Risk Management and Rural to Urban Migration Decisions in Indonesia

Hillary Caruthers*

March 29, 2013

Please do not cite without permission of the author

Abstract: This paper investigates the role of risk in rural to urban migration decisions using Indonesian household-level panel data. Specifically, I use consumption data and measures of household risk aversion to test whether rural to urban migration is a means of managing risk among uninsured households via the diversification of household income flows. Most previous studies of risk and migration do not analyze the migrant's choice of destination but instead focus on the relationship between risk aversion and the likelihood of migration; however, if migration is motivated, in part, by household risk management, then the level of risk aversion should impact both the propensity to migrate and the destination of migration. In this paper I generate predictions regarding the relationship between household risk aversion and the economic riskiness of receiving regions and test these predictions using a multinomial logit estimation. Empirical results generally affirm the predictions of the model. Households prefer to send migrants to destinations with lower consumption variability and, as predicted, this preference is stronger among households with higher risk aversion. Also, all households prefer destinations where average consumption is less correlated with home consumption.

1 Introduction

Rural to urban migration is an important feature of economic growth, especially in developing countries. Historically, per capita income is strongly correlated with the percentage of population living in the urban areas (Figure 1). In most developing countries, the rates of rural to urban migration are substantial (Bell and Muhidin, 2009), although the rate of urban population growth appears to

*University of Wisconsin-Madison, Department of Agricultural and Applied Economics.
Special thanks to Laura Schechter, Jenna Nobles, Jennifer Alix-Garcia, Dan Phaneuf, Ian Coxhead, Jeremy Foltz, Dylan Fitz, and UW AAE Development Workshop participants for their helpful comments.

have slowed slightly in many countries (Figure 2). Through remittances, rural to urban migration can be the means of distributing the gains of economic growth, which tend to disproportionately benefit urban and coastal regions. The movement of large numbers of people from sparsely populated rural regions to often overcrowded urban regions has significant economic and social consequences for both sending and receiving communities, including rural brain drain, increasing number of single-parent households, changes in income distribution (both between regions and within regions), improved access to education, increasing urban population density and demand for urban housing and public services, and increased interaction between different ethnic and religious groups. Understanding the determinants of rural to urban migration is of great importance since it enables social scientists to generate more accurate predictions regarding the magnitude and character of migratory flows and, consequently, better inform policy.

The majority of social scientists studying the motivations of rural to urban migration emphasize the importance of regional wage differentials. However, rural households with similar financial and demographic characteristics often have different migratory responses to the same urban/rural income gap, suggesting that other factors must also play a significant role in the migration decision. A growing body of migration research postulates that aversion to risk may play an important role in encouraging or inhibiting migration (Stark and Levhari, 1982; Stark, 1984; Katz and Stark, 1986; Taylor, 1987).

Some studies find inferential evidence supporting the household risk-management migration theory (Rosenzweig and Stark, 1989; Wouterse and Taylor, 2008); however, to my knowledge, no paper directly tests the theory using risk-aversion parameters. In response to the increased availability of risk preference survey data, more recent research investigates the relationship between risk attitudes and migration (Guiso and Paiella, 2004; Jaeger et al., 2010) and find that risk tolerance and the propensity to migrate are positively correlated. However these papers simply analyze the relationship between an individual's risk attitude and the propensity to migrate and rely the simplifying assumptions that individuals face only two migration options (migrate or don't migrate) and migration is associated with both higher risk and higher reward.

The reality is much more complicated since migrants must choose between multiple destinations, each associated with a different expected income and expected variance of income. If risk preferences are heterogeneous, then we expect these preferences to interact with heterogeneous regional characteristics to influence where migrants go. Many social scientists argue that migration is more accurately modeled as a household decision, in which case migration may be a means of diversifying household income flows. If this assumption is accurate, then the relationship between risk aversion and migration is much more complicated, as discussed in Section 3.

This paper develops a theoretical model of household migration decision-making where income-pooling households with heterogeneous levels of risk aversion must choose whether to send a migrant and, if so, the destination of that migration. Under these assumptions, the model predicts that the variance of consumption in the receiving region and the correlation of consumption in the home and receiving regions both interact with household risk preferences in influencing both the propensity to migrate and the destination of migration. I test the predictions of the model using household-level panel data from the Indonesian Family Life Survey (IFLS). This paper will proceed as follows: Section 2 summarizes the setting, Section 3 discusses the relevant literature, Section 4 describes the theoretical model, Section 5 describes the data, Section 6 presents the proposed estimation strategy, Section 7 summarizes the results, Section 8 discusses possible future directions, and Section 9 concludes.

2 Setting

The last few decades have been marked by incredible economic and demographic changes in Indonesia. The national population has grown rapidly from 119 million in 1971 to 230 million in 2008 (Farré and Fasani, 2011). Population growth, combined with disparate regional economic growth, have prompted massive urbanization. Between 1971 and 2005, urbanization increased from 17.3% to 43.1% (Anata and Arifin, 2008) which was mostly driven by rapidly growing internal migration flows. In 1971, only 4.9% of the population described themselves as interprovincial migrants, however this number steadily increased until, in 2000, it reached the high rate of 10.1% (Hill et al., 2008; Tirtosudarmo, 2009). Combined with an even higher intraprovincial rate, it's clear that the percentage of population participating in internal migration is much larger than those engaged in international migration which, in 2006, comprised only 1.5% of the total population (Ducanes and Abella, 2009).

The current interprovincial migratory flows usually take the form of rural to urban migration and with the largest cities, such as Jakarta, Surabaya, and Makassar, acting as popular destinations (Lu, 2010); however, the geographic pattern of sending and receiving provinces has certainly changed over time. In general, recent interprovincial migrants seem more willing or able to travel longer distances. Some provinces, such as Lampung and Jakarta, historically had net positive worker inflows but have recently changed to net outflows – whereas the opposite is true for other provinces like Bali and West Nusa Tenggara. Some provinces (North Sumatra, South Sulawesi, Central Java, and East Java) continue to be net senders of labor while others (Jambi and East Kalimantan) have remained net receivers (Anata and Arifin, 2008).

These shifting mobility trends may be, in part, motivated by Indonesia's recent volatile economy. Between 1960 and 1995, per capita income rose by 1500%, largely as a result of increased industrialization and oil exports. This growth created an optimistic outlook until 1997, when the Asian Financial Crisis devastated the economies of East and Southeast Asian countries. While Indonesia fared better than many other countries, it experienced a dramatic decrease in real GDP, which fell by more than 12% between 1997 and 1998, and high inflation rates of 80%. The economic troubles of the late 1990s spurred political instability and citizens in several cities took to the streets in protest. During this time, several policy changes were instituted, including the removal of subsidies and the resignation of President Suharto in May of 1998. In additional to economic shocks, Indonesia has also suffered from several devastating natural disasters, including the drought of 1997, the 2004 tsunami, and the 2006 earthquake. While every Indonesian was affected by these events, the benefits and costs were distributed very unequally across regions and socio-economic groups. For example, the 42% of the labor force which comprises the agricultural sector was particularly vulnerable to the stochastic shocks in weather and world food prices. So it comes as no surprise that the tumultuous events of the last four decades have been accompanied by large internal migratory flows. However, while the shifting regional wage gaps were probably a large determinant of internal migration, it is possible that other factors were at play. Specifically, the volatile economy may have prompted many families to adopt a new risk-management strategy via migrant remittances.

Remittances play a prominent role in the Indonesian economy, particularly in some regions. For example, in 1995 the aggregate remittances received in an East Javan kabupaten totaled to about $7.2 million USD which several times greater than the kabupaten's government budget (Sukamdi et al., 2004). Funds sent home during times of financial stress could definitely have a large impact on a household's ability to smooth consumption.

3 Literature Review

One of the first academics to develop theories of migration was a German cartographer by the name of Ernst George Ravenstein. Analysis of census records prompted Ravenstein's publication of the "Laws of Migration," which was essentially a list of empirical observations such as: "urban residents are often less migratory than inhabitants of rural areas" and "most migrants are adults" (Ravenstein, 1885). While the modern validity of many of these laws has been called into question (Davis, 1988), Ravenstein drew attention to the potentially predictable nature of migration flows and sparked the interest of many other academics. Subsequent researchers tended to categorize migratory forces into what Everett Lee formally characterized as "Push and Pull Factors" (Lee, 1966). Push factors tend to be unfavorable events in the source community which encourage out-migration and include, for example, natural disasters, lack of employment opportunities, and war. Pull factors attract in-migration and may include religious freedom, employment and educational opportunities. Many researches have applied the Push and Pull framework to empirical investigations and found that most changes in migratory flows can be explained by either a push or a pull factor (Thomas, 1973; Lowell, 1987).

While these early theories provided a structured way of thinking about migratory trends, they don't generally provide specific and testable predictions regarding the magnitude of labor flows. Neoclassical economists filled this void by developing models that focus on how migration responds to wage differentials and moving costs. Neoclassical migration models are generally characterized by the assumption that migration is driven by wage differentials. These models assume that individuals move in order to maximize individual income, thus driving up wages in the sending region and lowering wages in the receiving region until an equilibrium is reached, at which point migratory flows will cease (Lewis, 1954; Ranis and Fei, 1961; Harris and Todaro, 1970). Todaro emphasized the importance of accounting for unemployment rates in the receiving regions and claimed that migration flows were driven by disparities in expected wages (Todaro, 1969). Some neoclassical models focus on the microeconomics of the decision and characterize migration as an investment where individuals must pay an upfront migration cost which includes both pecuniary and non-pecuniary elements. Unlike the more macro models, these micro models allow individuals to have heterogeneous skill levels and, consequently, heterogeneous expected wage differentials (Sjaastad, 1962). Since a large part of the migration cost is a fixed cost, the temporal element plays an important role in these models. Specifically, it is assumed that an individual will choose to migrate only if the expected discounted net gains of migration are positive, where expected post-migration wages equal the wages an individual expects to earn, with his given skill set, multiplied by the probability of employment (Borjas, 1989).

The neoclassical emphasis on wage differentials is certainly important in the migration decision; however, observed empirical trends are sometimes at odds with the predictions of neoclassical models suggesting that other factors are simultaneously playing critical roles. Specifically, the neoclassical models predict that migratory flows between two regions should be highest when the expected wage differentials are highest, however, researchers often observe very low out-migration rates from the poorest regions and sometimes observe migratory flows persisting even when the expected wage gap is zero.

The New Economics of Labor Migration (NELM) attempts to explain these surprising results by emphasizing the role of risk and credit-constraints in the migration decision. NELM models assume that the migration decision is made at the household level and view migrations as a means of overcoming failures in insurance or credit markets (Stark, 1984; Katz and Stark, 1986; Taylor, 1987). The decision to send a migrant might be motivated by a risk-averse household's desire to diversify household income flows, thereby, lowering the variability of household income. If this is the case, then wage differentials may not be necessary if the sending and receiving wages are negatively correlated or have low correlation (Stark and Levhari, 1982).

The Neoclassical and NELM migration models imply very different policy prescriptions. Neoclassical models emphasize the role of labor markets, suggesting that migration is most responsive to policies that impact wages or employment opportunities. In contrast, NELM models assert that migration is motivated not only by expected wage differentials but also by consumption smoothing, credit constraints, and risk aversion, which implies that migration is responsive to policies that impact the redistribution of wealth within a community or improve access to insurance and credit markets. It is important to note that while the Neoclassical and NELM migration models clearly reflect different foci, they are not necessarily in contradiction with one another since it is possible that households are motivated by one set of factors and individuals within a household are motivated by other factors.

Clearly the decision to migrate is complicated and incorporates a large number of factors, however, in my analysis, I focus on the role of risk, which has not be adequately analyzed elsewhere. Specifically, I test whether the propensity to migrate and the choice of migration destination are impacted by the interaction of household risk aversion with risk in the sending and receiving regions. Other studies test multiple destination models but focus on the role of expected income gaps while either abstracting from risk and risk aversion (Davies et al., 2001; O' Keefe, 2004) or assuming homogenous risk aversion (Kennan and Walker, 2011).

Chen et al. (2003) develop a theoretical model of household migration that allows for the possibility of heterogeneous risk aversion, but in a single-location model that does not yield predictions about where migrants will go. While the predictions of the Chen et al. model are consistent with the notion that migration is a means of risk management, those predictions are not well tested since their main contribution is theoretical rather than empirical. While the authors highlight a historical example from Taiwan which is consistent with their model, they do not directly test the model's predictions or focus on the role that household risk aversion plays in the migration decision.

In a similar paper, Daveri and Faini (1999) use Italian data to investigate whether the destination decision is impacted by either the variability of income or the correlation of sending and receiving income. Specifically, they develop and test a model that allows migrants three migration options: i) stay home, ii) migrate internationally, or iii) migrate internally; however, they assume that households have homogenous risk preferences and, therefore, homogeneous responses to risk.

Other papers use an inferential approach to determine if migration is motivated by risk-management. Rosenzweig and Stark (1989) do not have data on risk aversion so instead they test the implications of the Katz and Stark (1986) model by examining trends in marriage-induced migration in southern India. Specifically, they postulate that more vulnerable households have stronger incentives to minimize the correlation between the income of the sending household and the income of the receiving household. In other words, there should be a positive relationship between the variability of the sending household income and the geographic distance between the sending and receiving households.

The authors find that the data is consistent with the predictions and that households experience a decrease in consumption variance after a marriage has taken place, however, they only are examining the migration destination conditional on marriage-induced migration (in other words, they are treating the migration as a given). Marriage migration is implicitly different than employment-motivated migration since an individual in a small village does not necessarily need to migrate in order to acquire employment.

Like Rosenzweig and Stark (1989), Wouterse and Taylor (2008) also lack risk-aversion data and must therefore use an inferential approach to test whether migration in rural Burkina Faso is a means of smoothing household consumption. The authors argue that if migration is a form of insurance, then we should observe migration acting as a substitute for other methods of consumption smoothing. Specifically, they postulate that if migration makes households less vulnerable, then households who have sent a migrant should be less concerned with diversifying their household activities and should shift production activities away from low risk and low reward investments into investments with a higher profit variance. The authors do not find a statistically significant impact of continental migration on the sending household farm activities; however, they find that inter-continental migration (which is generally associated with larger and longer-lasting remittances) is associated with a shift to riskier production activities and lower household income diversification net of remittances.

In response to the increased availability of risk-aversion data, more recent researchers have utilized risk-aversion survey data to investigate the relationship between risk and migration (Guiso and Paiella, 2004; Jaeger et al., 2010). However, these papers simply analyze whether an individual's level of risk aversion is correlated with the propensity to migrate and do not investigate how risk aversion impacts the choice of destination or whether migration is form of risk-management.

Bryan et al. (2012) use an experimental approach to investigate the hypothesize that the uncertainty associated with migration discourages internal seasonal migration in Northwestern Bangladesh. The authors test the theory by offering individuals in a treatment group a risk-free loan intended to cover migration costs. Bryan et al. also develop a theoretical model that relies on several assumptions: 1) employment in the home region is guaranteed while employment in the receiving region is uncertain, 2) expected income (net of migration costs) in the receiving region are higher than in the home region, and 3) wages in the sending and receiving region are independent. Unlike the first two assumptions, the last assumption is not stated overtly by the authors but is instead implicit in the model. The results of the experiment indicate that the risk inherent in migration discourages migration and results in a sub-optimal outcome. However the researchers do not have risk-aversion data and do not distinguish between the various migration destinations and, therefore, cannot test whether heterogeneity in the variability in receiving incomes and heterogeneity in risk preferences impact the destination choice.

4 Theory and Predictions

Consider a country composed of H rural regions and J urban regions and the two sets of regions are disjoint. Rural inhabitants are mobile and can migrate to urban sectors. Any region $h \in H$ or $j \in J$ is associated with a region-specific baseline income (f_h or f_j) which is subject to a region-specific shock (ε_h or ε_j). Individuals are homogenous in productivity and skills so every inhabitant of urban region j earns income $y_j = f_j + \varepsilon_j$ and every inhabitant of rural region h earns $y_h = f_h + \varepsilon_h$. The shocks are assumed to have zero mean so $E(\varepsilon_j) = E(\varepsilon_h) = 0$ and $E(y_j) = f_j$ and $E(y_h) = f_h$. Each region's shock has a region-specific variance σ_h^2 or σ_j^2. Each of the $H \times J$ pairwise combinations of rural and urban regions is associated with a correlation of incomes σ_{jh} and a migration cost d_{jh}.

Consider a household with n household members that is living in rural region h . For simplicity, assume that the household is deciding where to locate an individual household member. This one household member can migrate to one of J urban destinations or stay home in rural region h, so the household has $J + 1$ migration options. If the household member migrates to region j then he experiences a net expected income gain of $y_j - y_h - d_{jh}$ which may be negative. The household allocates the individual in order to maximize the household utility function:

$$U = E(I) - aVar(I) \tag{1}$$

where I is total household income and $a > 0$ is a measure of household risk aversion with higher values of a indicating higher aversion to risk. In their analysis of migration among risk-averse households, Chen et al. (2003) utilize this same utility function and argue that this simple utility function sufficiently captures the notion that all households benefit from higher average income and all risk-averse households derive negative utility from more volatile income. Notice also that this functional form assumes that households with higher values of risk aversion experience a greater marginal disutility for a given change in income variability.

In this analysis, I assume the default migration option is staying home and, therefore, I compare all other migrations destinations to the home option. If we define the net marginal utility of a household member migrating from region h to destination j , ΔU_{jh}, as utility associated with sending one household member to region j minus the utility associated with keeping all of the household members home, then a household will send a migrant to destination j if, and only if, the net expected marginal utility of sending an individual to j is both positive and greater than the expected net marginal utility of sending the individual to any other urban destination $l \neq j$.

$$\Delta U_{jh} > max\{0, \Delta U_{lh}\} \quad \forall l \neq j. \tag{2}$$

where

$$\Delta U_{jh} = U_{jh} - U_{hh} \tag{3}$$

Applying the functional form of the utility function in (1) yields:

$$U_{jh} = E\left(n\left(f_h + \varepsilon_h\right) + f_j + \varepsilon_j - d_{jh} - f_h - \varepsilon_h\right) - aVar\left(\left[n-1\right]\left(f_h + \varepsilon_h\right) + f_j + \varepsilon_j - d_{jh}\right) \tag{4}$$

and

$$U_{hh} = E\left(n\left(f_h + \varepsilon_h\right)\right) - aVar\left(n\left(f_h + \varepsilon_h\right)\right) \tag{5}$$

which reduces to:

$$U_{jh} = (n-1)f_h + f_j - d_{jh} - a\left[(n-1)^2\sigma_h^2 + \sigma_j^2 + 2(n-1)\sigma_{jh}\right] \tag{6}$$

and

$$U_{hh} = nf_h - a\left(n^2\sigma_h^2\right) \tag{7}$$

respectively. Substituting (6) and (7) into (3) yields:

$$
\begin{aligned}
\Delta U_{jh} &= \left(nf_h + f_j - d_{jh} - f_h - a\left[(n-1)^2\sigma_h^2 + \sigma_j^2 + 2(n-1)\sigma_{jh}\right]\right) - \left(nf_h - a\left(n^2\sigma_h^2\right)\right) \\
&= \left(f_j - d_{jh} - f_h - a\left[(n^2-2n+1)\sigma_h^2 + \sigma_j^2 + 2(n-1)\sigma_{jh}\right]\right) + a\left(n^2\sigma_h^2\right) \\
&= (f_j - d_{jh} - f_h) - a\left[\sigma_j^2 + 2(n-1)\sigma_{jh} - (2n-1)\sigma_h^2\right]
\end{aligned} \tag{8}
$$

4.1 Comparative Statics

Partial derivatives of condition (8) yields predictions regarding the marginal impact of changes in key parameters.

- $\frac{\partial \Delta U_{jh}}{\partial f_j} = 1 > 0$, meaning an increase in receiving income increases the expected net marginal utility of migrating to destination j
- $\frac{\partial \Delta U_{jh}}{\partial d_{jh}} = -1 < 0$, meaning an increase in the cost of migration decreases the net marginal utility of migration from h to j
- $\frac{\partial \Delta U_{jh}}{\partial f_h} = -1 < 0$, meaning an increase in home income will discourage out-migration
- $\frac{\partial \Delta U_{jh}}{\partial \sigma_j^2} = -a < 0$, meaning an increase in the variance of income in receiving location j will discourage migration to j.
- $\frac{\partial \Delta U'_{jh}}{\partial \sigma_h^2} = a(2n-1) > 0$, meaning an increase in the variance of home income will encourage out-migration
- $\frac{\partial \Delta U_{jh}}{\partial \sigma_{jh}} = -2a(n-1) < 0\, if\, n > 1$, meaning that an increase in the correlation of incomes in the sending and receiving region will decrease the likelihood of migration from h to j, only if there is more than one member in the household. This result is unique to household migration models that assume post-migration income sharing.

The previous comparative statics are straightforward; however, it is less obvious how changes in the household risk-aversion parameter impact the migration destination decision. The partial derivative with respect to household risk aversion equals the negative of the change in the variance of household income resulting from sending a migrant to destination j:

$$\frac{\partial \Delta U_{jh}}{\partial a} = -\left[\sigma_j^2 + 2(n-1)\sigma_{jh} - (2n-1)\sigma_h^2\right]$$

An increase in the household risk aversion will decrease the net marginal utility of migration to destination j if, and only if, the resulting change in the variance of household income is positive. In other words, if migration from h to j results in a more variable household income, then risk aversion has a negative impact on net marginal utility. If, instead, migration from h to j results in a less variable household income, then risk aversion has a positive impact on net marginal utility. This makes intuitive sense since households with higher levels of risk aversion receive greater marginal disutility from an increase in income variability or, alternatively, higher marginal utility from a decrease in income variability. In this context, migration can be used to reduce income variability.

Another way to express this is by taking the second derivative of the parameters which impact the variance of household income, with respect to risk aversion:

$$\frac{\partial \Delta U_{jh}^2}{\partial \sigma_j^2 \partial a} = -1 < 0 \tag{9}$$

$$\frac{\partial \Delta U_{jh}^2}{\partial \sigma_{jh} \partial a} = -2(n-1) < 0\, if\, n > 1 \tag{10}$$

$$\frac{\partial \Delta U_{jh}}{\partial \sigma_h^2 \partial a} = (2n-1) > 0 \tag{11}$$

meaning that an increase in the risk-aversion parameter will exacerbate any change in net marginal utility resulting from a given change in household income variability (both positive and negative). If households are using migration to mitigate risk, then we would expect to observe the interactive impacts in (9)-(11). Households with higher risk aversion should be more likely to choose destinations that lower the overall variance of household income, either because destination income is less variable or less correlated with home income. Less risk averse households are not as concerned about lowering the variance of household income and will, therefore, be more likely to choose destinations where income is more variable and more highly correlated with home income, controlling for receiving income. In summary, the theoretical model yields the following predictions:

- **H_0 1:** Households prefer destinations with higher income
- **H_0 2:** Households prefer destinations associated with lower migration costs
- **H_0 3:** Households prefer destinations with lower income variability
- **H_0 4:** Households prefer destinations with lower income inter-temporal correlation with home income
- **H_0 5:** Households with higher levels of risk aversion are less likely to choose destinations with more variable income
- **H_0 6:** Households with higher levels of risk aversion are less likely to choose destinations where income is highly correlated with home income

5 Data and Summary Statistics

I will be utilizing data from the Indonesian Family Life Survey (IFLS) which is an on-going longitudinal survey in Indonesia that is publicly accessible at rand.org. The RAND Corporation collaborated with various institutions in collecting four rounds of data in 1993, 1997, 2000, and 2007. The IFLS survey consists of individual, household, and community level questionnaires. In this section, I discuss the IFLS sampling methods, describe the tracking protocol and attrition rates, and define the structure of the key variables used in my empirical analysis

5.1 Sampling

The IFLS samples households from 13 of the 27 provinces (Figure 3) and is representative of 83% of the Indonesian population. This is remarkably high since designing a sampling protocol which is representative of Indonesia's total population is virtually impossible due to the country's expansive island terrain and diverse culture. Indonesia is spread across approximately 17,508 islands and its population encompasses over 300 distinct native ethnicities so the cost of implementing a survey which is both thorough and nationally representative would be prohibitively high. In order to design a feasible survey, the IFLS team developed a sampling method which stratified on a subset of provinces and urban/rural locations and then randomly sampled within these strata. Specifically, they selected 13 of the 27 provinces which both represented the Indonesian population and captured its cultural and economic diversity. Some of the remaining 14 provinces were excluded for pragmatic reasons.[1] The baseline IFLS team then randomly chose enumeration areas (EAs) across the 13 provinces and intentionally over-sampled urban EAs and EAs in smaller provinces in order to have sufficient data for urban to rural and Javanese to non-Javanese comparisons. These enumeration areas were derived from a nationally-representative sample frame designed by the Indonesian Central Bureau of Statistics (BPS) for their 1990 national census and a 1993 socio-economic survey (SUSENAS). Each EA consisted of 200-300 households and, within a selected EA, households were randomly selected using BPS listings. While the selection of provinces was not random, the selection of EAs within the provinces and the selection of households within each EA was random and, therefore, an unbiased representation of the true population within each province.

5.2 Tracking and Attrition

Since IFLS is a longitudinal survey, it was important to track the respondents in the subsequent survey rounds. The protocol for tracking and re-interviewing respondents varied slightly in each subsequent wave but followed the same basic protocol. The definition of a *Target Respondent* varies slightly in each wave but is essentially an individual who was, at one point, a member of a surveyed household and has subsequently split off to form a new household. If a household was an original IFLS1 household, then detailed surveys were administered to all of the household members, regardless of whether the individual was present in the IFLS1 survey. If the household was a split-off household, then only the target respondents and their spouses and their biological children were given detailed interviews and all remaining split-off household members were only included in the household roster. In IFLS1 7,224 households and 22,347 individuals were interviewed. Attrition was low and in IFLS2, 94% of IFLS1 households and 91%

[1] "The far eastern provinces of East Nusa Tenggara, East Timor, Maluku and Irian Jaya were excluded due to the high cost of fieldwork in these more remote provinces. East Timor is now an independent state. Aceh, Sumatra's northernmost province, was excluded out of concern for the area's political violence and the potential risk to interviewers." (Page 4 of Strauss et. al, 2004)

of IFLS1 target individuals were re-interviewed, including target respondents that had moved into new households. In IFLS3, 95.3% of IFLS1 households were re-interviewed and in IFLS4 the re-interview rate was 93.6%. Since many of the households experienced new births after the initial 1993 interview and because the survey team tracked down split-off households, the total sample size increases through time. In 1997, 2000, and 2007 the number of sampled households was respectively 7,620, 10,435, and 13,536.

This attrition rate is clearly quite low so it is unlikely that sample attrition has a significant impact on estimated coefficients; however a comparison of the characteristics of those who remained in the sample (for all four waves) to those who eventually fell out of the sample reveals that attrition varies a great deal conditional on the age and location of the respondent. For example, individuals who were age 40 or older in 1993 have much lower attrition rates than individuals who were teenagers in 1993, which results in a shifting sample age distribution through time. The IFLS team tries to adjust for the varying attrition rates by increasing the sampling rate of the younger demographic and generating longitudinal weights which allow researchers to re-weight the data. However, this re-weighting only accounts for disproportionate attrition across enumeration areas and age groups and does not account for other observed and unobserved characteristics that might be correlated with selecting out of the sample. Since attrition does not appear to be random, re-weighting will not completely remove the selection bias.

5.3 Data Structure and Key Variables

The empirical analysis concentrates on employment-driven rural to urban migration that took place between the last two rounds of the survey (2000 and 2007). Although the IFLS has migration data for all four survey rounds, I focus on recent migration because I want to analyze how migration responds to differences in inter-regional income correlation between 1993 and 2000. The IFLS samples households from 13 provinces (Figure 3); however, between 2000 and 2007, some individuals migrated to Indonesian provinces outside of the original sample (Figure 4). The majority of these individuals were successfully located and re-interviewed by surveyors in 2007, however, the number of observations within each of these out-of-sample provinces is limited. In order to assure a large number of observations for each region, I combine out-of-sample provinces with nearby sample provinces to create thirteen regions (indexed by r) which each include one sample province (Figure 5). Each of the 13 regions has two sectors:

rural (indexed by $h \in \{1, 2, ...13\}$) and urban (indexed by $j \in \{1, 2, ...13\}$). A household is considered to be part of region r's rural sector h if it has rural status and lives in region $r = h$. Similarly, each urban sector j is comprised of urban households living in region $r = j$.

An individual is considered an employment-motivated rural to urban migrant if he or she has satisfied the following three criteria: i) moved from a rural sector h to an urban sector j for at least six months between 2000 and 2007, ii) that move was associated with a change in kabupaten,[2] and iii) the move was primarily motivated by the employment of the migrant. Regarding the third criteria, migrating IFLS respondents were asked to describe the main reason for the move and given 24 response options, including "work-related (non-military)" and "job-seeking." If a respondent said the move was work-related, he was then asked whose worked cause the move and given seven possible responses, including "self." In my analysis I only consider migrants who were either job seeking or made a work-related move on his or her own behalf. The year 2000 sample is comprised of 6,990 rural households. Among these rural households, 785 had a member engage in employment-motivated rural to urban migration prior to the year 2000 and are, therefore, dropped from the sample. Among the remaining 6,205 households, 705 had a member engage in employment-motivated rural to urban migration between the year 2000 and 2007. In addition to reporting the motivation of the move, migrants were also asked to provide information regarding the destination and timing of the migration.

Recall from the model presented in Section 3, that the net marginal utility of sending a migrant from rural region h to urban region j depends on seven parameters: income in the sending rural region(f_h), income in the receiving urban region(f_j), cost of migrating from h to j (d_{jh}), household risk aversion (a), variance of income in the sending rural region (σ_h^2) , variance of income in the receiving urban region (σ_j^2), and the correlation between income in the sending and receiving region (σ_{jh}). The IFLS contains income data for both households and individuals, however, household survey income data is widely regarded as susceptible to measurement error so per capita household consumption is often used as a more reliable proxy. Another benefit of using consumption data is that the IFLS team used price data to convert all nominal values into real values.

The measure of regional income in rural region h should approximate what

[2]Also sometimes called regencies, a kabupaten is an Indonesian administrative division that is smaller than a province but typically larger than a city. If an individual moves to a new province then he or she must also be living in a new kabupaten

the average individual in region h would expect to earn in a given year between 2000 and 2007 if he or she stayed in region h. A simple proxy is the per capita household consumption averaged across all households in rural region h in the year 2000:

$$\bar{C}_{h,2000} = \frac{1}{N_h} \sum_{i=1}^{N_h} C_{ih,2000} \tag{12}$$

where $C_{ih,2000}$ is the per capita household consumption of household i in region h in the year 2000 and N_h is the number of households in rural region h. Similarly, the proxy for regional income in urban region j is:

$$\bar{C}_{j,2000} = \frac{1}{N_j} \sum_{i=1}^{N_j} C_{ij,2000} \tag{13}$$

Table 1 compares the rural and urban average per capita consumption for each of the 13 regions.

The measure of variability of income in the home region h is the standard deviation of per capita household consumption across households in region h in the year 2000:

$$sd_{h,2000} = \sqrt{\frac{1}{N_h} \sum_{i=1}^{N_h} (C_{ih,2000} - \bar{C}_{h,2000})^2} \tag{14}$$

and the measure of variability of income in the receiving region is $sd_{h,2000}$. Table 2 compares the standard deviation of rural and urban per capita consumption for each of the 13 regions. The measure of correlation between income in sending region h and receiving region j is calculated across time using the consumption data from earlier survey waves:

$$corr_{hj} = corr(\bar{C}_{h,t}\,, \bar{C}_{h,t}) \quad t \in \{1993,\, 1997,\, 2000) \tag{15}$$

Table 3 presents summary statistics for each of these previously defined variables.

One possible measure of risk attitudes exists in a section included in the most recent IFLS wave (2007) about individual risk preferences. In this section, individuals were asked a series of hypothetical questions where they must choose between two options a guaranteed monetary amount or a lottery over two amounts. As the questionnaire progresses, the expected value and variance of the lottery changes, while the guaranteed amount remains constant. The responses to these questions are used to divide individuals into five distinct categories associated with different levels of risk tolerance. I use the categories to construct two measures of risk aversion: the first is simply a category number ranging from 1 to 5 (where 5 is associated with the highest level of risk aversion) and the second is the lower-bound of the individual's coefficient of relative risk aversion.[3] I use the individual risk-preference parameters to generate measures of household risk aversion by simply averaging across all adult household members that completed the risk-preference survey. The risk-preference questionnaire was administered to every household head, every spouse of the household head, and a random subset of the remaining adults. In most households, this meant that all adult members successfully completed the risk-preference questionnaire. However, there are some households where one or more adult did not and, in these cases, I average over the adults who completed the questionnaire. The risk aversion summary statistics are presented in Table 4.

Overall, the low attrition of the IFLS coupled with it's thorough information regarding migration, consumption, and risk preferences make it a unique dataset.

6 Empirical Tests

Since I am modeling the migration decision as a household utility maximization problem where agents must choose one of several possible discrete destination options, a multinomial logit is an appropriate estimation tool. Other recent empirical studies of migration destination decisions also utilize this estimation strategy (Davies et al., 2001; O' Keefe, 2004; Christiadi and Cushing, 2007). A conditional logit (also called a McFadden Choice Model) is most appropriate since it allows for the inclusion of both household-specific and alternative-specific variables by interacting household-specific variables with alternative dummies (McFadden, 1974). In order to estimate any discrete choice model the set of alternatives must be finite, mutually exclusive, and exhaustive (Train, 2003). In other words, a household must choose one, and only one option. By design, the constructed urban migration regions are mutually exclusive but in order for the choice set to be exhaustive, it must also include the non-migration option.

Therefore, a household residing in rural region h has 14 migration location

[3] The Appendix 2 contains a more detailed explanation of the risk preference questions and how the answers are used to construct CRRA parameters.

options: the 13 urban regions and the home region h. These 14 location options are indexed by the variable $k\epsilon\{1, 2, ...14\}$where $k= j$ if the household sends a migrant to urban region j, and $k = 14$ if the household chooses the non-migration option. The conditional logit model assumes that people will choose the option that maximizes expected utility so household i, living in rural region h, will only choose migration option k if, and only if, the perceived utility associated with option k is greater than the utility associated with the other 13 options:

$$U_{ihk} > U_{ihl} \; \forall l \neq k \tag{16}$$

where U_{ihk} is the utility that household i, residing in rural region h, receives if it chooses migration option k. This utility is a function of household characteristics, destination characteristics, and the migration costs if a move is required.

However, we don't observe all relevant household characteristics, therefore, expected utility has a random component, ε, which the conditional logit framework assumes to be independently and identically distributed with an extreme value distribution. My estimation assumes that if household i residing in rural region h chooses migration option k then they experience the following baseline utility function:

$$\begin{aligned} U_{ihk} = \quad & \beta_1 \bar{C}_k + \beta_2 sd_k + \beta_3 corr_{hk} + \beta_4 (RA_i \times sd_k) \\ & + \beta_5 (RA_i \times corr_{hk}) + \beta_6 dist_{ik} + \alpha_k^{'} (v_k \times Z_i) + \varepsilon_{ihk} \end{aligned} \tag{17}$$

where ε is an unobservable random shock and all other elements are observable. $\bar{C}_k$ is per capita household consumption in region k and sd_k is the standard deviation of per capita consumption across households in region k. [4] Z is a vector of household characteristics including the household risk-aversion parameter, the number of household members, and the age and education of the household head. Household characteristics must be interacted with migration-choice dummy variables,v_k, yielding coefficients α_k. The coefficients on the interaction terms $RA_i \times var_k$ and $RA_i \times corr_{hk}$ are the parameters of key interest. RA_i is the level of household $i's$ risk aversion, and $corr_{hk}$ is the correlation between consumption in the home region h and migration location k where $corr_{hk} = 1$ when $k = 14$. Notice that both of our key parameters are identified by the unique combinations of households and migration locations. The variable $dist_{ik}$

[4] In the non-migration option, $k = 14$, these numbers are calculated within the home region (i.e. average per capita household consumption in region h and variance of per capita household consumption across households in region h)

is the geographic distance between a centroid of the sending kabupaten and a centroid in the receiving kabupaten. [5] The hypotheses presented in Section 4 are congruent with the following estimated coefficient predictions:

- $\mathbf{H_1}$**:** HHds prefer higher income $\leftrightarrow \hat{\beta}_1 > 0$
- $\mathbf{H_2}$**:** HHds prefer lower migration cost $\leftrightarrow \hat{\beta}_6 < 0$
- $\mathbf{H_3}$**:** HHds prefer lower income variability $\leftrightarrow \hat{\beta}_2 < 0$
- $\mathbf{H_4}$**:** HHds prefer lower income correlation $\leftrightarrow \hat{\beta}_3 < 0$
- $\mathbf{H_5}$**:** HHds with higher levels of risk aversion are less likely to choose destinations with more variable income$\leftrightarrow \hat{\beta}_4 < 0$
- $\mathbf{H_6}$**:** Households with higher levels of risk aversion are less likely to choose destinations where income is highly correlated with home income $\leftrightarrow \hat{\beta}_5 < 0$

For ease of notation, I will use the term V_{ihk} to describe all of the observable elements so the utility can be denoted as:

$$U_{ihk} = V_{ihk} + \varepsilon_{ihk} \tag{18}$$

We cannot observe the random shock ε or utility U but we do observe migration decisions which can used to generate probabilities associated with each possible migration destination. In other words, for every pairwise combination of household and destination, $i \times k$, we will generate a predicted probability of household i choosing migration location k:

$$\begin{aligned} Prob(Y_{ihk} = 1) &= Prob(U_{ihk} > U_{ihl})\ \forall l \neq k \\ &= Prob(V_{ihk} + \varepsilon_{ihk} > V_{ihl} + \varepsilon_{ihl})\ \forall l \neq k \end{aligned}$$

Under the conditional logit assumption that ε_{ihk} is independently and identically extreme value distributed, the probability that household i chooses migration option k is

$$Prob(Y_{ihk} = 1) = \frac{e^{V_{ihk}}}{\sum_{l=1}^{14} e^{V_{ihl}}} \tag{19}$$

[5] In the case of the non-migration option, this distance is assumed to equal zero.

Every household has 14 migration options so I create 14 observations for each household where the dependent variable, Y, is equal to one for the location selected by the household and is equal to zero for the remaining 13 options which were not chosen. Since we have 6,205 households in the sample, we have a total of 14*6,205=86,870 observations. The conditional logit framework assumes that households consider the costs and benefits of each migration option and chose the alternative that maximizes the household utility function. Under this assumption, parameters are estimated via maximum likelihood.

One limitation of a conditional multinomial logit model is the assumption of independence of irrelevant alternatives (IIA). This assumption implies that the relative odds of choosing one option over another are the same regardless of what other alternatives may be available or the characteristics of the other alternatives (Train, 2003). Specifically, the relative odds of choosing one option over another

$$\frac{Prob(Y_{ihk}=1)}{Prob(Y_{ihl}=1)} = \frac{e^{V_{ihk}}}{e^{V_{ihl}}} = e^{(V_{ihk}-V_{ihl})}$$

should not depend on any alternatives other than k and l (Train, 2003). IIA can only hold if the addition (or removal) of an alternative has the same proportional impact on the probability of choosing option k and the probability of choosing option l, $\forall\, l \neq k$. In some cases the IIA assumption is reasonable while in other scenarios the IIA property is clearly unrealistic. [6]

We can use a Lagrange multiplier test to determine whether the IIA assumption holds but, since the test is not entirely reliable, it is important to consider whether we would expect the IIA assumption to hold in this particular context.

[6] An example of IIA violation summarized by Train (2003): "Consider the famous red-bus–blue-bus problem. A traveler has a choice of going to work by car or taking a blue bus. For simplicity assume that the representative utility of the two modes are the same, such that the choice probabilities are equal: $P_c= P_{bb} = 1/2$, where c is car and bb is blue bus. In this case, the ratio of probabilities is one: $P_c/P_{bb} = 1$. Now suppose that a red bus is introduced and that the traveler considers the red bus to be exactly like the blue bus. The probability that the traveler will take the red bus is therefore the same as for the blue bus, so that the ratio of their probabilities is one: $P_{rb}/P_{bb} = 1$. However, in the logit model the ratio P_c/P_{bb} is the same whether or not another alternative, in this case the red bus, exists. This ratio therefore remains at one. The only probabilities for which $P_c/P_{bb} = 1$ and $P_{rb}/P_{bb} = 1$ are $P_c = P_{bb} = P_{rb} = 1/3$, which are the probabilities that the logit model predicts. In real life, however, we would expect the probability of taking a car to remain the same when a new bus is introduced that is exactly the same as the old bus.We would also expect the original probability of taking bus to be split between the two buses after the second one is introduced. That is, we would expect $P_c = 1/2$ and $P_{bb} = P_{rb} = 1/4$. In this case, the logit model, because of its IIA property, overestimates the probability of taking either of the buses and underestimates the probability of taking a car. The ratio of probabilities of car and blue bus, P_c/P_{bb}, actually changes with the introduction of the red bus, rather than remaining constant as required by the logit model" (46).

In the case of households choosing migration destinations, it seems unlikely that the inclusion of a new migration option has a proportional impact on the probability of choosing all other options. For example, suppose a household has a strong (and unobserved) preference to keep all household members home. In this case, an additional migration location will have a minimal impact on the probability of choosing the non-migration option but may have a large impact on other migration locations. Suppose instead that a household has a strong preference for destinations with a large Chinese-speaking population. In this case, the addition of another migration location with a large Chinese-speaking population will differentially impact the probability of choosing other locations with a large Chinese-speaking populations and locations where there is not a large Chinese-speaking population. If the IIA assumption is violated then the conditional logit may overestimate the probability of migrating to some destinations while underestimating the probability of other destinations.

I could account for this by attempting to control for any location characteristics which might influence the decision (i.e. the languages most commonly spoken in the receiving region, the average temperature of the receiving region, etc), however, the easiest way to control for the location characteristics is with a specific type of multinomial conditional logit called an alternative-specific conditional (ASC) logit. An ASC logit estimation assumes a default alternative and includes a dummy variable for the remaining alternatives.

Under the assumptions of the ASC logit model, the assumed household utility function would change slightly. Specifically, the alternative specific characteristics $\bar{C}_k$ and sd_k would subsumed by destination fixed effects v_k. The utility becomes:

$$U_{ihk} = \gamma_k^{'} v_k + \beta_3 corr_{hk} + \beta_4 (RA_i \times sd_k) + \beta_5 (RA_i \times corr_{hk}) + \beta_6 dist_{ik} + \alpha_k^{'} (v_k \times Z_i) + \varepsilon_{ihk} \tag{20}$$

While the utility function in (17) may yield biased estimated coefficients, the ASC logit will produce more reliable results.

7 Preliminary Results

Table 6 presents the results of the baseline conditional logit estimation based on the utility function in (17) and Table 7 presents the results of the ASC logit estimation based on the utility function in (20). While the ASC logit is more reliable, one benefit of the standard conditional logit estimation is it allows us to easily test predictions involving alternative-specific characteristics. In Table 6 we see that all of the conditional logit estimates indicate that the average per capita regional consumption is associated with a statistically significant odds ratio greater than one, which is consistent with Hypothesis 1 and suggests that households prefer to send migrants to destinations with higher income. The odds ratio for distance is statistically significant and less than one which is supportive of Hypothesis 2, confirming that households are less likely to send migrants to distant destinations. The results in column 1 are are inconsistent with Hypotheses 3, since we expected an odds ratio less than 1, however, the p-value raises dramatically when we expand the estimation to include correlation of consumption across regions. The impact of income correlation is not statistically significant, however when we examine the results in Table 7, we see that this is not the case in the more reliable ASC logit model.

The results of the ASC logit are consistent with Hypothesis 4 and 5 but not 6. In all four versions of the estimation, the income correlation odds ratio is statistically significant and less than one, suggesting that households prefer to send migrants to destinations with lower correlation with home income. The results are consistent with Hypothesis 5 which predicts that the odds ratio for the RA*SD interaction term should be less than one; however the statistical significance is not robust to the functional form used in the estimation. The results are incongruent with Hypothesis 6 which predicted that the RA*Corr interaction term should also have an odds ratio less than one; however in all versions of the estimation, the coefficient is not statistically different from zero.

8 Challenges to Identification and Alternative Estimation Strategies

The largest challenge to identification is the endogeneity between migration and risk attitudes. In my model, I assume that exogenous household risk preferences determine household migration decisions; however, this is problematic since it's possible that risk preferences are shaped by life experiences. For example Cameron and Shah (2011) also utilize the risk preference data from the Indonesian Family Life Survey and find that there is a strong positive correlation between an individual's exposure to a natural disaster and subsequent risk aversion. In the case of my model, it is possible that the act of sending a migrant impacts the household's attitude toward risk. This would be less of an issue if I had risk-preference data for each of the four waves of the survey but, unfortunately, the risk-preference questions were only added to the most recent wave. In other words, all of the migration documented in the IFLS data takes place prior to the collection of the risk-preference information. The validity of my exogenous risk-preferences assumption is strengthened by recent research on the stability of risk preferences through time. Sahm (2007) analyzes the responsiveness of risk preferences to economic shocks and find that risk preferences are not significantly impacted by changes in economic status.

Alternatively, I could test the predictions of the model by examining household consumption responses to economic shocks. One of the strengths of the IFLS dataset is that it covers a time period where Indonesia experienced several dramatic events including the extreme economic growth of the early 1990s followed by the devastating 1998 Asian Financial Crisis, the 2004 tsunami, and the 2006 earthquake. In each of these cases, I could compare saving and consumption responses across households and determine which households exhibit behavior consistent with high risk aversion. This approach would rely on the assumption that a household with relatively high risk aversion will have a greater tendency to save and will, therefore, experience lower variance of consumption across time. Because the IFLS surveyors asked thorough questions regarding household income, saving and expenditures in each wave, I can calculate variance of household consumption across the survey years prior to 2007:

$$var(C)_{(i,1993-2000)} = \frac{1}{3}\sum(C_{it} - \bar{C}_i)^2 \;\; where \;\; t \in \{1993,\, 1997,\, 2000\}$$

and use the consumption variance as a proxy for risk aversion. Specifically, I would estimate an ASC logit model in equation (20) but replace the previous measure of household risk aversion (RA) with the variance of household $i's$ consumption across time:

$$U_{ihk} = \beta^{'} v_k + \alpha^{'}_k (v_k \times Z_{i,2000}) + \rho(var(C)_{(i,1993-2000)} \times var(C)_{k,2000}) \quad (21)$$

$$+\pi(var(C)_{(i,1993-2000)} \times corr(C)_{hk}) + \gamma^{'} D_{hk} + \varepsilon_{ihk}$$

Under the assumption that variability of household consumption and risk aversion are inversely correlated, we would expect the estimated coefficients $\hat{\rho}$ and $\hat{\pi}$ to be negative in this case

The robustness of my results might be further validated by using risk-taking behavior (i.e. smoking, gambling) to construct alternative proxies for risk aversion. The IFLS(1-4) includes data on whether an individual currently uses tobacco, how frequently the individual consumes tobacco, and whether the individual has ever consumed tobacco. I could use this information to generate either a dichotomous measure of current tobacco use or a discrete measure indicating the frequency of use, and aggregate up to the household level. While these risk-taking proxies might be good robustness checks, it is important to note that they are a noisy measure of risk aversion since they are also an indication of an individual's discount rate which is also almost certainly correlated with migration decisions.

Another useful robustness check would be to test whether the results hold under alternative measures of income variability. Currently the variance of regional income is proxied with the standard deviation of year 2000 per-capita household consumption across all households within a region. While this method effectively captures variability across individuals, it does not capture income volatility through time. One alternative would be to calculate the temporal variance of per-capita consumption between 1993 and 2000 for each household and then average over households within each region. The major drawback of this method is that the number of sample points varies across households (i.e. some households are surveyed in all four waves while others have fewer than four sample points) and is not random. I could drop all households with fewer than four sample points but this is also problematic since I would be non-randomly selecting households (generally younger households) out of the sample. Another method of capturing regional income volatility is simply using the unemployment rate among migrants within a region.

9 Conclusion

Understanding the determinants of migration in Indonesia is important from a welfare perspective. While the remittances associated with migration often benefit the household, there are also several costs including the loss of utility due to the separation of family members. If migration is acting as a substitute for insurance, then the provision of formal insurance markets may increase household welfare since it enables households to insure against shocks without separating individuals from their families. Identifying the determinants of migration is also important from a policy perspective. The governments of countries with large rural to urban labor flows are often concerned with rapidly growing urban populations and the associated strains on urban resources and, consequently pursue policies that discourage migration to urban centers. Understanding the underlying motivators of migration is essential to accurately predicting which policies will be most effective.

This papers finds that household risk aversion and measures of regional economic risk interact to influence households' migration destination decisions in a way that is consistent with the theory that households use migration as means of managing risk. Specifically, I observe that households with higher levels of risk aversion are less inclined to send migrants to destinations with high consumption variability. Also, I observe that all households prefer destinations with low correlation between average consumption in the sending and receiving regions. However, contrary to the predictions of my model, household migration destination decisions are unaffected by the interaction between household risk aversion and the correlation of sending and receiving average consumption.

10 Figures and Tables

Figure 1: Urbanization and Economic Growth

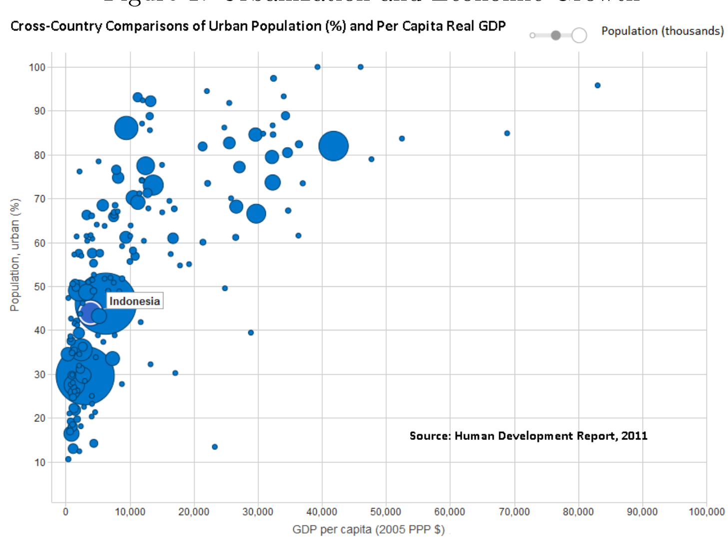


Figure 2: Urbanization in Southeast Asia

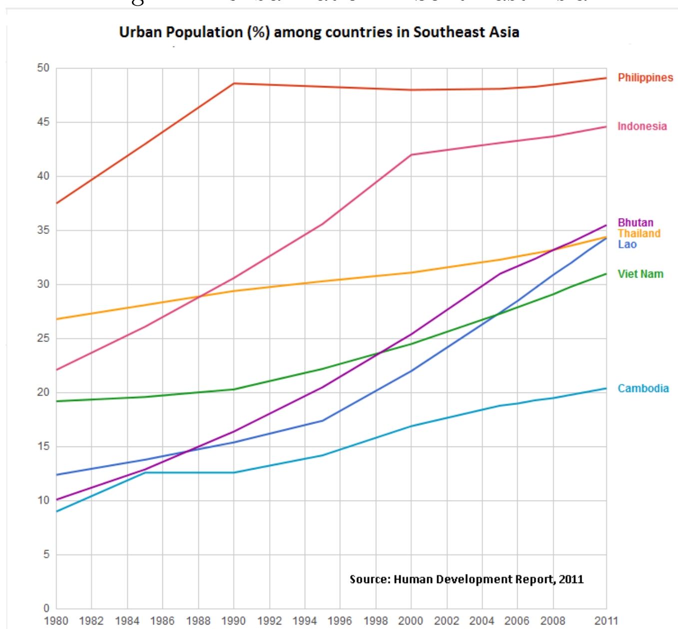


Figure 3: Sample Provinces

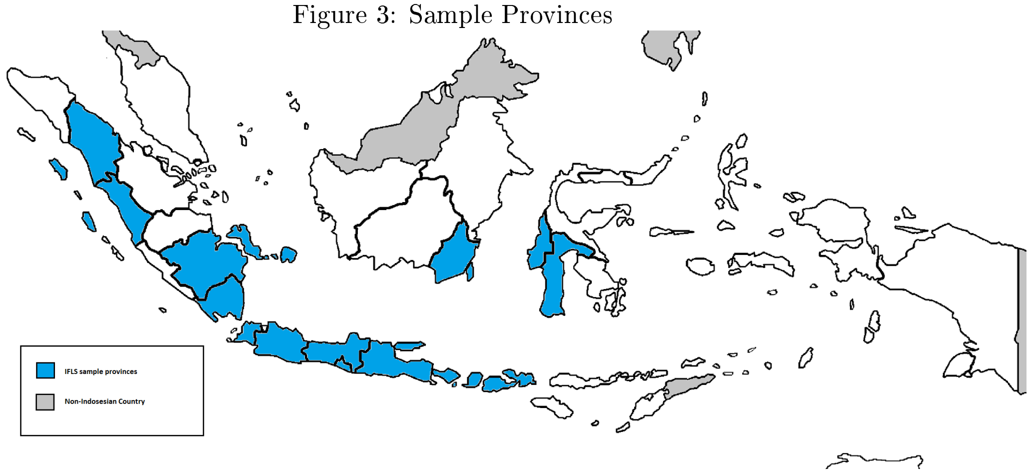


Figure 4: Out-of-Sample Migration

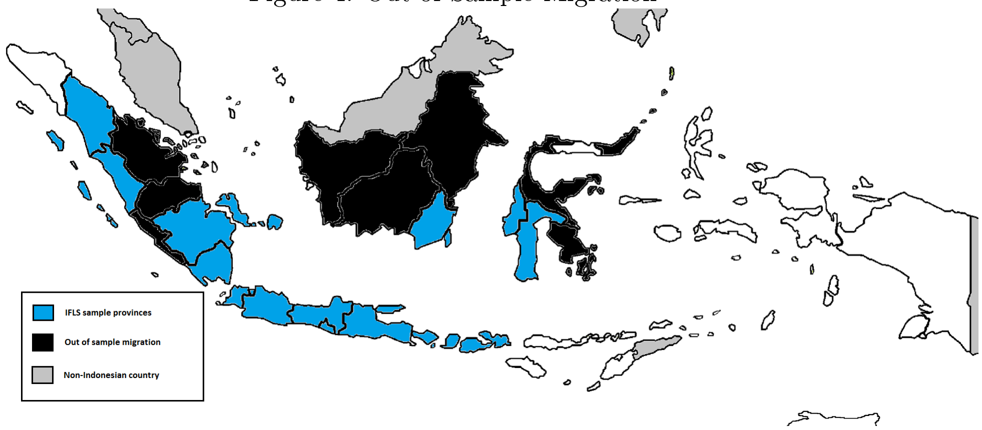
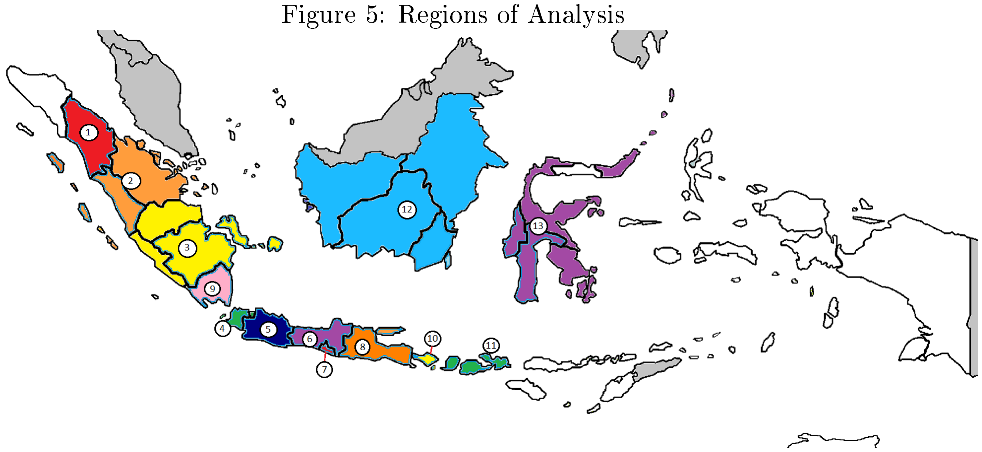


Figure 5: Regions of Analysis



List of Provinces in each Region: 1: N. Sumatera, 2. W. Sumatera and Riau 3. Jambi, S. Sumtera, Bengkulu, and Bangka-Belitung (Note: Banka-Belitung was part of S. Sumatera until 2000) 4. Jakarta and Banten (Note: Banten was part of Jarkarta until 2000). 5. W. Java 6. Central Java 7. Yogyakarta 8. E. Java 9. Lampung 10. Bali 11. W. Nusa Tenggara 12. W. Kalimantan, S. Kalimantan, C. Kalimantan, and E. Kalimantan 13. N. Sulawesi, C. Sulawesi, S. Sulawesi, SE. Sulawesi, and W. Sulawesi

Table 1: **Average Consumption by Region**

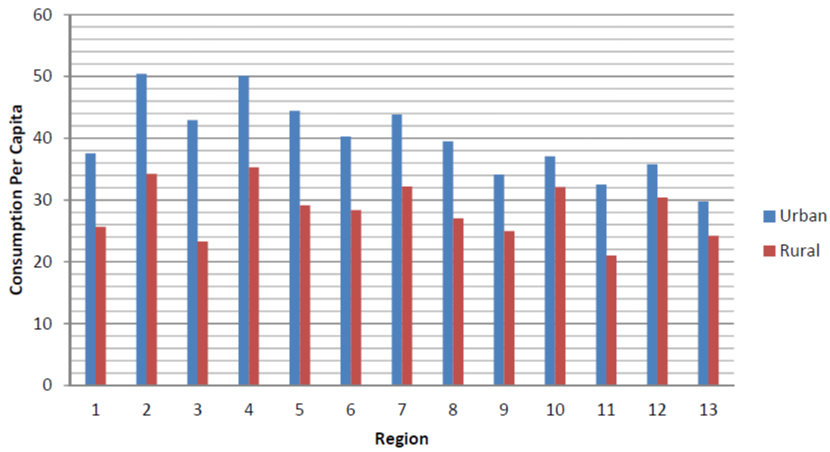


Table 2: **Average Standard Deviation of Consumption by Region**

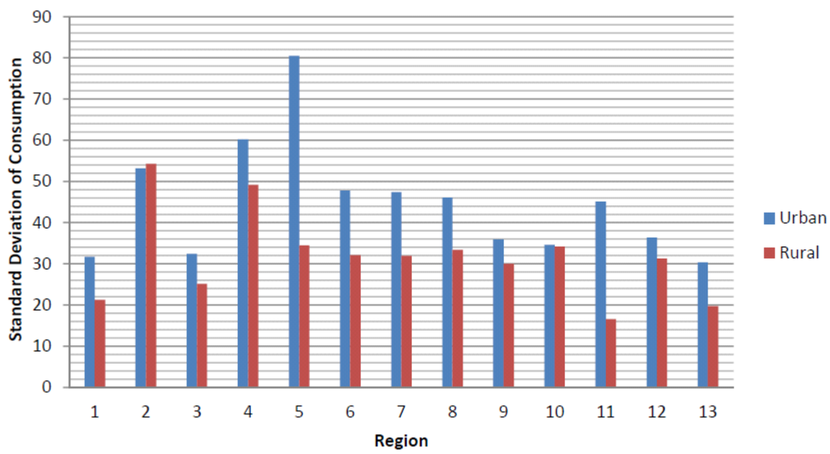


Table 3: **Year 2000 Consumption Summary Statistics Across Regions**

	Obs.	Mean	St. Dev.	Min	Max
Per Capita Consumption (in year 2000 USD)					
Urban	13	39.85	6.33	29.77	50.41
Rural	13	27.71	4.05	21.01	34.24
Standard Deviation of Consumption (in year 2000 USD)					
Urban	13	44.74	14.09	30.34	80.44
Rural	13	30.88	10.18	16.58	54.29
Correlation					
All	169	0.94	0.07	0.67	1.00

Table 4: **risk aversion Summary Statistics**

Measure of Risk Aversion	Obs.	Mean	St. Dev.	Min	Max
Coefficient of Relative Risk Aversion (CRRA)	7183	3.01	1.47	0.01	5.00
CRRA (dropping very risk-averse individuals)	4163	1.59	1.12	0.01	2.92
Risk-Aversion Category (1-5)	7183	3.73	1.10	1	5

Table 5: **Percent of Migrant Households by Region**

Out Migration

Region	Out Migration Rate
1	0.12
2	0.12
3	0.04
4	0.13
5	0.08
6	0.12
7	0.16
8	0.08
9	0.11
10	0.12
11	0.02
12	0.05
13	0.03

Table 6: Results of Conditional Logit Estimation

	(1) Odds Ratio	(2) Odds Ratio
Attribute Variables		
Average Consumption	**1.092**	**1.081**
	[0.012]***	[0.015]***
SD of Consumption		1.006
		[0.005]
Correlation of Consumption	0.195	0.243
	[0.202]	[0.253]
Distance	**0.777**	**0.778**
	[0.013]***	[0.014]***
Observations	4,368	4,368

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

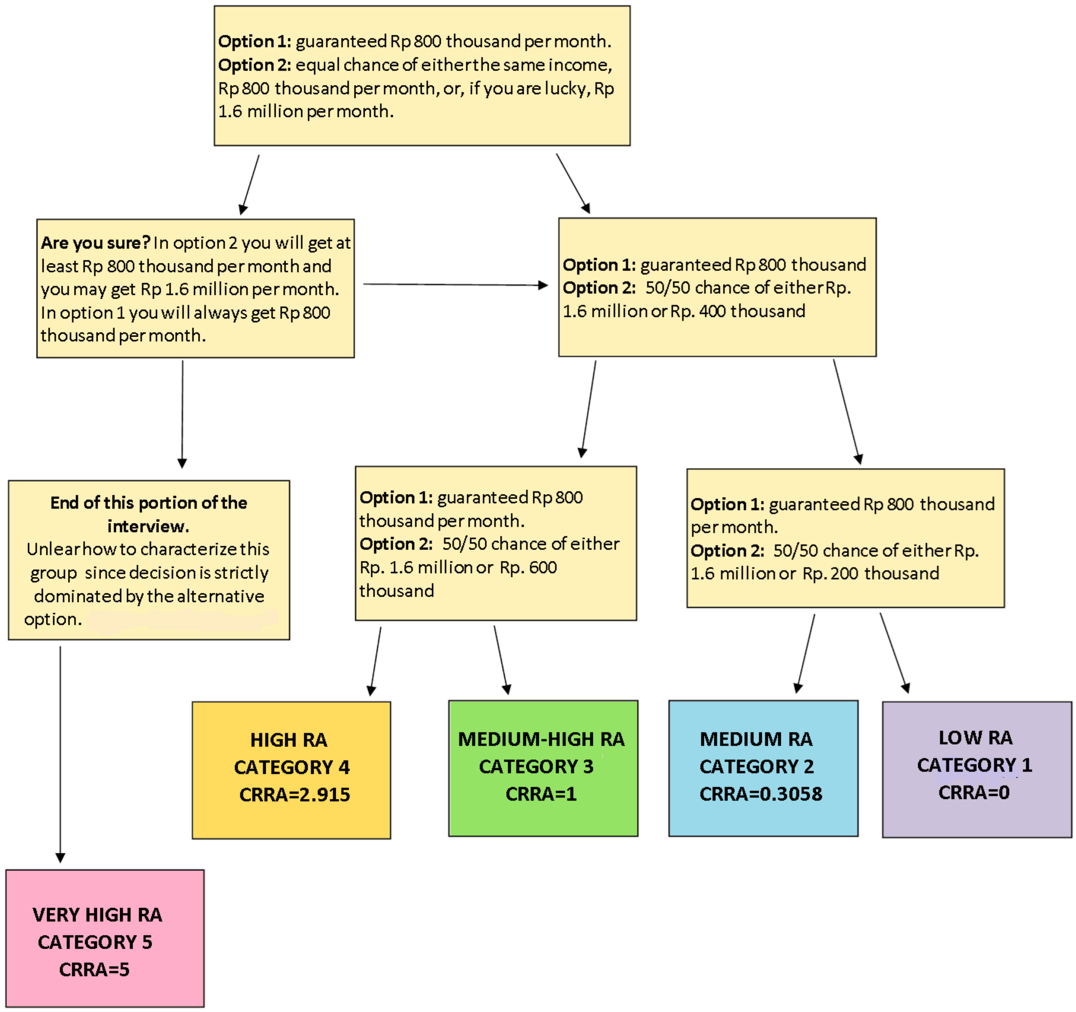
Table 7: Results of Alternative-Specific Conditional Logit Estimation

ASC Logit Odds Ratios

	Using RA 1	Using RA 1	Using RA 2	Using RA 1 (drop highly risk averse)
	(1)	(2)	(3)	(4)
Correlation of Consumption	**0.001**	**0.000**	**0.001**	**0.006**
	[0.002]**	[0.001]**	[0.002]***	[0.009]***
RA * SD of Consumption	**0.996**	**0.996**	**0.995**	0.999
	[0.002]*	[0.002]*	[0.002]**	[0.003]
RA * Correlation of Consumption	2.685	3.165	2.830	2.525
	[2.518]	[2.945]	[2.006]	[2.073]
Distance	**0.760**	**0.760**	**0.760**	**0.770**
	[0.014]***	[0.014]***	[0.014]***	[0.015]***
Origin Fixed Effects	Yes	No	Yes	No
Observations	86,128	86,128	86,128	64,568

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1
Other Controls: HHd Risk-Aversion, HHd size, Education Level of HHd Head -- all interacted with alternative-fixed effects

Figure 6: Risk Preference Response Tree



11 References

Ananta, A., & Arifin, E. N. (2008). Demographic and population mobility transition in Indonesia. *PECC-ABAC Conference on "Demographic Change and International Labor Mobility in the Asia Pacific Region: Implications for Business and Cooperation". Seoul, Korea, March,* 25-26.

Badan Pusat Statistik. (2005). *Statistik 60 Tahun Indonesia Merdeka. Jakarta: Badan Pusat Statistik.*

Badan Pusat Statistik. Population of Indonesia: Results of the 2000 Population Census. (2001). *Series L.2.2. Jakarta: Badan Pusat Statistik.*

Bell, M., & Muhidin, S. (2009). Cross-national comparison of internal migration. *Report Prepared for the 2009 United Nations Human Development Report Office.*

Borjas, G. J. (1989). Economic theory and international migration. *International Migration Review, 23*(3, Special Silver Anniversary Issue: International Migration an Assessment for the 90's), pp. 457-485.

Bryan, G., Chowdhury, S., & Mobarak, A. M.,. (2012). *Seasonal migration and risk aversion.* Unpublished manuscript.

Cameron L. & Shah., M. (2011). *Risk-taking behavior in the wake of natural disasters.* Unpublished manuscript.

Cassar, A., Healy, A., & von Kessler, C. (2011). *Trust, risk and time preferences after a natural disaster: Experimental evidence from thailand.* Unpublished manuscript.

Chen, K., Chiang, S., & Leung, S. F. (2003). Migration, family, and risk diversification. *Journal of Labor Economics, 21*(2), pp. 353-380.

Christiadi, & Cushing, B. (2007). *Conditional logit, IIA, and alternatives for estimating models of interstate migration.* Unpublished manuscript.

Daveri, F., & Faini, R. (1999). Where do migrants go? *Oxford Economic Papers, 51*(4), 595-622.

Davies, P. S., Greenwood, M. J., & Li, H. (2001). A conditional logit approach to U.S. state-to-state migration. *Journal of Regional Science, 41*(2), 337-360.

Davis, K. (1988). Social science approaches to international migration. In M. S. Teitelbaum, & J. M. Winter (Eds.), *Population and resources in western intellectual traditions* (pp. 245-261) Population Council.

Ducanes, G., & Abella, M. (2009). *Prospects for future outward migration flows: China and Southeast Asia* ILO.

Eckel, C. C., El-Gamal, M. A., & Wilson, R. K. (2009). Risk loving after the storm: A Bayesian-network study of Hurricane Katrina evacuees. *Journal of Economic Behavior & Organization, 69*(2), 110-124.

Farré, L., & Fasani, F. (2011). Media exposure and internal migration: Evidence from Indonesia.

Guiso, L., & Paiella, M. (2004). *The role of risk aversion in predicting individual behaviors.* Unpublished manuscript.

Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: A two-sector analysis. *The American Economic Review, 60*(1), pp. 126-142.

Hill, H., Resosudarmo, B. P., & Vidyattama, Y. (2008). Indonesia's changing economic geography. *Bulletin of Indonesian Economic Studies, 44*(3), 407-435.

Jaeger, D. A., Dohmen, T., Falk, A., Huffman, D., Sunde, U., & Bonin, H. (2010). Direct evidence on risk attitudes and migration. *Review of Economics & Statistics, 92*(3), 684-689.

Katz, E., & Stark, O. (1986). Labor migration and risk aversion in less developed countries. *Journal of Labor Economics, 4*(1), pp. 134-149.

Kennan, J., & Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica, 79*(1), 211-251.

Kimball, M. S., Sahm, C. R., & Shapiro, M. D. (2008). Imputing risk tolerance from survey responses. *Journal of the American Statistical Association, 103*(483), 1028-1038.

Knapp, T. A., White, N. E., & Clark, D. E. (2001). A nested logit approach to household mobility. *Journal of Regional Science, 41*(1), 1-22.

Lee, E. S. (1966). A theory of migration. *Demography, 3*(1), pp. 47-57.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School, 22*(2), 139-191.

Lowell, B. L. (1987). *Scandinavian exodus: Demography and social development of 19th-century rural communities*. Boulder: Westview Press.

Lu, Y. (2010). Rural-urban migration and health: Evidence from longitudinal data in Indonesia. *Social Science & Medicine, 70*(3), 412-419.

O'Keefe, S. (2004). Locational choice of AFDC recipients within California: A conditional logit analysis. *Journal of Public Economics, 88*(7–8), 1521-1542.

Ranis, G., & Fei, J. C. H. (1961). A theory of economic development. *The American Economic Review, 51*(4), pp. 533-565.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Statistical Society of London, 48*(2), pp. 167-235.

Rosenzweig, M. R., & Stark, O. (1989). Consumption smoothing, migration, and marriage: Evidence from rural India. *Journal of Political Economy, 97*(4), pp. 905-926.

Sahm, C. (2007). *How much does risk tolerance change?* Unpublished manuscript.

Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of Political Economy, 70*(5, Part 2: Investment in Human Beings), pp. 80-93.

Stark, O. (1984). Migration decision making: De Jong, Gordon F. and Robert W. Gardner, eds., (Pergamon, New York, 1981). *Journal of Development Economics, 14*(1), 251-259.

Stark, O., & Levhari, D. (1982). On migration and risk in LDCs. *Economic Development and Cultural Change, 31*(1), pp. 191-196.

Stark, O., & Taylor, J. E. (1989). Relative deprivation and international migration. *Demography, 26*(1), pp. 1-14.

Stark, O., Taylor, J. E., & Yitzhaki, S. (1986). Remittances and inequality. *The Economic Journal, 96*(383), pp. 722-740.

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herawati, Y., & Witoelar, F. (2004). *The third wave of the Indonesia family life survey (IFLS3): Overview and field report*

. Unpublished manuscript.

Sukamdi, E. S., & Haris, A. (2004). Impact of remittances on the Indonesian economy. *International Migration in Southeast Asia*, 137.

Taylor, J. E. (1987). Undocumented Mexico-U.S. migration and the returns to households in rural Mexico. *American Journal of Agricultural Economics, 69*(3), pp. 626-638.

Thomas, B. (1973). *Migration and economic growth; a study of Great Britain and the Atlantic economy* (2d ed.). Cambridge Eng.: University Press.

Tirtosudarmo, R. (2009). Mobility and human development in Indonesia.

Todaro, M. P. (1969). A model of labor migration and urban unemployment in less developed countries. *The American Economic Review, 59*(1), pp. 138-148.

Train, K. (2003). *Discrete choice methods with simulation*. Cambridge, MA: Cambridge University Press.

Wouterse, F., & Taylor, J. E. (2008). Migration and income diversification:: Evidence from Burkina Faso. *World Development, 36*(4), 625-640.

.

Appendix 1: Possible Alternative Measures of Regional Income and Variance of Regional Income

A1.1: Alternative Measures of Regional Income

While the household consumption data is more reliable, there are some benefits associated with using income data. In this section, I first summarize the IFLS income data and then discuss how that data might be used to construct alternative measures of regional income.

The IFLS does not contain a single, comprehensive household income variable but does, however, provide enough information for household income to be calculated. Specifically, household income is comprised of the following four elements: Farm Income, Non-Farm Business Income, Labor Income, and Non-labor Income. In the farm income portion of the survey, households are asked to report their "approximate amount in rupiah of net profit generated by the farm business during the past 12 months" including produce for own consumption. In the non-farm business income portion of the survey, households are asked if any members operated a non-farm business during the last year and, if so, to report the "approximate amount in rupiah of net profit generated by the business during the past 12 months". Regarding labor income, there are two possible sources. In the roster, the household head (or some adult respondent who is knowledgeable about the characteristics of the household members) is asked to report the total earning of each member in the last 12 months. In a separate questionnaire, a subset of adult household members are asked more detailed questions about their employment status including "approximately how much net profit did you gain last year, after taking out all your business expenses" (if the individual was self-employed) and "approximately how much gross income did you gain last year, including all your business expenses" (if the individual was employed by a firm or the government).

An alternative proxy for regional income could be generated by summing the three income elements for each household and then averaging per-capita household income across the households in each region.

One major drawback of using either mean regional consumption or mean regional income is that it includes many high skill jobs that the average rural-to-urban migrant worker might not reasonably expect to acquire.The IFLS wage data could allow us to overcome this problem in one of three ways. First, we could examine rural to urban migration to each province in order to get a sense of what type of employment a rural migrant to j might reasonably expect to procure. Specifically, for all rural to urban migrations to province j , I could determine the percentage of migrants that successfully procured primary employment within each of the eight employment categories[7] and use the percentages to generate a weighted average income:

$$\bar{I}_j = \sum_{k=1}^{8} \left[\left(\frac{\rho_{jk}}{N_{jk}} \sum_{njk=1}^{Njk} I_{njk} \right) \right]$$

where k indexes the eight employment categories and njk indexes the adult individuals living in province j and working in employment sector k. $\rho_{jk}$is the percentage of rural-to-urban migrants to province j that procured primary employment in sector k. The main shortcoming of this measure is its lack of information regarding the unemployment rate; the calculation does not incorporate unemployed individuals since it relies upon the average income of people working within each sector. I could attempt to account for unemployment rates by multiplying the weighted average with a province-level unemployment rate:

$$\bar{I}_j = \frac{U_j}{N_j} \sum_{k=1}^{8} \left[\left(\frac{\rho_{jk}}{N_{jk}} \sum_{njk=1}^{Njk} I_{njk} \right) \right]$$

where U_j is the number of adult individuals in province j who are in the labor force but did not work at least one hour in previous week. I could further refine this measure by calculating the labor force rate among the subset of individuals who are rural to urban migrants in region j, however this may understate the expected unemployment rate since migrants who do not successfully procure employment before the official six month threshold may get discouraged and return home and thus not satisfy the migrant criteria. Alternatively, I could

[7]The employment categories are: i) agriculture, forestry, fishing and hunting ii) mining and quarrying iii) manufacturing iv) electricity, gas, water v) construction vi) wholesale, retail, restaurants and hotels vii) transportation, storage and communications vii) finance, insurance, real estate and business services viii) social services

multiply the weighted average with the average number of hours worked by urban individuals within the province. A simpler measure of receiving income could be constructed by average income of rural to urban migrants to province j and abstracting from the employment sectors all together. In other words, I would calculate the average post-migration real income of individuals who undertook a rural to urban migration to province j sometime between 1993 and 2000.

Alternatively, I could use the characteristics of the household to generate a first-stage prediction for the wage that an individual from household i in region h might expect to earn if he or she migrated to destination j. These characteristics might include the education and literacy of the household head, the age of the household head, and the primary language of the household member.

Appendix 2: Description of Household Risk Aversion

The risk-preference questionnaire was patterned after questions used in the Mexican Family Life Survey and is composed of two series of questions, each consisting of up to four questions. Each question series can be used to construct a CRRA. Specifically, within each question series, respondents can be divided into four distinct groups based on their answers to these questions. It should be noted that at the time of interview, the exchange rate between the U.S. Dollar and the Indonesian Rupiah was roughly \$1: Rp 10,000; so Rp 800,000 was approximately 80USD. The per capita nominal GDP in Indonesia is approximately 3,000USD.

The sequence of the questions in the first series is summarized in Figure 6. The surveyor starts by asking respondents: "Suppose you are offered two ways to earn some money. With option 1, you are guaranteed Rp 800 thousand per month. With option 2, you have an equal chance of either the same income, Rp 800 thousand per month, or, if you are lucky, Rp 1.6 million per month. Which option would you choose." Under the assumption of rationality, we would expect them to choose option 2 since it strictly dominates option 1. However, contrary to expectations, a large number of respondents chose option 1 (13,079 of 29,054) which makes me question whether the respondents fully understood or trusted this scenario. If a respondent chose option 1, then the surveyor repeated the question and pointed out that option 2 guarantees at least Rp 800 and gave the respondent the opportunity to change his or her previous answer. Surprisingly, only a small number of respondents opted to switch (1,023 of 13,079). If the respondent reaffirmed his/her choice of option 1, then this portion of the questionnaire ended.

Figure 6 illustrates how the responses can be used to separate people into five risk-tolerance categories. The 4,451 individuals with the highest tolerance for risk are characterized as having "low risk aversion". There are 2,309 individuals with "medium risk aversion", 1,593 individuals with "medium-high risk aversion", and 8,464 individuals with "high risk aversion". Characterizing the "irrational" responses is problematic. In one set of estimations I assume that individuals who chose the dominated option are simply extremely risk averse and unwilling to engage in any gamble (even if it is guaranteed to yield a higher payoff). In another set of estimations, I drop all individuals who choose the dominated option. Low, medium, medium-high, high, and very high risk aversion are associated with risk category values of 1,2,3, 4, and 5 respectively.

I also calculate an alternative measure of risk aversion by using the responses to calculate coefficients of relative risk aversion (CRRA) for each respondent. Specifically, the responses are used to calculate a range of possible CRRA coefficients and I assume that each individual's CRRA is equal to the lower bound of this range.