mathcal{I}(\text{race} = q)$ are indicator functions that equal 1 if the expression in parentheses is true and 0 otherwise. The term r_{jq} is the fraction of the population in location ℓ of race q. The first term in the utility function represents the effect of income on utility. The second term

12

is used to account for the preference for living at one's home location. If a person is living at his home location, his utility will increase by the amount α_H. I define the home location as the state that a person is living in at age 15. I allow for a preference for living at a location with more people of a person's race. The term $r_{\ell q}$ gives the percent of people of race q in location ℓ. Data from 1988 on the percent of each state of each ethnicity is shown in Figure 3.[14] This shows wide variation across states. I use data from 1978-1988 in estimating the model to allow for changes over time, which are small.[15]

Substituting in for Y_ℓ in equation (10) using equation (8),

$$u(Y_\ell, \ell, X) = \alpha_0 w_\ell + \alpha_0 \lambda_1 ik_1 + \alpha_0 \lambda_2 ik_2 + \alpha_0 \lambda_3 k_3 + \alpha_H \mathcal{I}\{\ell = H\} + \alpha_r r_{\ell q} \mathcal{I}\{\text{race} = q\}$$

(11)

I now calculate expected utility. When people move to new locations, they do not know their actual utility because they do not know their earnings. There is uncertainty over whether or not they will be employed, their wage draw, and whether or not they will receive each type of in-kind payment. They have expectations over these factors and therefore we can define expected utility as follows.

$$Eu(Y_\ell, \ell, X) \quad = \quad \alpha_0 EY_\ell + \alpha_H \mathcal{I}\{\ell = H\} + \alpha_r r_j \mathcal{I}\{\text{race} = r\}$$

Substituting in for expected income using equation (9)

$$\begin{aligned} Eu(Y_\ell, \ell, X) \quad = \quad & (1 - \chi_u(\ell, X))\alpha_0 \Big(Ew_\ell + \chi_1(\ell, X)\lambda_1 + \chi_2(\ell, X)\lambda_2 + \chi_3(\ell, X)\lambda_3 \Big) \\ & + \quad \alpha_H \mathcal{I}\{\ell = H\} + \alpha_r r_{\ell q} \mathcal{I}\{\text{race} = q\} \end{aligned}$$

If a person is unemployed, then he receives zero earnings. If he is employed, then he expects to earn the expected wage. Then there is also some probability that he gets each type of in-kind payment. The expected utility increases by α_H if a person is living in his home location. It increases by α_r times the fraction of the state that is of the same ethnicity as the individual.

5.5 Moving Costs

I assume that the cost of moving equals some fixed cost plus a variable cost, which is linear in the distance between locations. The cost of moving from location x to location y is given by $c(x, y, X)$, where

---

[14]This data was provided by the Malaysian Department of Statistics.

[15]I do not have data for 1978-1979 for WP Kuala Lumpur, so I use the 1980 values for those years. For projections beyond 1988 in the model, I assume that individuals expect these values to remain the same in the future.

$$c(x, y, X) = c_0 + c_1 Chinese + c_2 Indian + c_3 d(x, y) + \sum_{e=2}^{5} \delta_e education_e,$$

The constant c_0 is the fixed cost of moving that is paid by all people who decide to move. The variables *Chinese* and *Indian* are dummy variables that equal 1 if a person is of that ethnicity. This allows for the fixed cost of moving to vary with race. I test whether the moving cost is lower for Malays than the other groups because of government policies. However, since I observe differences in the data across all three ethnic groups, I allow for specific fixed costs for each group. Since I see different migration behavior across education levels, I allow the fixed cost of moving to vary with education. The sample is split into 5 education groups. I estimate an extra fixed cost for all groups except those with no education, setting the fixed cost for that group at 0. The variables $education_e$ are dummy variables that equal 1 if a person is in education group e. The distance between two locations x and y is denoted as $d(x, y)$.[16] I assume that the cost of moving varies linearly with the distance between locations.

5.6 Likelihood Function

I estimate the utility and cost parameters using maximum likelihood. Denote the parameters to be estimated as θ, where

$$\theta = \{\alpha_0, \alpha_H, \alpha_r, \lambda_1, \lambda_2, \lambda_3, c_0, c_1, c_2, c_3, \delta_2, \delta_3, \delta_4, \delta_5\}$$

.

The likelihood function is calculated by using the probability that a person picks the location that he lives in each period, conditional on his current location and earnings. I write the log-likelihood function as follows (assuming a sample size N):

$$L(\theta) = \sum_{i}^{N} \sum_{t=1978}^{1988} \left[\log \left(P_t(\ell_{it} | Y_{i,t-1}, \ell_{i,t-1}, X_{it}) \right) \right] \tag{12}$$

In equation (12), the probability at each time and for each person is given by equation (6).

6. Results

I maximized the likelihood function in equation (12) with respect to the parameters θ, assuming a discount rate $\beta = 0.95$. I assume that each person works up to age 65.[17]

---

[16]The distance is defined as the distance between state capitals. If a person moves between a urban and rural area in the same state, the distance is defined as 0 and the only cost paid is the fixed cost.

[17]I assume perfect expectations about changes in the income distribution from 1978 to 1988. However, since the data ends in 1988, there is no income data past that year. Therefore when calculating expected

Table 9 shows the utility and cost parameter estimates.[18]

The wage parameter is positive and significant, indicating that wages affect migration decisions. This means that people who are earning lower wages are more likely to migrate and that people move from low to high wage locations. The results on in-kind payments are a bit confusing, as I find a negative effect of being provided meals. People prefer to live in locations where a larger share of the population is of the same race.

Moving costs are also significant. I find that the cost of moving for Malays or Indians is less than that of Chinese. In addition, age increases the moving cost, making younger people more likely to immigrate. The human capital model of migration implies that younger people are more likely to move because they have more time to earn higher wages. However, after accounting for this in the model, I find that ages still affect moving costs, implying that there is some other factor that affects both immigration and age.

To test the fit of the model, I compared predicted and actual migration probabilities. I calculate the probability that a person moves in each period, splitting the sample by whether or not they are at their home location at the start of the period, age, and education. Overall, the model fits the data fairly well. The biggest problems are for the group with no education. However, the sample size here is quite small.[19]

The model predicted that an increase in wages will decrease the likelihood that a person migrates, and that people who are living at their home location will be less likely to move. So far, the parameter estimates demonstrate qualitative support for these predictions. I can use the model to calculate how each of these factors affects migration. Also, I can use the model to see how migration varies with race.

I simulated migration probabilities to see how wages, location, and ethnicity quantitatively affect migration behavior. I calculated migration probabilities for a 35 year old male with the average level of education. I assume that he is living in urban areas in the state of Kelantan, and vary whether or not this is his home location. According to the model, people who earn wages above their expected wages should be less likely to move. Thus I simulated migration probabilities for 3 values of the realized wage in the original location (low, medium, and high). For each of these situations, I calculated the probability that a person moves away from his original location at least once over the 10 year period. I vary whether or not the initial location is a person's home location in order to test the effects of the home bias. I also vary the ethnicity of the individual to see how behavior varies with race. These results are shown in Table

---

values in the future past 1988 agents assume that the income distribution is the same as in 1988.

[18]The coefficients on the in-kind parameters are the total effect of each in-kind payment, i.e. $\alpha_0 \lambda_k$ for each in-kind payment k.

[19]In future work, I will combine the group with no education and those with just primary school for the education fixed effects in the moving costs. This should help to identify these costs better.

11. Each entry shows the probability that a person will move at least once over the 10 year period from 1978 to 1988.

In Table 11, I see that if a person gets a positive wage differential (a high wage draw), the probability of migration decreases. As expected, due to the home bias in the utility function, a person who is not living at his home location is much more likely to migrate than a person living at his home location. This is shown to be quite important as it drastically effects migration probabilities. Due to the differential moving costs, Malays are more likely to move than Chinese.

6.1 Wage Growth from Migration

In this section I focus on how migration leads to wage growth. I use the parameter estimates from the model to see how wage growth is affected by migration. I simulate behavior of individuals in the sample to see the average wage growth.

I simulate the behavior of each individual over this time period and calculate their earnings in each period according to the simulated decisions. Because wages do not change, if there was no migration, a person's wage would only change due to returns to increased experience.[20]

Figure 4 shows wages over a lifetime in two scenarios: when they can and cannot move between locations. This shows that wages are higher when people are allowed to move, as they move from low to high wage locations and also move when the realized wage draw in the current location is low. Most of the increased earnings due to migration come at younger ages, which is driven by the fact that younger people are more likely to move. Overall, this simulation shows that lifetime wages are reduced by about 10% when people are not allowed to migrate.

7. Conclusion

In this paper, I estimated a model to explain migration trends in Malaysia between 1978 and 1988. In this period, close to 40% of the sample migrated at least once, and repeat and return migration were common.

I estimate a discrete choice dynamic programming model of migration to explain these trends. People receive payoff shocks to living in each location in each period. They know the wage distribution in each location but do not realize their wage draw in a location without moving there. There is a bias in preferences to account for the empirical fact that people prefer to live at their home location. There is also a cost of moving between locations. The model predicts that if a person's wage decreases,

---

[20]This is driven by the assumption that a person cannot get a wage draw if they stay in their current location.

then he will be more likely to migrate. People living at their home location will be less likely to migrate.

I estimated the parameters of the model using data from the Malaysia Family Life Survey. The estimates show that wages affect migration decisions. A strong finding from the estimation is that people prefer to live at their home location. This is shown to have a significant effect on migration, as people who are not living at home are more likely to migrate. Government policy at this time made it easier for Malays to migrate, so we allowed the fixed cost of moving to vary with race. I found that the moving costs for Malays are lower than those of Chinese.

I use the estimated model to see how migration contributes to wage growth over a lifetime. I find that migration leads to a 10% increase in lifetime earnings. Even though people are likely to move, they also have a strong tendency to return to their home location within a short period of time. Therefore they only earn higher wages for a short period of time.

References

GEMICI, A. (2011): "Family Migration and Labor Market Outcomes," Working Paper.

KENNAN, J., AND J. WALKER (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79(1), 211–251.

MCFADDEN, D. (1973): *Frontiers in Econometrics*chap. Conditional Logit Analysis of Qualitative Choice Behavior. Academic Press.

PESSINO, C. (1991): "Sequential Migration Theory and Evidence from Peru," *Journal of Development Economics*, 36, 55–87.

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55 (5), 999–1033.

Table 1: Characteristics of Sample

Education Group	Malay	Chinese	Indian
No Education	3.4	1.9	0.7%
Primary School	38.0	33.0	37.5%
Lower Secondary School	16.2	15.1	27.6%
Secondary School	28.1	25.3	25.7
Post Secondary School	14.3	20.8	8.6
Average age	34.8	35.7	36.0
Number of observations	587	106	152

Table 2: Migration and Race, 1978-1988

Race	Percent that Move	Average Number of Moves*
Malay	39.9%	2.4
Chinese	34.9%	2.2
Indian	32.2%	2.5
Total	37.6%	2.4

*Taking the average over the migrant population.

Table 3: Probability of Moving, 1978-1988

Variable	Coefficient Estimate
Wage residual	0.00001
	(0.0001)
Home	-0.41***
	(0.11)
Primary School	0.65
	(0.40)
Lower Secondary School	0.74*
	(0.41)
Secondary School	0.94**
	(0.41)
Post-Secondary School	1.41***
	(0.41)
Malay	0.17
	(0.13)
Chinese	-0.03
	(0.18)
Age	-0.07*
	(0.04)
Age-squared	0.0005
	(0.0005)
Urban	0.40***
	(0.12)
Constant	0.32
	(0.80)
State Fixed Effects	yes
Pseudo R-squared	0.21

Standard errors in parentheses

*** .01 significance, ** .05 significance, * .10 significance

Table 4: Average Wages

	Malay	Chinese	Indian
No education	193.7	366.8	272.0
Primary School	210.2	386.1	281.1
Lower Secondary School	256.0	435.3	257.1
Secondary School	269.1	406.0	281.6
Post-secondary school	447.1	443.1	380.9

Wages given as monthly earnings, in Malaysian dollars adjusted for inflation across years.

Table 5: In-Kind Payments

State	Rural			Urban		
	Meals	Housing	Food	Meals	Housing	Food
Johor	5.0%	174%	13.2%	10.5%	14.8%	13.4%
Kedah	4.9%	5.7%	14.1%	1.5%	1.0%	1.0%
Kelantan	4.1%	12.4%	20.9%	7.9%	21.4%	2.9%
Melaka	6.8%	10.0%	17.4%	10.4%	11.9%	20.9%
N. Sembilan	8.1%	10.4%	6.1%	9.1%	31.2%	3.8%
Pahang	3.3%	21.6%	9.8%	13.2%	15.7%	9.6%
Pulau Pinang & S. Prai	8.8%	5.9%	10.4%	4.6%	18.3%	15.2%
Perak	7.0%	16.3%	7.3%	4.5%	7.1%	6.1%
Perlis	0.66%	0%	23.7%	0%	0%	0%
Selangor	5.3%	9.5%	13.3%	5.0%	9.0%	4.7%
Trengganu	3.9%	9.2%	10.8%	13.4%	14.3%	0%
W. Persekutuan	12.9%	16.8%	10.1%	7.0%	13.7%	6.5%

Gives the percent of the sample living in rural areas of each state that receives each type of in-kind payment.

Table 6: Wage Regressions

	Coefficient Estimate
Experience	2.10**
	(0.93)
Experience-Squared	-0.02
	(0.02)
Primary School	24.7**
	(12.36)
Lower Secondary School	61.4***
	(13.2)
Secondary School	75.2***
	(12.9)
Post Secondary School	214.5***
	(13.8)
Urban	22.9***
	(5.7)
Malay	-21.9***
	(5.8)
Chinese	113.8***
	(8.0)
Constant	185.7***
	(17.9)
State Fixed Effects	yes
Time Fixed Effects	yes
R-squared	0.17

For education, excluded group is those with no education.

Standard errors in parentheses

*** .01 significance, ** .05 significance, * .10 significance

Table 7: Probability of Receiving In-Kind Payments

Variable	Coefficient Estimates		
	Meals	Housing	Food
Urban	0.06	-0.18***	-0.24***
	(0.05)	(0.04)	(0.05)
Primary School	0.81***	0.42***	-0.18**
	(0.23)	(0.13)	(0.09)
Lower Secondary School	0.82***	0.67***	-0.10
	(0.24)	(0.13)	(0.10)
Secondary School	0.74***	0.78***	-0.11
	(0.24)	(0.13)	(0.10)
Post Secondary School	0.43*	0.33**	-0.70***
	(0.24)	(0.13)	(0.11)
Malay	-0.09	-0.24***	0.57***
	(0.06)	(0.04)	(0.05)
Chinese	0.38***	-0.39***	-0.09
	(0.07)	(0.06)	(0.08)
Constant	-2.10	-0.94***	-1.12***
	(0.25)	(0.14)	(0.12)
Year Fixed Effects	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes
Pseudo R-squared	0.05	0.06	0.07

For education, excluded group is those with no education.

Standard errors in parentheses

*** .01 significance, ** .05 significance, * .10 significance

Table 8: Probability of Unemployment

Variable	Coefficient Estimate
Urban	0.14***
	(0.05)
Malay	-0.23***
	(0.05)
Chinese	-0.01
	(0.06)
Migrant	0.30***
	(0.07)
Constant	-1.69***
	(0.14)
Primary School	0.17
	(0.12)
Lower Secondary School	0.01
	(0.12)
Secondary School	-0.18
	(0.12)
Post Secondary School	0.22*
	(0.12)
Constant	-1.69***
	(0.14)
Year Fixed Effects	Yes
State Fixed Effects	Yes
Pseudo R-Squared	0.04

For education, excluded group is those with no education.

Standard errors in parentheses

*** .01 significance, ** .05 significance, * .10 significance

Table 9: Parameter Estimates

Utility Parameters	
Wage Parameter	.0005
	(0.00004)
Home Bias	0.32
	(0.01)
Moving Cost Parameters	
Fixed Cost	2.17
	(0.17)
Chinese	0.20
	(0.05)
Indians	-0.13
	(0.10)
Distance Parameter	0.0022
	(0.00019)
Age	0.086
	(0.006)
Lower Secondary School	0.48
	(0.13)
Secondary School	-0.015
	(0.0062)
Post Secondary School	0.0033
	(0.0016)
Log Likelihood	-2,455.5

Table 10: Model Fit

	Probability of Moving			
	Not at home location at start of period		At home location at start of period	
	Model	Data	Model	Data
Age <30	16.9%	15.1%	5.2%	7.2%
Age 30-40	7.3%	5.5%	2.2%	2.7%
Age 40+	3.3%	3.2%	0.86%	1.4%
No education/Primary School	8.7%	4.7	2.5%	1.6%
Lower Secondary School	8.0%	7.6%	2.3%	4.3%
Secondary School	15.4%	18.4%	4.5%	7.5%
Post Secondary School	19.4%	18.4%	4.7%	8.2%

Table 11: Simulated Probability of Moving between 1978 and 1988

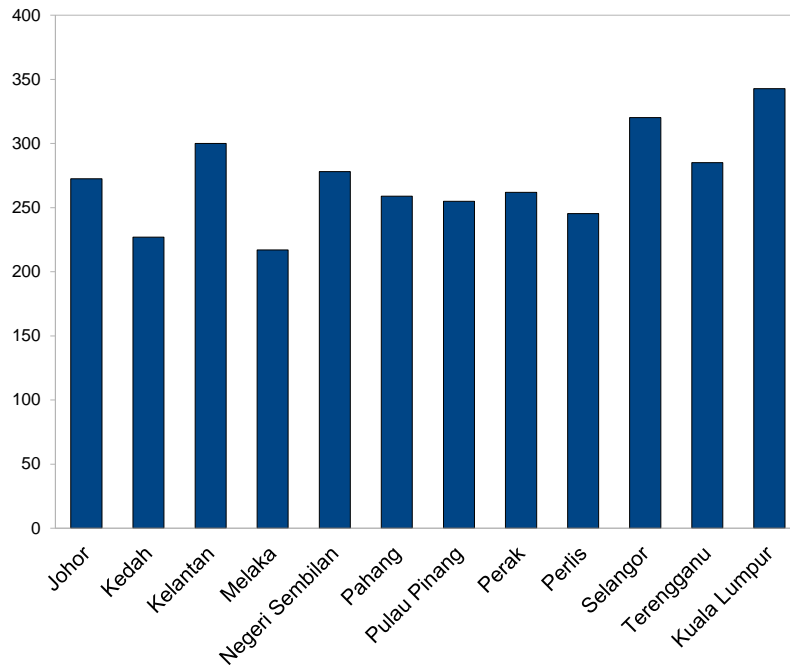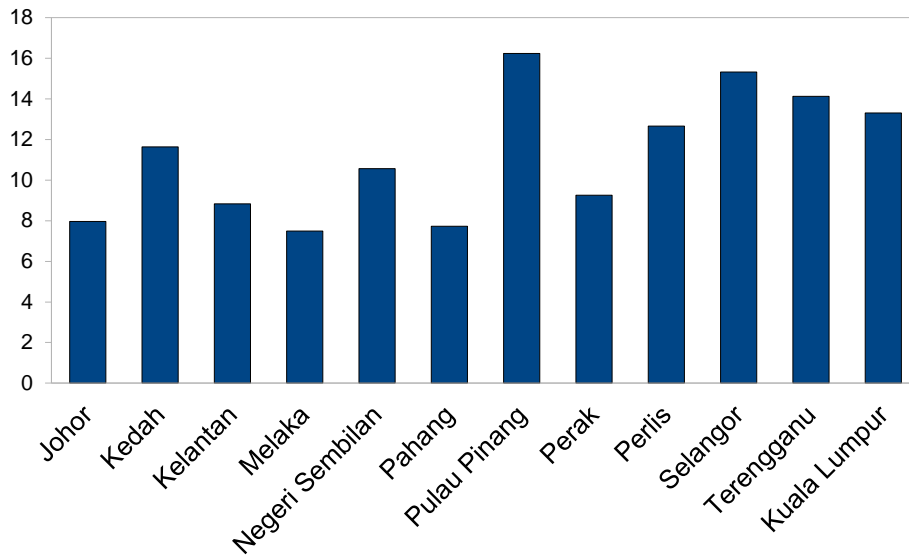| Initial State: Kelantan			
		Malay	Chinese
	Low Wage	21.7%	12.0%
At Home Location in 1978	Medium Wage	19.9%	11.0%
	High Wage	18.3%	10.0%
	Low Wage	57.7%	37.4%
Not At Home Location in 1978	Medium Wage	53.7%	33.9%
	High Wage	50.0%	30.7%

Figure 1: Average Wages

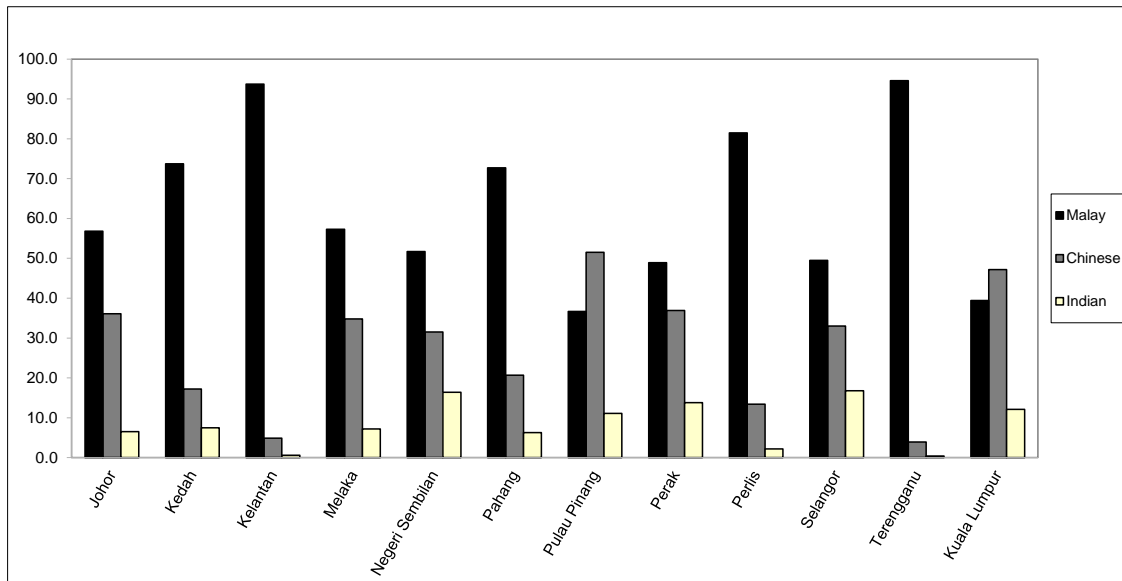Figure 2: Percent Unemployed

Figure 3: Ethnicity by State, 1988

Figure 4: Wage Growth from Migration