Title
Essays of Economic Development and Migration

Author
Kleemans, Maria Adriaantje

Publication Date
2015

Peer reviewed|Thesis/dissertation

Essays on Economic Development and Migration

By

Maria Adriaantje A Kleemans

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the Graduate Division of the

University of California, Berkeley

Committee in charge:

Professor Jeremy R. Magruder

Professor Alain de Janvry

Professor Edward Miguel

Fall 2015

Essays on Economic Development and Migration

By

Maria Adriaantje A Kleemans

Abstract

Essays on Economic Development and Migration

By

Maria Adriaantje A Kleemans

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkley

Professor Jeremy Magruder, Chair

This dissertation is composed of three chapters and studies issues related to economic development and migration. The first chapter looks at migration choice in an environment where people face risk and liquidity constraints. The second chapter, which is co-authored with Jeremy Magruder, studies the labor market impact of immigration in Indonesia. The third chapter is written together with Joan Hamory Hick and Edward Miguel and examines selection into migration in Kenya.

The first paper develops and tests a migration choice model that incorporates two prominent migration strategies used by households facing risk and liquidity constraints. On the one hand, migration can be used as an ex-post risk-coping strategy after sudden negative income shocks. On the other hand, migration can be seen an as investment, but liquidity constraints may prevent households from paying up-front migration costs, in which case positive income shocks may increase migration. These diverging migratory responses to shocks are modeled within a dynamic migration choice framework that I test using a 20-year panel of internal migration decisions by 38,914 individuals in Indonesia. I document evidence that migration increases after contemporaneous negative income shocks as well as after an accumulation of preceding positive shocks. Consistent with the model, I find that migration after negative shocks is more often characterized by temporary moves to rural destinations and is more likely to be used by those with low levels of wealth, while investment migration is more likely to involve urban destinations, occur over longer distances, and be longer in duration. Structural estimation of the model reveals that migration costs are higher for those with lower levels of wealth and education, and suggests that the two migration strategies act as substitutes, meaning that those who migrate to cope with a negative shock are less likely to invest in migration. I use the structural estimates to simulate policy experiments of providing credit and

subsidizing migration, and I explore the impact of increased weather shock intensity in order to better understand the possible impact of climate change on migration.

The second paper studies the labor market impact of internal migration in Indonesia by instrumenting migrant flows with rainfall shocks at the origin area. Estimates reveal that a one percentage point increase in the share of migrants decreases income by 1.22 percent and reduces employment by 0.26 percentage points. These effects are different across sectors: employment reductions are concentrated in the formal sector, while income reduction occurs in the informal sector. Negative consequences are most pronounced for low-skilled natives, even though migrants are systematically highly skilled. We suggest that the two-sector nature of the labor market may explain this pattern.

The third paper exploits a new longitudinal dataset to examine selective migration among 1,500 Kenyan youth originally living in rural areas. More than one-third of individuals report moving to an urban area during the study period. Understanding how this migration differs for people with different ability levels is important for correctly estimating urban-rural wage gaps, and for characterizing the process of "structural transformation" out of agriculture. We examine whether migration rates are related to individual "ability", broadly defined to include cognitive aptitude as well as health, and then use these estimates to determine how much of the urban-rural wage gap in Kenya is due to selection versus actual productivity differences. Whereas previous empirical work has focused on schooling attainment as a proxy for cognitive ability, we employ an arguably preferable measure, a pre-migration primary school academic test score. Pre-migration randomized assignment to a deworming treatment program provides variation in health status. We find a positive relationship between both measures of human capital (cognitive ability and deworming) and subsequent migration, though only the former is robust at standard statistical significance levels. Specifically, an increase of two standard deviations in academic test score increases the likelihood of rural-urban migration by 17%. Results are robust to conditioning on household demographic and socioeconomic measures that might capture some aspect of credit constraints or household bargaining. In an interesting contrast with the existing literature, schooling attainment is not significantly associated with urban migration once cognitive ability is accounted for. In contrast, academic test score performance is not correlated with international migration to neighboring Uganda. Accounting for migration selection due to both cognitive ability and schooling attainment does not explain more than a small fraction of the sizeable urban-rural wage gap in Kenya, suggesting that productivity differences across sectors remain large.

Chapter 1

Migration Choice under Risk and Liquidity Constraints

Abstract

This paper develops and tests a migration choice model that incorporates two prominent migration strategies used by households facing risk and liquidity constraints. On the one hand, migration can be used as an ex-post risk-coping strategy after sudden negative income shocks. On the other hand, migration can be seen an as investment, but liquidity constraints may prevent households from paying up-front migration costs, in which case positive income shocks may increase migration. These diverging migratory responses to shocks are modeled within a dynamic migration choice framework that I test using a 20-year panel of internal migration decisions by 38,914 individuals in Indonesia. I document evidence that migration increases after contemporaneous negative income shocks as well as after an accumulation of preceding positive shocks. Consistent with the model, I find that migration after negative shocks is more often characterized by temporary moves to rural destinations and is more likely to be used by those with low levels of wealth, while investment migration is more likely to involve urban destinations, occur over longer distances, and be longer in duration. Structural estimation of the model reveals that migration costs are higher for those with lower levels of wealth and education, and suggests that the two migration strategies act as substitutes, meaning that those who migrate to cope with a negative shock are less likely to invest in migration. I use the structural estimates to simulate policy experiments of providing credit and subsidizing migration, and I explore the impact of increased weather shock intensity in order to better understand the possible impact of climate change on migration.

[1] I am sincerely thankful to my advisors, Jeremy Magruder, Edward Miguel and Alain de Janvry, for their guidance and support. This paper has greatly benefited from comments by Michael Anderson, Sam Bazzi, David Card, Michael Clemens, Fred Finan, Meredith Fowlie, Svenn Jensen, Zhimin Li, Ethan Ligon, Yongdong Liu, Aprajit Mahajan and Melanie Morten, as well as seminar participants at College of William and Mary, University of Delaware, NEUDC at Boston University, PacDev at UCSD, University of San Francisco, U.C. Berkeley, GARESC at U.C. Davis and the 2014 Nordic Conference in Development Economics, University of Notre Dame, IFPRI, University of Illinois at Urbana-Champaign, Univerity of Ottawa, University of Amsterdam, Uppsala University, University of Rochester. I gratefully acknowledge financial support from the AXA Research Fund. All errors are my own.

1 Introduction

Approximately 230 million individuals in the world are currently characterized as international migrants, and another 763 million as internal migrants, moving within the borders of their country (Bell and Muhidin, 2013). This migration is partially motivated by large income differences between countries, as well as between areas within a country, for example rural and urban areas. In Indonesia, the focus of this study, those living in rural areas earn 32 percent less than those living in urban areas and this number accounts for differences in prices and employment between rural and urban areas.

While a wide range of reasons may explain the choice to migrate, two primary rationales are often highlighted – especially in a developing country context – as reasons to migrate and, more broadly, as roles that migration can play in the process of economic development. On the one hand, migration can be used to cope with negative income shocks. If a household is hit by a negative shock, for example an agricultural shock due to drought, the household may decide to send a household member elsewhere to earn additional income. This migration strategy can be seen as an alternative to other ex-post risk-coping strategies, such as reducing savings, selling assets, increasing labor supply locally and decreasing consumption.

Alternatively, migration can be used as an investment strategy with the goal of increasing and diversifying future expected income and benefiting from higher wages elsewhere, for example in urban areas. However, as with any investment, this often requires large up-front costs. If a household is liquidity-constrained, it may not be able to make this investment, even if it would be profitable. Therefore, in the presence of liquidity constraints, an increase of wealth – for example due to one or more positive income shocks – may relax liquidity constraints and so increase migration.

While both migration strategies are closely related, they have opposite predictions in terms of the migratory response to shocks. When moving in order to cope with negative shocks, a strategy I will refer to as survival migration, migration increases after negative contemporaneous income shocks. Alternatively, if individuals are liquidity-constrained, migration may increase after (an accumulation of) positive income shocks that help relax liquidity constraints that prevented migration initially. I will refer to this strategy as investment migration.

Both migration strategies are widely observed and documented empirically but described as separate phenomena and in different papers. The survival rationale of migration is described for example in Kleemans and Magruder (2014) and Morten (2013), who find that sudden negative rainfall shocks induce people to migrate internally.[1] Evidence of the investment strategy is documented by Bryan, Chowdury and Mobarak (2014) and by Bazzi (2014), who find that beneficial migration is prevented by liquidity constraints and that overcoming these constraints by subsidizing migration or through positive income shocks increases out-migration. The difference between Kleemans and Magruder (2014) on the one hand and Bazzi (2014) on the other hand seems puzzling as both papers study the Indonesian context but find opposite responses to rainfall shocks. However, the discrepancy may be understood by recognizing that different types of migration are observed: Kleemans and Magruder (2014)

[1] Other papers that empirically observe increased migration after negative income shocks include Mueller, Gray and Kosec (2014), De Weerdt and Hirvonen (2013) and Boustan, Fishback and Kantor (2010).

focus on internal, short-distance migration, while Bazzi (2014) studies international migration that requires large up-front migration costs, making liquidity constraints more likely to be binding.

This paper provides a unified framework of migration choice that incorporates both survival and investment rationales for migration. I develop a migration choice model that encompasses both migration strategies and that improves on previous migration models by allowing for multiple moves over time, between multiple locations, and by incorporating wealth as an important determinant of migration choice. This model is dynamic in nature, to allow for people to plan future migrations and save up for migration over time to overcome liquidity constraints. It builds on the dynamic savings model by Deaton (1991), in which people have a certain amount of wealth and, after receiving a stochastic wage draw in each time period, must decide how much to save in order to smooth consumption and maximize utility over time. I extend this to become a migration choice model by including the current location as an additional state variable and migration choice as an additional control variable. The basic intuition can be explained by a simple three-location model in which a household can decide to migrate away from its home location to either a nearby rural area at a low migration cost, but where wages are only slightly higher than at home, or to a further-away urban area with higher costs and higher wages.

In each period, the household observes a wage draw at its current location from a known distribution. If the household receives a bad wage draw and does not have sufficient savings built up, they may prefer to move to another location to receive a different wage. To avoid high migration costs, the household would likely prefer to move to a nearby rural location just to get another wage draw. I explicitly model a disutility of being away from the home location, which predicts that survival migration will be short in terms of distance as well as duration.

On the other hand, households may try to save up for migration as an investment to benefit from higher wages in a further-away city. If they are liquidity-constrained, then an accumulation of positive shocks may push them over the barrier, after which they are able to cover migration costs. The model therefore predicts that this type of migration is more likely to occur over longer periods of time.

I solve the dynamic migration choice model numerically and test the predictions of this model using a rich dataset of internal migrants in Indonesia. As part of the Indonesia Family Life Survey, all migration moves of 38,914 individuals were recorded over a 20-year period. Individuals were carefully tracked as they changed location, allowing me to study all migration decisions that individuals made, even if they are of short duration and over short distances. After showing that rainfall shocks are good proxies for income shocks, and that a sequence of positive rainfall years helps households accumulate wealth, I study the migration response to rainfall shocks. In line with the model, I find that migration increases both after contemporaneous negative rainfall shocks and after an accumulation of previous positive shocks. Also in agreement with the model, I find that survival migration is more likely to be temporary, have a rural destination, and be used by those with low levels of wealth. Investment migration, on the other hand, is more likely to occur over longer distances and to urban areas, and is longer in duration.

I then structurally estimate the model using maximum likelihood estimation in a mixed logit framework in order to retrieve individual migration cost parameters. The average migration costs of going to nearby rural area locations, which are used mostly for survival migration, are approximately equal to 20 percent of annual income. Investing in migration to a more distant, urban area is about 4 times as costly, slightly more than average annual income. Examining heterogeneous effects reveals that migration is about 30 percent more costly for those with lower levels of wealth and education, and approximately 50 percent less costly for younger individuals.

Studying the benefits of migration in terms of increased consumption and wages, I find that both migration strategies have positive returns to the mover. However, the magnitude of these benefits depends strongly on the migration rationale: those who migrated to cope with negative income shocks benefit to a lesser extent than those who invested in migration. Predicted consumption increases by 8 percent after survival migration and by 35 percent after investment migration; comparable numbers for wage increases are 8 and 46 percent for survival and investment migration, respectively. Comparing individuals with various degrees of prior migration experience moreover suggests that the two migration strategies act as substitutes, meaning that those who migrate to cope with a negative shock are less likely to invest in migration.

Taken together, these findings may have important policy implications. Those with lower levels of wealth and education pay higher migration costs while earning less. In addition, they are more likely to engage in the type of migration that yields lower returns, which reduces the opportunity to invest in migration to the extent that the two strategies act as substitutes. This may have important distributional implications and resonates with a recent debate on the existence of geographical poverty traps. Jalan and Ravallion (2002) introduced this term, defining it as a situation in which the characteristics of a household's area of residence are such that the household's consumption cannot rise over time, while an otherwise identical household that lives in a better-endowed area would enjoy a rising standard of living. In a recent paper, Kraay and McKenzie (2014) survey the empirical evidence on poverty traps. While finding sparse evidence in support of poverty traps in general, they argue that geographical poverty traps form an exception, stating that the evidence most consistent with poverty traps comes from poor households in remote rural regions. While not specifically testing for the existence of poverty traps, I find that liquidity constraints prevent profitable migration (as also shown by Bryan, Chowdhury, and Mobarak (2014) and Bazzi (2014)) that poor individuals face higher migration costs and engage in less profitable migration, which may subsequently limit their chances of investing in migration.

A policy instrument that may mitigate these distributional challenges and promote profitable migration is the provision of credit. In my model environment, where part of the population faces liquidity and credit constraints, I examine a policy experiment of providing credit at various interest rates. I find that, on the one hand, credit reduces the need for survival migration, as it provides an alternative ex-post risk-coping strategy by allowing individuals to borrow in order to finance consumption. On the other hand, credit increases the use of investment migration by allowing individuals to borrow the up-front cost of migrating, thereby confirming that liquidity constraints initially prevented migration with positive expected returns.

The findings in this paper also have implications for the expected future impacts of climate change on migration. Weather patterns are expected to change due to global warming, and rainfall shocks will likely increase in intensity. This may adversely impact those living rural areas, for whom weather shocks are a major source of income variation. While there is still considerable uncertainty about the impact of climate change on migration, this paper addresses a piece of the puzzle by studying how individual migration choices respond to weather shocks. I run a counterfactual experiment to examine the predicted change in migration patterns and welfare in response to increased intensity of weather shocks. I find that more extreme weather shocks increase the need to engage in survival migration as an ex-post risk-coping strategy while simultaneously limiting the opportunity to save up for profitable investment migration. This leads to a predicted reduction in overall welfare and disproportionately affects those at the bottom of the wealth distribution.

This paper advances our understanding of what drives people to migrate, a question that has engaged development economists for decades (e.g. for early references: Lewis, 1954 and Harris and Todaro, 1970). Still, existing income differences between countries and areas within a country, combined with evidence of profitable returns to migration, have led people to wonder why more people do not migrate.[2] Moreover, empirical evidence shows that those who migrate for longer distances and duration tend to benefit to a larger extent, which has made people wonder why these migration patterns are not observed more frequently (e.g. Banerjee and Duflo, 2007 and Munshi and Rosenzweig, 2005).

By bringing together two often-cited and empirically observed migration strategies, this paper contributes to the understanding of why people migrate, where they migrate to, and how long they stay at their destination.[3] In an environment in which people face risk and liquidity constraints, I model these two strategies within a dynamic migration choice framework. The dynamics of the model allow for updating of preferred migration strategies in each period, making the model flexible by incorporating moves between various locations as well as multiple moves over time. As such, the model incorporates commonly observed migration patterns such as return migration and circular migration, which are not easily explained in models where people migrate merely in search of the best employment opportunity or models in which migration is treated as a one-shot decision. The importance of including multiple moves and a choice between multiple locations was also recognized by Kennan and Walker (2011), who develop a detailed dynamic model of optimal migration that explains migration choice based on expected income differentials in their data. There are considerable differences between their model and the model presented in this paper, primarily that Kennan and Walker (2011) consider a model in which wealth does not affect migration decisions. As such, individuals can borrow and lend without restriction to finance the cost of migration. This assumption may be warranted for their target group – young white males with a high school education in the United States – but has much less validity in the context of rural Indonesia. The model in this paper is therefore presented as an alternative model of migration choice applicable to developing country contexts in which wealth and liquidity constraints profoundly limit migration and destination choices.

This paper is structured as follows: First, I will present the dynamic migration choice model in Section 2. The data and empirical strategy are described in Section 3, and Section 4

[2] This question has been examined in the international context for example by Clemens, Montenegro, and Pritchett (2008) and McKenzie, Gibson, and Stillman (2010), and in the context of internal migration for example by Bryan, Chowdhury, and Mobarak (2014) and Beegle, De Weerdt and Dercon (2011).

[3] The optimal duration of migration has been explored by Dustmann and Kirchkamp (2002) in relation to return migration, see also Dustmann (1997), Dustmann (2003) and Dustmann and Weiss (2007).

provides reduced-form results. Section 5 introduces the structural estimation of the model, after which Section 6 presents the structural results. Various policy and counterfactual experiments are considered in Section 7, and Section 8 concludes.

2 Dynamic Migration Choice Model

This section develops a model incorporating both the survival and investment rationales for migration. This approach improves on previous models by allowing for multiple migration choices over time and between multiple locations, and incorporating wealth as an important determinant of migration choice. The model is dynamic in nature, to allow people to save up for migration and to acknowledge the forward-looking nature of migration choice. It extends the dynamic savings model from Deaton (1991) by adding location as an additional state and control variable. In Deaton's savings model, individuals are not permitted to borrow to finance consumption. The model has one state variable, wealth, and one control variable, consumption. In each period, the decision maker receives an income draw from a known distribution and chooses how much to consume and how much to save for the next period in order to maximize utility. As such, savings serve as a precautionary motive to smooth consumption and maximize lifetime utility.

Recently, Bryan, Chowdury and Mobarak (2014) developed a migration model that also builds on Deaton (1991) by incorporating liquidity constraints. In their model, migration is risky while individuals find out whether or not they are good at migrating. If they are not, they lose the cost of migrating; for those close to subsistence, this will lead to underinvestment in migration in order to avoid the cost of failed migration. As such, their model incorporates liquidity constraints that may be relaxed by a migration incentive, which they randomly distribute in villages in rural Bangladesh. Indeed, the 8.50 US dollar incentive induces 22 percent of households to send a migrant. While they find empirical evidence in support of their model, the large magnitude of their effects is not fully accounted for. As is common in migration choice models, they focus on the binary choice of whether to migrate. In order to incorporate different migration strategies, I also include the choice of which location to migrate to. Therefore, I extend Deaton's dynamic savings model by adding current location as a state variable and next location as an additional control variable. Unlike wealth and consumption, which are continuous variables, there is a finite number of discrete locations to choose from. Initially, I will set up the model in which locations are defined as a function of distance from a $Home$ location, which is defined as the location where the person lives at age 18. After presenting this general set-up, I will introduce a three-location model upon which the main predictions are based, and that will later be structurally estimated.

Migration is modeled as an individual decision but can alternatively be thought of as a household decision problem, in which in each period, the household chooses whether or not to send a household member to another location. By treating the household as one unit, intra-household transfers and remittances are not modeled explicitly. I focus on individual migration choices in order to reduce the computational time needed to numerically solve the model, without losing its main objective of incorporating survival and investment migration.

This is a partial equilibrium model and assumes that wages are exogenous to the individual decision maker, which matches the micro-level focus of the data. Wages are furthermore assumed to be stationary, so the model does not account for upward trends in wages. In the empirical analysis, all monetary values are converted to their year 2000 equivalent using the Indonesian consumer price index and time fixed effects are included to account for annual variation that is the same across individuals.

The timing of the model is as follows: In the beginning of each period, the individual is at a certain location l and is endowed with wealth x. Then, a wage draw w_l is revealed from a known distribution. The person chooses to either accept this wage draw or to migrate to another location with a known wage distribution, but where the wage draw has not yet been revealed. In case of the latter, the individual has to pay the up-front migration cost that is a function of the current and next location, and in particular, a function of the distance between them: $m(l, l^{'}) = f(d)$. I assume that migration costs increase monotonically with the distance traveled:

$$\frac{\partial m(l, l^{'})}{\partial d} > 0 \quad \text{with} \quad m(l, l^{'}) = 0 \quad \text{if} \quad l^{'} = l \tag{1}$$

In case the person decides to move, he or she first pays the migration costs, then moves to the next location $l^{'} \neq l$ and, upon arrival at $l^{'}$, observes the new wage draw $w_l^{'}$. I will refer to the final wage received as $w_l^{'}$, which is equal to the original wage draw if the person decided not to migrate, $w_l = w_l^{'}$, and will generally be different if the person migrated to a different location.

Finally, based on the wage received and current wealth, the person chooses consumption c in order to maximize utility U. At the end of the period, he or she is left with wealth $x^{'}$ and at location $l^{'}$, which are the starting values of the state variables in the next period. Note that the primes indicate the next period's values, so $l^{'} = l$ if the person stayed in the same location, and $l^{'} \neq l$ if he or she migrated.

The equation of motion describes the evolution of wealth:

$$x^{'} = (1 + r)(x - c - m(l, l^{'}) + w), \tag{2}$$

where r is the interest rate and w is the wage at the location the individual lives when receiving the wage. Similar to Deaton (1991), the liquidity constraint is modeled as a borrowing constraint:

$$x \geq 0 \tag{3}$$

This gives the following Bellman equation:

$$V(x, l) = \max_{c, l^{'}} \left\{ U(c, l^{'}) + \beta \int V(x^{'}, l^{'} dF(w_{l^{'}}) \right\} \tag{4}$$

In line with Deaton (1991) and Bryan, Chowdury and Mobarak (2014), an isoelastic utility function is chosen that exhibits constant relative risk aversion. In addition to consumption, utility is a function of the location chosen in each period. This input argument is added as a constant disutility y of being away from home to reflect that, ceteris paribus, individuals prefer to be at home, and also to avoid the unrealistic scenario where everyone would migrate.

$$U(c, l^{'}) = \frac{c^{1-\rho}}{1 - \rho} - y\mathbf{1}(l^{'}) \tag{5}$$

$$\text{with } \mathbf{1}(l') = 1 \text{ if } l' \neq Home \tag{6}$$

While in the basic set-up the disutility y is incurred every period throughout the duration of the stay, one of the model extensions defines the disutility relative to the person's previous location (not necessarily where he or she lived at age 18) to reflect that the perception of $Home$ changes over time as people settle at a new location.

I consider migration decisions in which individuals are given the opportunity to choose between multiple locations. Each of these locations is associated with a certain migration cost and independent wage distribution. I focus on migration decisions driven by economic rationales, such that people will not migrate to locations that are both more costly and provide lower wages, allowing me to consider only locations that are not dominated by both costs and wages. Thus, assuming that people are optimally migrating, paying higher costs of migration must be associated with larger wage gains. While in principle this model allows for any finite number of locations, in practice, the model becomes computationally unfeasible when using all distinct locations in the data.[4] Therefore, I will now turn to a simplified three-location model that is sufficient to provide the main predictions and intuition underlying the survival and investment migration strategies.

First, I define a $Home$ location as the location where a person resides at age 18. While moves at younger ages are observed in the data as well, these migration choices are likely made by the individual's parents. For the vast majority of individuals in the data, the $Home$ location is characterized as a rural location, so, throughout the model and empirical implementation, I restrict my focus to individuals whose $Home$ location is rural (though results are robust to including all $Home$ locations). We can then think of the migration decision as choosing between the best nearby rural area (with low migration costs, but wage draws that are not much better than at $Home$) or migrating to a further-away city with higher costs and higher wages. In the description of the model, I will therefore interchangeably use the nearby and rural location on the one hand and the far and urban location on the other hand, and all analyses will be carried out using both distinctions. As such, each person's location choice set consists of three entries: $\{H, R, U\}$, corresponding to $\{Home, Rural, Urban\}$, or alternatively, $\{H, N, F\}$, corresponding to $\{Home, Near, Far\}$.

I solve this three-location model numerically in discrete time with an infinite time horizon using value function iteration, following Miranda and Fackler (2002). More details on the model solution are given in the computational Appendix. While I also solve the model in finite time horizon using backward induction, the infinite time horizon is preferred because this lines up directly with the data I observe. In the finite time horizon model solution, individuals no longer migrate or save as the last period approaches, when the value is zero. In the panel data, I observe individuals during 20 years at various stages of their lives, so there is no equivalent of the final period, which makes the infinite time horizon model more appropriate.

The model solution is shown in the form of three model realizations in Figures 1, 2 and 3. The individuals depicted in each graph start off at $Home$ with wealth equal to 2, and the model is solved for each period to obtain the individual's optimal choices. For illustration purposes, $Rural$ wages are only slightly higher than wages at $Home$, while wages in $Urban$

[4]As described in the next section, there are 3,317 separate locations observed in the data and using all of these would take approximately 400 days to solve the model.

areas are significantly higher, as shown in Figure 4. While the wage distributions are certain and known to the decision maker, each wage draw is random. Figures 1, 2 and 3 give examples of an individual's behavior as predicted by the model under different wage draw trajectories during 20 time periods. Each figure shows cash-on-hand (in blue), consumption (in green), wage received (red squares) and original wage draw (grey crosses). In periods in which the individual does not migrate, the wage received (red squares) is equal to the wage draw in that period (grey crosses). When he or she migrates, however, the original wage draw in the starting location is usually not equal to the wage draw at the new location.

Figure 1 shows that wages follow a stochastic process over time and that cash-on-hand acts as a buffer to smooth consumption. Indeed, consumption is fairly constant during the first 13 periods. During this time period, wage draws are relatively good, allowing the individual to save and slowly increase his or her cash-on-hand up to the moment that, in period 14, wealth is high enough to cover migration costs to the urban area. As shown in the lower panel, the person moves from $Home$ to $Urban$ in period 14, where wages are higher, as shown earlier in Figure 4. The costs to cover this move are shown as a drop in cash-on-hand in the top panel. From period 14 onward, the individual indeed enjoys higher consumption and wages.

Figure 2 shows an individual with the same starting conditions but who is less fortunate with the wage draws he or she receives. In the first periods, poor wage draws prevent the individual from building up wealth. In period 6, cash-on-hand is not sufficient to buffer against the bad wage draw he or she receives at $Home$. As a result, the person would have to reduce consumption and therefore utility. To avoid this, he or she decides to migrate to a $Rural$ location in hope of receiving a better wage draw. Indeed, in period 6, the individual receives a better $Rural$ wage draw (red squares) than he or she would have received if he or she decided to stay at $Home$ (grey cross). As wealth remains low after migrating to a $Rural$ area, the situation reoccurs in period 10 and 11, and again in period 15. A person who stays at $Home$ throughout the 20 time periods is shown in Figure 3. He or she does not build up enough wealth to move to the $Urban$ location, nor does this person experience negative shocks that require migrating to a $Rural$ area.

These three examples of wage realization and resulting migration and consumption choices provide the basic intuition for the two migration motives often observed and studied. As in Figure 2, people may experience negative shocks, and, if they do not have sufficient wealth or savings to avoid the need to reduce consumption, migration can be used as an ex-post risk-coping strategy to allow the person to receive another random wage draw. To avoid high migration costs, it is preferred to migrate to a $Rural$ location under these circumstances. The long-term benefits of this strategy are limited because $Rural$ wages are only slightly higher than at $Home$ and being away from home comes at a utility cost y.

The investment potential of migration is illustrated in Figure 1. Migrating to the $Urban$ location is beneficial because wages are significantly higher than at $Home$, allowing migrants to increase consumption. However, if liquidity constrained, the individual may not be able to pay the up-front migration costs. As shown in Figure 1, individuals may be able to overcome liquidity constraints through positive wage draws, allowing them to build up wealth over time. If able to move, they would want to continue benefiting from higher wages, making this type of migration a longer-term strategy.

As shown by these realizations of the model solution, survival migration is used when individuals lack wealth to buffer against negative wage shocks. Investment migration is only possible after sufficient funds have been accumulated. It furthermore follows that migration to the *Rural* area is more likely to be of short duration, while there are incentives to stay longer after migration to an *Urban* area. As noted earlier, migration increases with contemporaneous negative shocks as well as with an accumulation of past positive shocks. Therefore, those who move in response to negative contemporaneous shocks stay at their destinations for shorter periods on average than those who move in response to an accumulation of previous positive shocks.

As individuals save up for migration further away or to urban areas, the migratory response to an accumulation of past positive shocks is predicted to be stronger for longer distances. The expected distance traveled after contemporaneous negative shocks is ambiguous. One the one hand, survival migration occurring after sudden negative shocks predicts short distance or rural migration to avoid high migration costs. On the other hand, conditional on having accumulated sufficient funds to invest in migration, it is still preferred to migrate when experiencing a negative shock that has reduced the opportunity costs of staying. So, while the migration response to an accumulation of positive shocks is expected to dominate for urban and faraway destinations, migration response after negative shocks is expected to occur at all destination types.

3 Data

I use the Indonesia Family Life Survey (IFLS) to study migration choice under risk and liquidity constraints and to test the migration choice model described in the previous section. Data was collected from the same households and individuals in four waves: 1993, 1997, 2000 and 2007. This panel dataset is particularly suitable to study migration due to its intensive efforts to track respondents and its resulting low rates of attrition: In the last wave in 2007, the recontact rate of original households interviewed in 1993 was 93.6 percent (Strauss et al., 2009 and Thomas et al., 2012). This longitudinal survey is representative of about 83 percent of the Indonesian population (Strauss et al., 2004). The analyses are based on all four waves of the IFLS, allowing me to construct a 20-year panel from 1988 through 2007 of 38,914 individuals.

3.1 Household panel dataset

Using the migration modules of the IFLS, a dataset is obtained of 38,914 individuals, who recorded when and where they migrated after the age of 12. All moves longer than 6 months are included. In addition to migration data based on recall between the four survey waves, the dataset contains information on where respondents were born and where they lived at age 12. This information is transformed into a panel dataset that reports the person's location in each year from 1988 to 2007. Children may move with their parents for reasons not included in the model, so for the main analyses I study people between age 18 and 65 and as noted earlier, a person's *Home* location is where he or she resides at age 18. As women move for marriage more often than men, robustness checks are performed for men only. To improve the balance of the panel, the year 1988 is the first year for all individuals in the panel, even though all moves after age 12 are recorded, including those before 1988 if the individual was already old enough. This results in a panel dataset of individual location decisions of 38,914 individuals age 18 and above during the period 1988 – 2007, with a total of 558,425 individual-year observations.

More than 99 percent of moves in the sample took place within the borders of Indonesia, so this study focuses primarily on internal migration. Location information is available at three geographical levels. The largest level is the province, of which there are 34 in Indonesia, and these are further divided into kabupaten (districts) and kecamatan (sub-districts). To be able to study all migration choices, including those over short distances, this study uses all three geographical levels. As such, a migrant is someone who resides in a kecamatan different from the one he or she lived in at age 18. There are 3,317 separate kecamatan observed in the data, each having corresponding latitude and longitude coordinates, making it possible to calculate all distances travelled between kecamatan, some of which are only short distances.

Figure 5 gives an example of the migration choices observed in the data. Each line represents an individual's move observed in the data, starting at a red dot and ending at a green dot. In total, more than 22,500 moves are observed in the data, so this map only shows a subset of the moves, namely those taking place in August of 1995. This map shows that a large share of the moves occur over short distances and within islands. Figure 6 illustrates this more clearly by using pie charts to show migration within and between islands. The colors of the pie chart correspond with the colors of the destination islands. For example, the pie chart for Sumatra in the west of Indonesia shows that, during the study period of 20 years, 1565 individual migrants originated from Sumatra. Of those, 49.7 percent migrated to the islands in the south (marked in darker colors), including Java and Bali, and another 49.3 percent migrated to destinations within Sumatra. The remaining one percent migrated from Sumatra to Kalimantan and Sulawesi in the north and north-east. This map shows that a large share of moves takes places within islands, which is especially true for the prosperous areas in Java and Bali, where more than 90 percent of individuals migrated within the island group. Figures 7 and 8 show that the large cities are popular destinations. 5.9 percent of individuals residing in the Java-Bali area migrate to Jakarta, the capital and largest city, at least once during the study period; 4.3 of those residing in Sumatra make a trip to Jakarta; and comparable numbers for Kalimantan and Sulawesi are 1.1 and 1 percent, respectively. Indonesia's third largest city, Medan, located in the north-west of Sumatra, attracts 2.8 percent of individuals residing in Sumatra but fewer people from island groups that are farther away from Medan.

Table 1 provides summary statistics of this dataset. In almost 38 percent of the individual-year pairs, the person does not reside in the kecamatan in which he or she lived at age 18, which defines the migrant stock. The migrant flow is lower at 4.37 percent, which includes only the individual-year pairs in which a person changed location. The majority of these moves were away from the location at age 18, defined as $Home$. The median move lasted 4 years and took place over a distance of 100 km. In 64 percent of the moves, individuals traveled by themselves, and, in the cases when they did move together, they traveled on average with 2.58 persons.

In addition to detailed information on migration, data are available on individual and household characteristics as well as labor market outcomes. Similar to the construction of the annual migration panel, an annual panel of individual income is created using recall data between the survey years. Income in both formal and informal sectors is included, as well as income from both main and side jobs. Although an imperfect measure, assets are used to approximate wealth. Asset data from the individual and household asset module are summed up and, following Haagenars et al. (1994), the adult equivalent of assets is used to create the individual-level wealth variable. Asset data is only collected during the survey years, so wealth observations are available for 1993, 1997, 2000 and 2007. To facilitate interpretation, all annual monetary values are reported in 100,000 Indonesian Rupiah and converted to their year 2000 equivalent, using the Indonesian consumer price index that is part of the International Financial Statistics collected by the International Monetary Fund.

3.2 Weather data

Weather data are obtained from the Center for Climatic Research at the University of Delaware (Matsuura and Willmott, 2009). Monthly estimates of precipitation and temperature are available for grids of 0.5 by 0.5 degree, which corresponds to about 50 by 50 kilometers in Indonesia. These data are based on interpolated weather station data and are matched to IFLS respondent locations using GPS coordinates. Figure 9 shows all individuals' locations on a map of Indonesia as red dots, with blue grids representing the weather data to which each location is mapped. While this study explores various weather measures, precipitation levels are used as the main weather variable. This is in line with Maccini and Yang (2009), who argue that rainfall is the most important source of weather variation in Indonesia. Temperature shows less variation over time due to Indonesia's equatorial location. Instead of using annual data from each calendar year, all measures are created from July until June of the following year to reflect the growing seasons in Indonesia. In addition to precipitation, this study carries out robustness checks with various other weather variables. This includes precipitation z-scores and temperature, as well as precipitation squared and cubed to allow for nonlinear effects. To capture unusual weather patterns, robustness checks are performed with deviations from the mean, precipitation growth and extreme weather events.

4 Empirical Results

In order to study the migratory response to weather shocks, I estimate the following equation:

$$migrate_{it} = \alpha + \beta weather_{iT} + \delta_t + \lambda_i + \varepsilon_{it} \tag{7}$$

$migrate_{it}$ is a dummy variable for whether person i migrates in year t. $weather_{iT}$ is the precipitation level at the location of individual i at time T, in which T can take various values. In case of contemporaneous shocks, precipitation at time t, $weather_{i,t}$, is used, while previous shocks are accumulated over preceding time periods, for instance from time period $t-1$ until $t-3$: $weather_{i,(t-1,t-3)}$. All regression analyses include time fixed effects, δ_t, and individual fixed effects, λ_i, and are clustered at the location level, which is the level at which weather shocks are observed. In order to justify the use of accumulated previous rainfall shocks to proxy for wealth accumulation, I first regress wealth on previous rainfall shocks (see Table 3). These results confirm that wealth increases in the presence of positive weather shocks in the current year t, as well as in previous years.

The main empirical results on the migratory response to contemporaneous and preceding weather shocks are presented in Table 4. This table shows that migration away from the individual's rural Home location increases in response to both negative contemporaneous shocks and to an accumulation of previous positive shocks. This is true for various sets of previous shocks, ranging from year $t-1$ in column 1 to year $t-5$ in column 4, and is highly significant whether analyzed in the same regression or separately (not shown).

This confirms that the two diverging migratory responses to income shocks, which have been studied as separate phenomena, can be observed in the same dataset. The magnitudes of these effects are economically meaningful. Average rainfall in Indonesia is about 150 mm per month. If the equivalent of one month of rain is missed in a particular year, this induces 2.25 percent of individuals to leave. Similarly, an extra month of rain annually in previous years induces about 1.5 percent of individuals to migrate. Compared to the average migrant flow of 4.37 percent, such changes in weather patterns would account for almost half of the moves observed in the data.

To further examine differences in migration patterns in response to income shock, Table 5 is split between moves that lasted less than the median duration of four years, and those that lasted longer. The choice of how long to stay at a new destination is an endogenous choice that may be affected by subsequent income shocks. In order to account for possible bias, duration is recorded in the year of migration, when subsequent employment outcomes and shocks were unknown. This table confirms that individuals who move in response to negative contemporaneous shocks stay at their destination for a shorter period than those who move in response to an accumulation of previous positive shocks. Comparing columns 1 and 2 indicates that people save up for migration that lasts longer than 4 years, while there is no evidence of savings accumulation for shorter moves.

Before comparing changes in migration patterns to rural and urban destinations induced by various weather shocks, Table 2 shows the transition matrix between the locations $Home$, $Rural$ and $Urban$. The first row shows that, in 62.26 percent of the individual-year pairs, a person lives at his or her $Home$ location and decides to stay there; in 1.25 percent of individual-year pairs, a person migrates from the $Home$ location to a $Rural$ location and, in 1.01 percent of pairs, he or she moves from $Home$ to an $Urban$ area. The diagonal shows that, on average, people stay at a $Rural$ destination in 17.82 percent of individual-year pairs and at an $Urban$ destination in 15.88 percent of individual-year pairs. Technically, the diagonal also includes the situation in which a person moves between $Rural$ destinations or between $Urban$ destinations, but these moves are uncommon. Summing up all off-diagonal matrix entries gives a migration flow of 4.04 percent. The total migration flow as reported in Table 1 is 4.37 percent, so the remaining 0.33 percent can be attributed to moves between $Rural$ areas or between $Urban$ areas. The bottom panel of Table 2 shows the absolute number of individual-year pairs in each matrix cell. These sum up to 558,425, the number of individual-year pairs observed in the data.

Table 6 compares the migratory response to weather shocks when migrating to a rural (column 1) versus an urban (column 2) destination. As described in Section 2, investment migration is more likely to have an urban destination, while survival is expected to induce individuals to migrate to a nearby rural location. The second row confirms that accumulation of wealth through preceding positive shocks dominates for urban destinations, to which migration costs are higher. As expected, migration after a negative shock encourages individuals to leave *Home* regardless of their destination, though individuals seem to be slightly more likely to move to a rural area. Table 7 repeats this exercise by comparing migratory responses to weather shocks when migrating less than 100 km (column 1) versus more than 100 km (column 2). As described in Section 2, investment migration is expected to dominate longer distance migration, while migration at all distances is expected to respond to current negative income shocks. This is confirmed by Table 7. An empirical challenge for comparing migration at various distances is that the distance itself is an endogenous choice. If there is positive serial correlation in rainfall patterns, a negative shock would tend to induce people to migrate farther away, while a positive shock would induce people to stay closer. However, the opposite pattern is observed in Table 7, so this is less of a concern, as it would merely bias the results downward.

5 Structural Estimation

In addition to the reduced-form evidence provided in the previous section, this section will structurally estimate the model in order to test its validity, estimate various model parameters and perform counterfactual policy analyses using these parameter values. The following table summarizes how the variables and parameters of the model match those observed in the data.

Location

Location is both a state and control variable in the model and, as described in Section 3, locations are defined at the kecamatan (sub-district) level, 3,317 of which are observed in the data. In line with the model, and in order to reduce computation time, the structural estimation distinguishes between three locations. *Home* is the kecamatan where the person lives at age 18, and in the basic version of the model, the definition of *Home* does not change over time. In accordance with the model and reduced-form results, two sets of criteria are used to distinguish migration destinations: *Rural* and *Urban* destinations on the one hand, and locations *Near* and *Far* on the other hand, which are defined as those less and more than 100 km away from *Home*. Robustness checks are performed with alternative distance cut-offs ranging from 50 to 200 km. Figure 10 confirms the model assumption that *Urban* wages first order stochastically dominate *Rural* wages and Figure 11 shows that the same is true for wages *Far* compared to wages *Near*. To account for any possible selection in migrant status, Figures 12 and 13 show that the first order stochastic dominance still holds when only including individuals who ever migrate.

Wealth

As described in Section 2, asset data are used to approximate wealth. These data are available only during the survey years (no recall data on assets were collected), so wealth is observed in the four survey years: 1993, 1997, 2000 and 2007. Because this is an infinite time horizon model in which each individual-year pair is treated equally, and given that each period uses only data from that period, the structural estimation will be restricted to the four years in which survey data was collected.

Wages

Model parameter	Symbol	Empirical variable
State and control variables		
Location	l	Location $\{H, R, U\}$
Wealth	x	Adult-equivalent of household assets
Wages		
Wage distribution at *Home*	μ_h	Mean income non-migrants at *Home*
	σ_h	Variance income non-migrants
Wage distribution *Rural*	μ_r, μ_n	Mean income migrants *Rural*, *Near*
	σ_r, σ_n	Variance income migrants *Rural*, *Near*
Wage distribution *Urban*	μ_u, μ_f	Mean income migrants *Urban*, *Far*
	σ_u, σ_f	Variance income migrants *Urban*, *Far*
Wage draw at time t	w_t	Predicted wage model
Exogenous parameters		
Discount factor	β	Set exogenously ranging from 0.90 - 0.99
Interest rate	r	Set exogenously ranging from 0.01 to 0.10
Structurally Estimate		
Cost migrating *Rural*	m_n	One-time cost of moving to *Rural*, *Near*
Cost migration *Urban*	m_f	One-time cost of moving to *Urban*, *Far*
Disutility away from *Home*	y	Constant disutility of being away from *Home*
Coeff relative risk aversion	ρ	Set exogenously or structurally estimate

While wages are observed for each individual in each year, these wages correspond to the wages at the location where the person chooses to be. For the model, the original wage draw is essential in determining whether or not a person migrates in response to a bad wage draw at the starting location that reduced the opportunity costs of moving. Original wage draws are not observed, however, and are likely lower than accepted wages if people indeed migrate away from negative shocks, making observed wages inadequate to use as wage shocks.

Instead, I use a wage model to predict original wage draws. Following Mincer (1974), I run basic Mincer regressions for the locations, $Home$, $Rural$, $Urban$, $Near$ and Far to predict income. In line with the reduced-form evidence presented in Section 4, I add current and lagged rainfall shocks to the common regressors, including education level, gender, age and age squared. Table 9 presents the results for all locations in columns 1, for $Home$, $Rural$ and $Urban$ in columns 2, 3, and 4, respectively, and columns 5 and 6 repeat the analyses for $Nearby$ and $Faraway$ destinations. As expected, those with higher levels of education earn higher wages, and this relationship is the strongest in $Urban$ and $Faraway$ areas. A similar pattern is observed for men compared to women. Income increases with age at a declining rate, as indicated by the negative squared term. Current and lagged precipitation are reported as z-scores to facilitate interpretation. In line with earlier reduced-form results, precipitation terms are positive and slowly reduce predictive power as precipitation from earlier time periods is used. Comparing the Mincer regression at $Home$ in column 2 to those in other locations (columns 3 to 6) reveals that the precipitation terms are predictive of income at $Home$, but to a much lesser extent in other locations. As the main source of exogenous variation, this limits the use of income shocks at locations away from $Home$. Therefore, while the model is flexible in allowing for moves in any direction, the main estimation will focus on structural parameters estimated using wage shocks at $Home$. Robustness checks are performed with a broader range of structural parameters and are consistent with the main results but computation time increases sharply with the number of parameters estimated.

Parameter values estimated structurally

The cost of migrating has a fixed and variable component. In order to finance a move, a one-time migration cost m_r needs to be paid to move to a $Rural$ location, and m_u needs to be paid to migrate to an $Urban$ area. Note that these migration costs include all one-time costs incurred when moving, such as transportation costs, as well as the cost of forgone income when employment is not immediately found. In addition to these one-time migration costs, individuals incur continuous costs, modeled directly in the utility function as a disutility of being away from home, y. While the migration cost and disutility of being away from home have different interpretations in the model, the main distinction in the structural estimation is provided by the difference in timing: migration cost is incurred only in the year of the move, while the disutility of being away from home is incurred as long as the individual is not present at $Home$. Both types of costs can be structurally estimated because people are observed in the year they move as well as in years they decide to stay at their destination.

5.1 Maximum likelihood estimation

I use maximum likelihood estimation in order to find the model parameter values underlying a series of simulated data that matches the observed data as closely as possible. For each set of parameter values, I solve the model, which leads to predicted choices for all state variable combinations.

Naturally, the predicted choices will not always correspond to actual choices I observe in the data. Following Rust (1987), I attribute deviations from predicted model decisions to unobserved state variables, ϵ, that are observed by the decision maker but unobserved by the econometrician. I assume that ϵ is distributed as a multivariate extreme value distribution, which leads to the logit formula as shown by Luce and Suppes (1965) and McFadden (1974). The conditional choice probabilities can then be expressed in the following closed form, in which x are the observed state variables and $d \in D$ the discrete decision variables:

$$P(d|x) = \frac{e^{(V(x,d))}}{\sum_{d' \in D(x)} e^{(V(x,d'))}} \tag{8}$$

I employ a nested fixed point algorithm that loops over various sets of parameter values in the outer loop. For a given set of parameter values, the inner loop solves the model by finding the value function of each location as a fixed point of the contraction mapping. Focusing on the discrete location choice consisting of the choice set, H, R, U, the following log likelihood function is calculated for each set of parameter values:

$$f = 1/N \sum_{i=1}^{N} log \left(\frac{e^{V_l}}{e^{V_h} + e^{V_r} + e^{V_u}} \right) \tag{9}$$

where N is the number of individual-year pairs in the data, which are all treated equally in the infinite time horizon model. The outer loop finds the set of parameter values that maximizes the log likelihood function. The computational Appendix describes the nested fixed point algorithm in more detail.

Migration costs likely vary across individuals. Therefore, after estimating the multinomial logit, I follow Berry, Levinsohn and Pakes (1995), Train (2003), and others by using a mixed logit model with random coefficients on migration costs. By applying a mixing distribution on the model parameters, I account for heterogeneity between individuals and exploit the panel structure of the data.

The choice probabilities for person i are

$$P(d|x)_{i\mathbf{j}} = \int \rho_{i\mathbf{j}}(\beta|\theta) f(\beta) d\beta, \tag{10}$$

where $\mathbf{j}$ is the sequence of choices $\mathbf{j} = \{j_{1,}j_2, j_3, j_4\}$ observed in the four survey years and $\rho_{i\mathbf{j}}$ is defined as before:

$$\rho_{i\mathbf{j}} = \frac{e^{(V_l^{'})_{i\mathbf{j}}}}{e^{V_h} + e^{V_r} + e^{V_u}} \tag{11}$$

and $f(\beta|\theta)$ is the mixing distribution, which I take to be normal to allow for the cost parameters θ to be either positive or negative. The probabilities are approximated through simulation for any given value of θ.

Following Train (2003), I use simulated log likelihood to estimate the mean and standard deviation of the mixing distribution. This accounts for differences in migration costs between individuals, keeping these constant for the same individual over time. As such, the mixed logit framework uses the panel structure in which individuals are observed up to four times across all survey years.

Table 10 shows how well the structurally estimated model fits the data in terms of predicted migration patterns. Comparing the observed migration patterns in the top panel to the predicted ones in the bottom panel reveals that the structurally estimated model closely predicts the actual location choices. The model does particularly well when $Home$ is the starting location, which is expected given that the exogenous weather shocks are most predictive of wages at the $Home$ location, as shown earlier in Table 9. Indeed, model predictions deviate further from the observed data when predicting return migration. The next section will therefore focus on structural estimation of the costs of migrating away from $Home$.

6 Structural Results

The main results of the structural estimation are presented in Table 11. The top panel shows the estimated costs of migrating to a rural versus an urban area and the bottom panel uses the distinction between migrating more or less than 100 kilometers. To facilitate interpretation, the monetary values are presented in units of 100,000 Indonesian Rupiah, converted to their equivalent values in the year 2000. Multiplying each number by 12 gives approximate comparable values in US dollars.

Column 1 gives the estimated migration costs for the full sample. The average migration cost of moving to a rural area is 11.09, which is about 133 US dollars. With an average annual income of 56, people have to spend about 20 percent of their annual income on average to move to another rural area. The estimated costs of moving to an urban area are considerable larger at 72.32, which is equivalent to 868 US dollars and is more than average annual income. The disutility of being away from $Home$ is denoted y in the utility function and presented as the equivalent amount of consumption that people are willing to forgo to maintain the same level of utility. At 7.53, the disutility away from $Home$ is about 90 US dollars and this amount is incurred every year the person is away from $Home$.

The bottom panel shows that the migration cost of moving to a $Near$ area is 15.86 (190 US dollars) and moving to a Far area is 69.82 (838 US dollars). Compared to the rural-urban division, the difference in migration costs is smaller between $Near$ and Far. The disutility cost is slightly higher at 8.74 (105 US dollars).

Columns 2 to 5 show migration cost estimates by wealth quartile and reveal that, in general, migration costs are considerably larger for those with less wealth, as approximated by the adult-equivalent of household assets. Note that this discrepancy, on top of the fact that those with lower levels of wealth earn lower income, makes it harder for those at the bottom of the wealth distribution to migrate to other locations. Average migration costs for the poorest quartile are 16.87 (202 US dollars) to move to a rural area and 95.95 (1150) to move to an urban area. Both costs are higher than for the general population in column 1, and the rural-urban difference is also greater, indicating that moving to an urban area is particularly costly for those at the bottom of the wealth distribution. Migration costs initially decrease with greater wealth, but are slightly higher again for those in the highest wealth quartile, which may be explained by the fact that those with very high levels of wealth in rural areas own large landholdings and migrate less often. The structural estimation will attribute this pattern to high migration costs for this subgroup.

6.1 Heterogeneous migration costs

Table 11 showed that migrating is considerably more costly for those with lower levels of wealth. This section re-estimates the model to study how migration costs differ across various subgroups of the society.

Age and education

Table 12 compares migration costs across education and age groups. Those with no education or only primary education are defined as having low education and those with any secondary education (even if they did not complete their degree) are classified as having high education. Structural estimates reveal that migrating is about twice as costly for those with lower levels of education. Once they arrive at the destination, however, it seems less costly to stay away from $Home$, because the disutility is smaller for the group with low education. Given the difference in timing used in the structural estimates, this indicates that, once they have migrated, people with lower levels of education are more likely to stay.

Columns 4, 5 and 6 compare age groups and reveal that migration costs increase rapidly with age. For those age 25 and below, migrating to a rural area cost only 6.44 (77 US dollar), and migrating to an urban area cost 20.43 (245 US dollar). For individuals above age 50, these costs are about 3 and 5 times higher for rural and urban areas, respectively.

Gender and marital status

Differences across gender and marital status are explored in Table 13. At first, it may seem surprising that migrating is about twice as costly for men as for women. In Indonesia, however, women are more likely to migrate for marriage, which increases the number of moves by women and leads to lower structurally estimated costs. Comparing migration costs for married versus single individuals indeed reveals that migration is almost four times as costly for those who are married. When conditioned on being married in columns 6 and 7, the gender difference in migration costs is considerably reduced.

Prior migration experience and connection at destination

Migration costs are likely to be lower for those with previous experience migrating, especially to the same or similar areas, as people have gained information and potentially established contacts at the destination. This is explored in Tables 14 and 15, though it should be noted that the results in these tables are merely correlations and cannot be interpreted causally. To facilitate comparison, columns 2 through 6 of Table 14 include only individuals currently at $Home$ after having gained migration experience, if any. Comparing columns 2 and 3 shows the somewhat surprising finding that urban migration costs are higher for those with some prior migration experience compared to no migration experience. This is further explained by comparing columns 4 and 5. Migrating to rural areas is less costly for those who migrated to rural areas before, while migration to urban areas is considerably more costly. The ratio of migration costs for those moving to urban versus rural areas is 6.5 in the general population (72.32/11.09) and increases to 10.3 for those with prior rural migration experience. This may suggest that those who used survival migration to a rural area are more likely to invest in urban migration subsequently. However, this can only be interpreted as a correlation, as low migration costs to rural areas may be precisely the reason these individuals migrated to rural areas before. The opposite pattern is observed for those with prior migration experience to urban areas, shown in column 5. The ratio of migration costs between urban and rural areas is merely 1.43 for this subgroup, indicating that those with prior migration experience to urban areas have much lower costs of migrating to an urban area again. In addition to the importance of migration experience, these results highlight the interaction between the two migration strategies. Those who have rural migration experience have higher urban migration costs and vice versa, which provides suggestive evidence that the migration strategies act as substitutes rather than complements.

This may have important policy implications. As shown earlier, those with lower levels of wealth and education pay higher migration costs while earning less. Moreover, the rural and nearby areas to which they are more likely to migrate yield lower returns to migration. To the extent that the migration strategies act as substitutes, this may further reduce their opportunity to investment in migration that would more likely improve their livelihoods. This may be one of the factors contributing to geographical poverty traps. As pointed out by Jalan and Ravallion (2002), geographical poverty traps exist when the characteristics of an area are such that household's consumption cannot rise while a similar household living in a geographically preferred area would enjoy rising standards of living. While migration can help people escape geographical poverty traps, the results in this study may indicate that the type of migration matters as well. If they engage in survival migration with low returns and low future opportunities to invest in migration, this may further perpetuate their disadvantageous rural position.

Finally, Table 15 compares migration costs between those who knew somebody at their destination prior to arrival and those who did not. This question was only included in the first survey wave of the IFLS, so this analysis is carried out for the year 1993 only. As a result of the small sample size, the standard errors are larger, but the pattern remains clear: those who report knowing someone at their destination have lower migration costs than those who did not. This is true for all four types of destinations considered, though the differences are not always statistically significant. Despite the small sample size, this result echoes the findings of earlier studies that emphasize the importance of migrant networks in determining migration choice.

6.2 Benefits of migration to the mover

The migration costs estimated in the previous section can be used to calculate predicted benefits of migration in terms of consumption and wage gains, as done in Table 16. Consumption changes on the top panel are those predicted by the model for the individual who moves and are relative to the model prediction if the person had not moved. Changes in annual consumption are calculated separately for those moving to a $Rural$ and an $Urban$ destination, and are comparable when using the distance metric of moving $Near$ and Far (results not shown). The first-year consumption changes are presented in the first two columns and show that those migrating to a $Rural$ area experience an average consumption increase of 2.04 percent and those who move to an $Urban$ area consume on average 3.62 percent more in the first year. Columns 3 and 4 repeat this analysis for annual consumption levels five years after the move and are not conditioned on whether the individual is still at the destination. Hence, although the person could have returned, the predicted consumption level is still observed five years after the move. These columns show that those who migrated to a *Rural* area have 5.82 percent higher annual consumption five years after moving; some of them may still be at the *Rural* destination. Those who migrate to an *Urban* area see a larger consumption gain at 30.71 percent, presumably because migrants tend to stay at their *Urban* destination for longer periods of time, enjoying higher wage and consumption levels.

The bottom panel of Table 16 calculates the change in wages for individuals who moved. Unlike consumption, the actual wages after moving are observed, and are used to estimate wage changes after migrating. The first two columns show that, in the year of migrating, wages increase by 8.36 percent when migrating to a rural area and by 38.44 percent when migrating to an urban area where wages are considerably higher. Wage gains after five years after 3.36 and 43.35 percent for rural and urban areas, respectively. Note that these wage gains are compared to predicted wages had the person stayed at *Home*. As shown earlier, migrants tend to move to rural areas when experiencing negative income shocks, which may explain why, when migrating to a rural area, the immediate wage gain is larger than the wage gain after five years.

7 Policy Experiments

The dynamic migration choice model with estimated migration cost parameters is used to explore predicted changes in welfare and migration, using counterfactual scenarios and policy experiments. First, I examine predicted changes in response to the provision of credit that allows people to borrow to fund migration and consumption. I then study the predicted effects of changes in migration costs, providing a subsidy to migrate, and restricting migration to urban areas. Finally, a counterfactual experiment of increased shock intensity is examined, in order to better understand the possible impact of climate change on migration.

7.1 Providing credit

The availability of credit may affect both migration strategies. Survival migration is used as a coping strategy after negative shocks to safeguard adequacy of consumption levels. When credit is available, people may prefer to borrow to guarantee sufficient consumption, thereby reducing the need for survival migration. Credit may furthermore relax liquidity constraints that prevented investment migration. So far, the model used here has followed Deaton (1991) by including a liquidity constraint stating that wealth needs to be weakly positive. In the numerical solution algorithm, any negative wealth value would results in zero consumption, causing utility and value functions to equal minus infinity, a value that will always be avoided by the decision maker. As such, an individual will be able to migrate only if he or she can cover the up-front migration cost. By providing credit, the policy maker can relax this constraint and allow people to borrow funds needed to invest in migration. Credit may therefore reduce the need for survival migration while increasing opportunities to engage in investment migration.

This section will explore changes in predicted migration rates and average consumption when providing credit at different rates and will distinguish between credit that can only be used to finance migration and credit provision for any purpose. All results are presented in panel A of Table 17.

The first row shows predicted effects of providing credit that can be used for any purpose, at an annual interest rate of 5 percent. As expected, survival migration is reduced considerably, from 1.25 percent as observed in the data to a predicted 0.87 percent. Note that these numbers refer to the migration flow, which counts only the year in which the move takes place. Also as expected, investment migration increases sharply from 1.01 to 1.61. As a results, columns 3 and 4 show that credit availability increases average consumption. The second row shows that, as the interest rate doubles, changes in migration rates and increases in consumption shrink disproportionately.

Various NGOs and development organizations have launched programs providing credit conditional on migrating. This policy experiment is examined in rows 3 and 4 at interest rates of 5 and 10 percent, respectively. Compared to unconditional credit provision, the impact is concentrated on investment migration because this is the type of migration where credit constraints are likely to be binding. The need for survival migration is only slightly reduced, which leads to lower welfare benefits, especially when interest rates are higher. Overall, these policy experiments reveal that credit acts as a substitute for survival migration and a complement to investment migration, and is welfare-enhancing overall.

7.2 Changes in migration costs

A number of policies can be used to directly affect migration costs, such as subsidizing migration. One such program was studied by Bryan, Chowdury and Mobarak (2014) and consisted of subsidies of 8.50 US dollars conditional on migrating. Panel B of Table 17 examines predicted changes in migration and consumption as a result of various changes in migration costs.

The first row shows that a 10 percent increase in migration costs reduces survival migration from 1.25 to 1.20 percent, while investment migration is reduced by a much greater degree, from 1.01 to 0.62 percent. This may be explained by a larger absolute change in $Urban$ migration costs, while the need to use survival migration remains large and consistent. Column 3 shows that average consumption increases initially as funds are used for consumption instead of for migration. However, the consumption effect turns negative in column 4 because fewer migration opportunities reduce long-term welfare, as measured by consumption.

The opposite pattern is observed for a 10 percent migration cost reduction in row 2 of Panel B. Survival migration increases from 1.25 to 1.35 percent, while investment migration sees a larger increase to 1.66 percent, after which there is more investment migration than survival migration. Consumption after 5 years increases significantly, though there is a small initial decrease as more people pay for migration.

Row 3 is a quantitative comparison with the experiment studied by Bryan, Chowdury and Mobarak (2014). The migration incentive of 8.50 US dollars has only a small effect on migration and consumption patterns. This stands in contrast to the 22 percent increase in migration rates they observed, but may be explained by the fact that, in addition to the price incentive, their intervention included information sessions that likely increased knowledge and awareness.

This migration choice model allows for estimating the welfare costs of restricting migration such as in the Chinese Hukou system. This system of household registration assigns rural or urban status to all citizens based on place of birth; if people move to another region, they lose access to various benefits and services, such as schooling. As such, the predicted migration rate to urban areas drops to zero. There is more need for survival migration as shown in column 1. Although there is a short period of increase from the lack of *Urban* migration, long-term consumption decreases by a considerable 21.3 percent.

7.3 Increases in weather shock intensity

While there is still considerable uncertainty about the impact of climate change on migration, this paper addresses a piece of the puzzle by studying how individual migration choices respond to weather shocks. Those living in rural areas in developing countries with limited asset holdings are often particularly susceptible to large income fluctuations. This is especially true for those working in agriculture, for whom weather shocks are a major source of income variation. Weather patterns are expected to change due to global warming, and rainfall shocks will likely increase in intensity.

I run a counterfactual experiment to examine the predicted change in migration patterns and welfare in response to increased intensity of weather shocks. Panel C of Table 17 predicts changes in migration and consumption in response to a 5 and 10 percent increase in the standard deviation of weather shock. The first row shows that a 5 percent increase in weather shock intensity increases survival migration by 0.12 percentage points from 1.25 to 1.35 percent, and reduces investment migration from 1.01 to 0.89 percent. The second row reports the predicted changes for the poorest half of the wealth distribution and reveals that changes in migration rates are almost twice as large as for the general population. This suggests that poor individuals carry the largest burden of increased need for survival migration and reduced opportunities for investment migration. Welfare effects, as approximated by changes in consumption, are negative at 3.19 percent on average and are more than twice as negative for the poorest half of the population. The last two rows repeat the counterfactual analysis for a 10 percent increase in the standard deviation of shocks and reveal that this increases survival migration from 1.25 to 1.74 percent and reduces investment migration from 1.01 to 0.74 percent. Adverse effects, especially in terms of the need for survival migration, are considerably larger for poor individuals. Overall reductions in welfare, as approximated by consumption changes, range from 8.63 percent on average to 15.43 percent for the poorest half of the wealth distribution.

Overall, these counterfactual experiments reveal that more extreme weather shocks increase the need to engage in survival migration as an ex-post risk-coping strategy while simultaneously limiting the opportunity to save up for profitable investment migration. This leads to a predicted reduction in overall welfare and disproportionately affects those at the bottom of the wealth distribution.

8 Conclusion

This paper studies migration choice in the face of risk and liquidity constraints. On the one hand, households can use migration as an ex-post risk-coping strategy by moving after sudden negative shocks, such as agricultural crop loss. On the other hand, migration can be seen as an investment, but liquidity constraints may prevent households from paying the up-front migration costs. While both migration strategies have been observed and described in the literature, they have diverging predictions in terms of the migratory response to shocks. In the case of survival migration, the occurrence of contemporaneous negative shocks may induce people to migrate, while, in the presence of liquidity constraints, an accumulation of preceding positive shocks may relax those constraints and increase out-migration.

This paper develops a dynamic migration choice model that incorporates both migration strategies. It builds on Deaton's (1991) savings model and adds current location as a state variable and migration choice as an additional control variable. Predictions are derived based on the types of shocks that induce migration, characteristics of the move – including distance and duration – and characteristics of those who migrate. The main contributions of the model are that it allows for multiple choices over time and between multiple locations, and that it incorporates wealth as an important determinant of migration choice. My approach goes beyond that of Kennan and Walker (2011), who, as noted above, did not include wealth constraints among an educated cohort of migrants in a developed country. The model in this paper is therefore presented as an alternative model of migration choice applicable to developing country contexts in which wealth and liquidity constraints profoundly limit migration and destination choices.

The model is tested using a rich panel of more than 38,000 individuals in Indonesia, for whom all migration choices were recorded over a 20-year period. I document evidence of both migration strategies. In agreement with the models predictions, I find empirically that survival migration is more often characterized by temporary moves to rural destinations and is used by those with low levels of wealth. Investment migration, on the other hand, is more likely to involve urban destinations, occur over longer distances, and be longer in duration.

I structurally estimate the model and find the model parameter values underlying a series of simulated data that match the observed data as closely as possible. Migration costs are structurally estimated and average about 20 percent of annual income for survival migration, while corresponding costs for investing in migration average slightly more than annual income, making it reasonable that people have to save to afford such moves. There is considerable heterogeneity in the cost to migrate, which is 30 percent higher for those with lower levels of wealth and education and 50 percent higher for individuals above the median age. Furthermore, costs are lower for women than men, which seems to be driven by migration for marriage, as gender differences decrease sharply when conditioned on being married.

While both migration strategies have positive returns to migrants, those who invest in migration benefit to a greater degree. Suggestive evidence moreover reveals that the two migration strategies act as substitutes, meaning that those who migrate to cope with a negative shock are less likely to invest in migration. This may have important distributional implications and contributes to the debate on geographical poverty traps. While not testing for poverty traps directly, I find that liquidity constraints prevent profitable migration, and that poor individuals face higher migration costs while engaging in less profitable migration, which may limit their chances of investing in migration subsequently.

A policy instrument that may mitigate these distributional challenges and promote profitable migration is credit provision. I use the structural estimates to perform policy experiments and find that providing credit reduces the need for survival migration and increases the opportunity to invest in migration.

Finally, I explore how changes in the intensity of weather shocks affect migration patterns, which has implications for predicted migratory responses to climate change. I find that more extreme weather shocks increase the need to engage in survival migration while limiting the opportunity to invest in migration. This leads to an overall reduction in welfare and disproportionately affects those at the bottom of the wealth distribution.

References

Abramitzky, R., L. Boustan, and K. Eriksson (2012) "Have the Poor Always Been Less Likely to Migrate? Evidence From Inheritance Practices During the Age of Mass Migration," NBER Working Papers 18298, National Bureau of Economic Research, Inc.

Angelucci, M. (2013) "Migration and Financial Constraints: Evidence from Mexico," IZA Discussion Papers 7726, Institute for the Study of Labor (IZA).

Banerjee, A. and E. Duflo (2006) "The Economic Lives of the Poor," CEPR Discussion Papers 5968, C.E.P.R. Discussion Papers.

Bazzi, S. (2014) "Wealth Heterogeneity and the Income Elasticity of Migration," *working paper*.

Bazzi, S., A. Gaduh, A. Rothenberg, and M. Wong (2014) "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," *working paper*.

Beegle, K., J. De Weerdt, and S. Dercon (2010) "Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey," CEPR Discussion Papers 7759, C.E.P.R. Discussion Papers.

Bell, M. and E. Charles-Edwards (2013) "Cross-national comparisons of internal migration: An update on global patterns and trends," *United Nations, Department of Economic and Social Affairs*, Vol. Technical Paper No. 2013/1.

Berry, S., J. Levinsohn, and A. Pakes (1995) "Automobile Prices in Market Equilibrium," *Econometrica*, Vol. 63, pp. 841–90.

Bohra-Mishra, P., M. Oppenheimer, and S. Hsiang (2014) "Nonlinear permanent migration response to climatic variations but minimal response to disasters," *PNAS*.

Borjas, G. (1990) "Self-Selection and the Earnings of Immigrants: Reply," *American Economic Review*, Vol. 80, pp. 305–08.

Boustan, L., P. Fishback, and S. Kantor (2010) "The Effect of Internal Migration on Local Labor Markets:American Cities during the Great Depression," *Journal of Labor Economics*, Vol. 28, pp. 719–746.

Boustan, L., M. Kahn, and P. Rhode (2012) "Moving to Higher Ground: Migration Response to Natural Disasters in the Early Twentieth Century," *American Economic Review*, Vol. 102, pp. 238–44.

Bryan, G., S. Chowdhury, and A. Mobarak (2014) "Under-investment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," NBER Working Papers 20172, National Bureau of Economic Research, Inc.

Card, D. and T. Lemieux (2001) "Can Falling Supply Explain The Rising Return To College For Younger Men? A Cohort-Based Analysis," *The Quarterly Journal of Economics*, Vol. 116, pp. 705–746.

Carrington, W., E. Detragiache, and T. Vishwanath (1996) "Migration with Endogenous Moving Costs," *American Economic Review*, Vol. 86, pp. 909–30.

Clemens, M. (2014) "Does Development Reduce Migration? - Working Paper 359," Working Papers 359, Center for Global Development.

Clemens, M., C. Montenegro, and L. Pritchett (2008) "The place premium : wage differences for identical workers across the US border," Policy Research Working Paper Series 4671, The World Bank.

Deaton, A. (1991) "Saving and Liquidity Constraints," *Econometrica*, Vol. 59, pp. 1221–48.

Deaton, A. and G. Laroque (1996) "Competitive Storage and Commodity Price Dynamics," *Journal of Political Economy*, Vol. 104, pp. 896–923.

Deb, P. and P. Seck (2009) "Internal Migration, Selection Bias and Human Development: Evidence from Indonesia and Mexico," Human Development Research Papers (2009 to present) HDRP-2009-31, Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

Dustmann, C. (1997) "Return migration, uncertainty and precautionary savings," *Journal of Development Economics*, Vol. 52, pp. 295–316.

——— (2003) "Return migration, wage differentials, and the optimal migration duration," *European Economic Review*, Vol. 47, pp. 353–369.

Dustmann, C. and O. Kirchkamp (2002) "The optimal migration duration and activity choice after re-migration," *Journal of Development Economics*, Vol. 67, pp. 351–372.

Dustmann, C. and Y. Weiss (2007) "Return Migration: Theory and Empirical Evidence," CReAM Discussion Paper Series 0702, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.

Frankenberg, E. and L. Karoly (1995) "The 1993 Indonesian Family Life Survey: Overview and Field Report.," *RAND, Santa Monica, CA.*

Frankenberg, E. and D. Thomas (2000) "The Indonesia Family Life Survey (IFLS): Study Design and Results from Waves 1 and 2," *RAND, Santa Monica, CA. DRU-2238/1-NIA/NICHD.*

Grogger, J. and G. Hanson (2011) "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, Vol. 95, pp. 42–57.

Hagenaars, A., K. de Vos, and M. Zaidi (1994) "PovertyStatistics in the Late 1980s: Research Based on Micro-data, Office for Official Publications of the European Communities.," *Luxemborg.*

Halliday, T. (2006) "Migration, Risk, and Liquidity Constraints in El Salvador," *Economic Development and Cultural Change*, Vol. 54, pp. 893–925.

Harris, J. and M. Todaro (1970) "Migration, Unemployment & Development: A Two-Sector Analysis," *American Economic Review*, Vol. 60, pp. 126–42.

Hicks, J. Hamory, M. Kleemans, and E. Miguel (2014) "Individual Ability and Selection into Migration in Kenya," MPRA Paper 19228, University Library of Munich, Germany.

Hosegood, V., A. Case, and C. Ardington (2009) "Labor Supply Responses to Large Social Transfers: Longitudinal Evidence from South Africa," *American Economic Journal: Applied Economics*, Vol. 1, pp. 22–48.

Jalan, J. and M. Ravallion (2002) "Geographic poverty traps? A micro model of consumption growth in rural China," *Journal of Applied Econometrics*, Vol. 17, pp. 329–346.

Jayachandran, S. (2006) "Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries," *Journal of Political Economy*, Vol. 114, pp. 538–575.

Keane, M. and K. Wolpin (1995) "The career decisions of young men," Working Papers 559, Federal Reserve Bank of Minneapolis.

——— (2007) "Exploring The Usefulness Of A Nonrandom Holdout Sample For Model Validation: Welfare Effects On Female Behavior," *International Economic Review*, Vol. 48, pp. 1351–1378.

Kennan, J. (2013) "Open Borders," *Review of Economic Dynamics*, Vol. 16, pp. L1–L13.

Kennan, J. and J. Walker (2011) "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, Vol. 79, pp. 211–251.

Kleemans, M. and J. Magruder (2014) "Labor market changes in response to immigration: evidence from internal migration driven by weather shocks," *working paper*.

Klein, P. and G. Ventura (2009) "Productivity differences and the dynamic effects of labor movements," *Journal of Monetary Economics*, Vol. 56, pp. 1059–1073.

Kraay, A. and D. McKenzie (2014) "Do Poverty Traps Exist? Assessing the Evidence," *Journal of Economic Perspectives*, Vol. 28, pp. 127–48.

Levine, D. and D. Yang (2006) "The Impact of Rainfall on Rice Output in Indonesian Districts," *Mimeo,University of California, Berkeley and University of Michigan.*

Lewis, A. (1954) "Economic Development with Unlimited Supplies of Labour," *The Manchester School*, Vol. 22, pp. 139–191.

Luce, D. and P. Suppes (1965) "Preferences, utility and subjective probability," *Handbook of Mathematical Psychology, John Wiley and Sons, New York*, pp. 249–410.

Maccini, S. and D. Yang (2009) "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall," *American Economic Review*, Vol. 99, pp. 1006–26.

Matsuura, K. and C. Willmott (2009) "Terrestrial Air Temperature and Precipitation: Monthly Climatologies," *Center for Climatic Research, Department of Geography, University of Delaware.*

McKenzie, D., J. Gibson, and S. Stillman (2010) "How Important Is Selection? Experimental vs. Non-Experimental Measures of the Income Gains from Migration," *Journal of the European Economic Association*, Vol. 8, pp. 913–945.

Mckenzie, D. and H. Rapoport (2007) "Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico," *Journal of Development Economics*, Vol. 84, pp. 1–24.

McKenzie, D., C. Theoharides, and D. Yang (2014) "Distortions in the International Migrant Labor Market: Evidence from Filipino Migration and Wage Responses to Destination Country Economic Shocks," *American Economic Journal: Applied Economics*, Vol. 6, pp. 49–75.

Miller, D., A. Cameron, and J. Gelbach (2006) "Bootstrap-Based Improvements for Inference with Clustered Errors," Working Papers 621, University of California, Davis, Department of Economics.

Mincer, J. (1974) *Schooling, Experience, and Earnings* in , NBER Books, No. minc74-1: National Bureau of Economic Research, Inc.

Miranda, M. and P. Fackler (2002) *Applied Computational Economics and Finance*: The MIT Press.

Morten, M. (2013) "Temporary Migration and Endogenous Risk Sharing in Village India," *working paper*.

Mueller, M., C. Gray, and K. Kosec (2014) "Heat stress increases long-term human migration in rural Pakistan," *Nature Climate Change*, Vol. 4, pp. 182–185.

Munshi, K. (2003) "Networks In The Modern Economy: Mexican Migrants In The U.S. Labor Market," *The Quarterly Journal of Economics*, Vol. 118, pp. 549–599.

Munshi, K. and M. Rosenzweig (2005) "Economic development and the decline of rural and urban community-based networks," *The Economics of Transition*, Vol. 13, pp. 427–443.

Olken, B. (2009) "Merged Indonesia Kecamatam Codes," *NBER electronic data set*.

Programme, United Nations Development (2009) *HDR 2009 - Overcoming barriers: Human mobility and development* in , Human Development Report (1990 to present), No. hdr2009: Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

Rosenzweig, M. and O. Stark (1989) "Consumption Smoothing, Migration, and Marriage: Evidence from Rural India," *Journal of Political Economy*, Vol. 97, pp. 905–26.

Rust, J. (1986) "Structural estimation of markov decision processes," in R. F. Engle and D. McFadden eds. *Handbook of Econometrics*, Vol. 4 of Handbook of Econometrics: Elsevier, Chap. 51, pp. 3081–3143.

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

Strauss, J., K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawati, and F. Witoelar (2004) "The Third Wave of the Indonesia Family Life Survey (IFLS): Overview and Field Report," *WR-144/1-NIA/NICHD*.

Strauss, J., D. Thomas, F. Witoelar, E. Frankenberg, B. Sikoki, C. Sumantri, and W. Suriastini (2010) "Cutting the costs of attrition: Results from the Indonesia Family Life Survey," Working Papers id:2652, eSocialSciences.

Strauss, J., F. Witoelar, B. Sikoki, and A. Wattie (2009) "The Fourth Wave of the Indonesian Family Life Survey (IFLS4): Overview and Field Report," *WR-675/1-NIA/NICHD*.

Thomas, D., E. Frankenberg, and J. Smith (2001) "Lost but Not Forgotten: Attrition and Follow-up in the Indonesia Family Life Survey," *Journal of Human Resources*, Vol. 36, pp. 556–592.

Todaro, M. (1969) "A Model for Labor Migration and Urban Unemployment in Less Developed Countries," *American Economic Review*, Vol. 59, pp. 138–48.

Townsend, M. (1994) "Risk and Insurance in Village India," *Econometrica*, Vol. 62, pp. 539–91.

Train, K. (2009) *Discrete Choice Methods with Simulation*, Cambridge Books: Cambridge University Press.

Weerdt, J. De and K. Hirvonen (2013) "Risk sharing and internal migration," Policy Research Working Paper Series 6429, The World Bank.

Wolpin, K. (2007) "Ex Ante Policy Evaluation, Structural Estimation and Model Selection," *American Economic Review*, Vol. 97, pp. 48–52.

Yang, D. (2006) "International Migration, Remittances, and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks," NBER Working Papers 12325, National Bureau of Economic Research, Inc.

——— (2008) "Risk, Migration, and Rural Financial Markets: Evidence from Earthquakes in El Salvador," *Social Research*, Vol. 75.

Figures

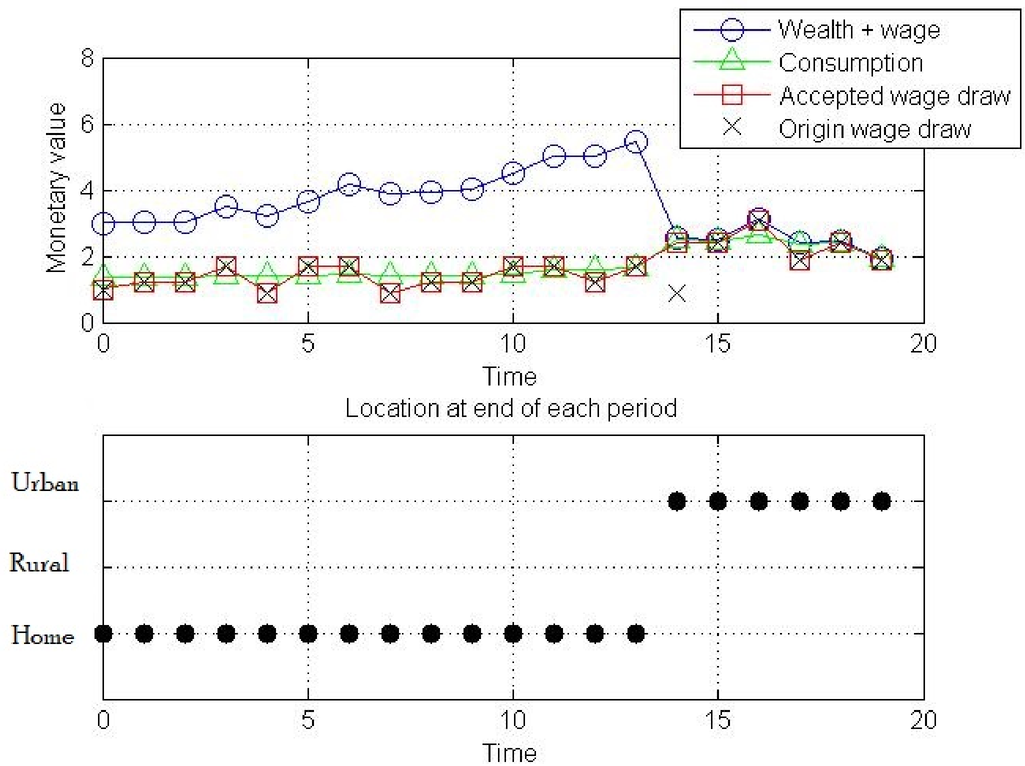


Figure 1: Model Solution: Moving to an Urban area

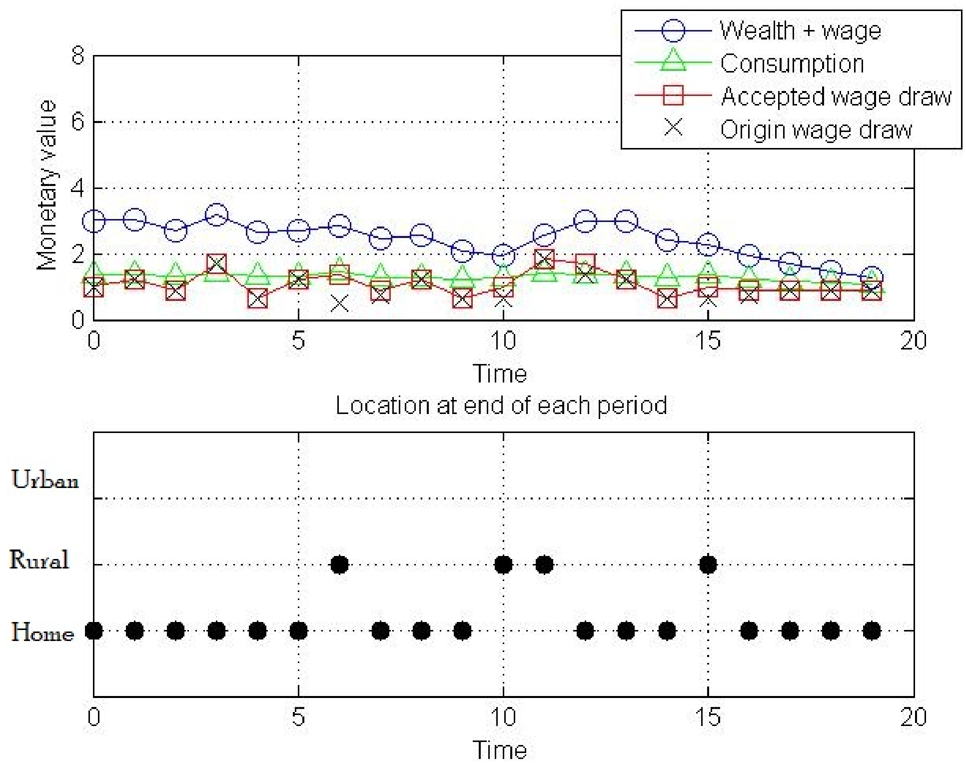


Figure 2: Model Solution: Moving to a Rural area

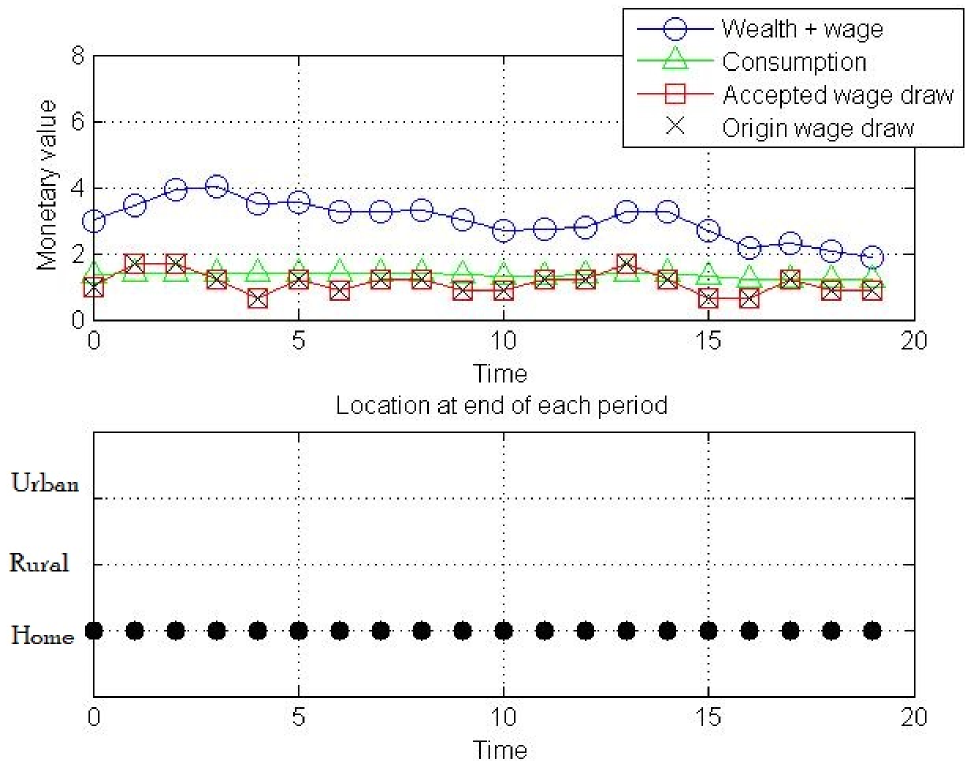


Figure 3: Model Solution: Staying Home

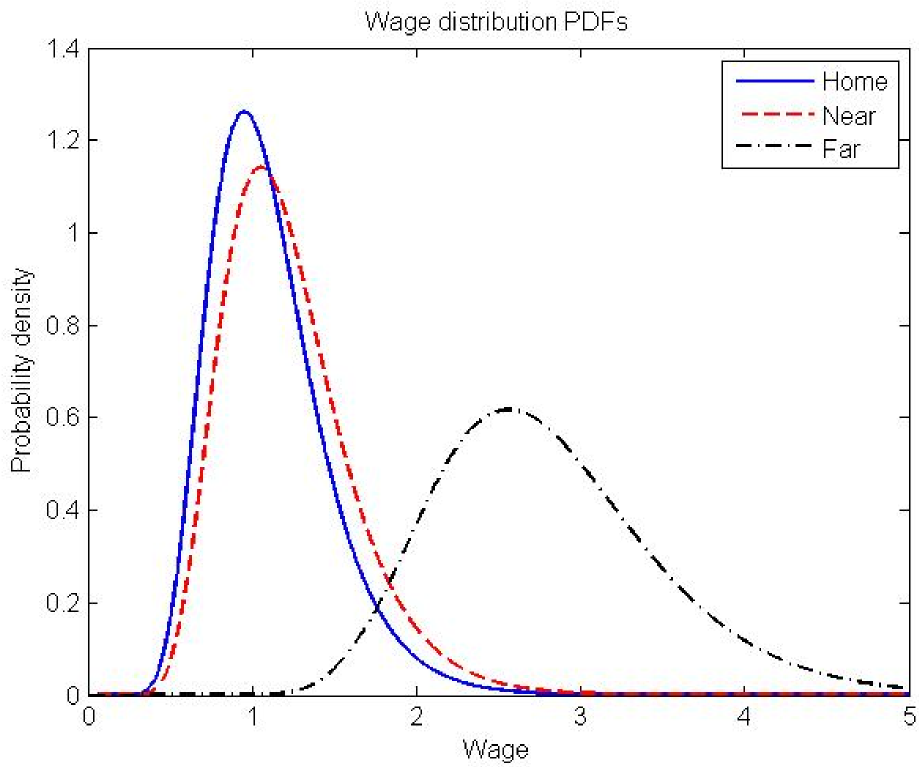


Figure 4: Wage Distributions used for model solutions shown in Figures 1, 2 and 3

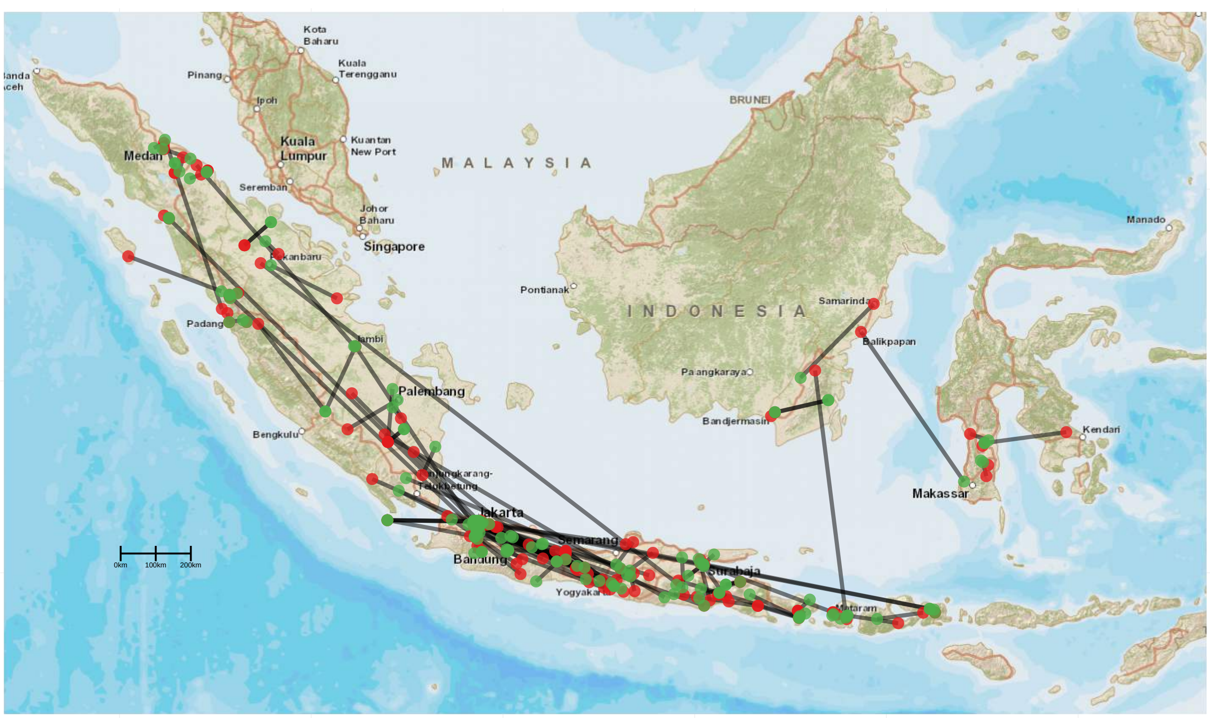


Figure 5: Monthly migration flow in August 1995

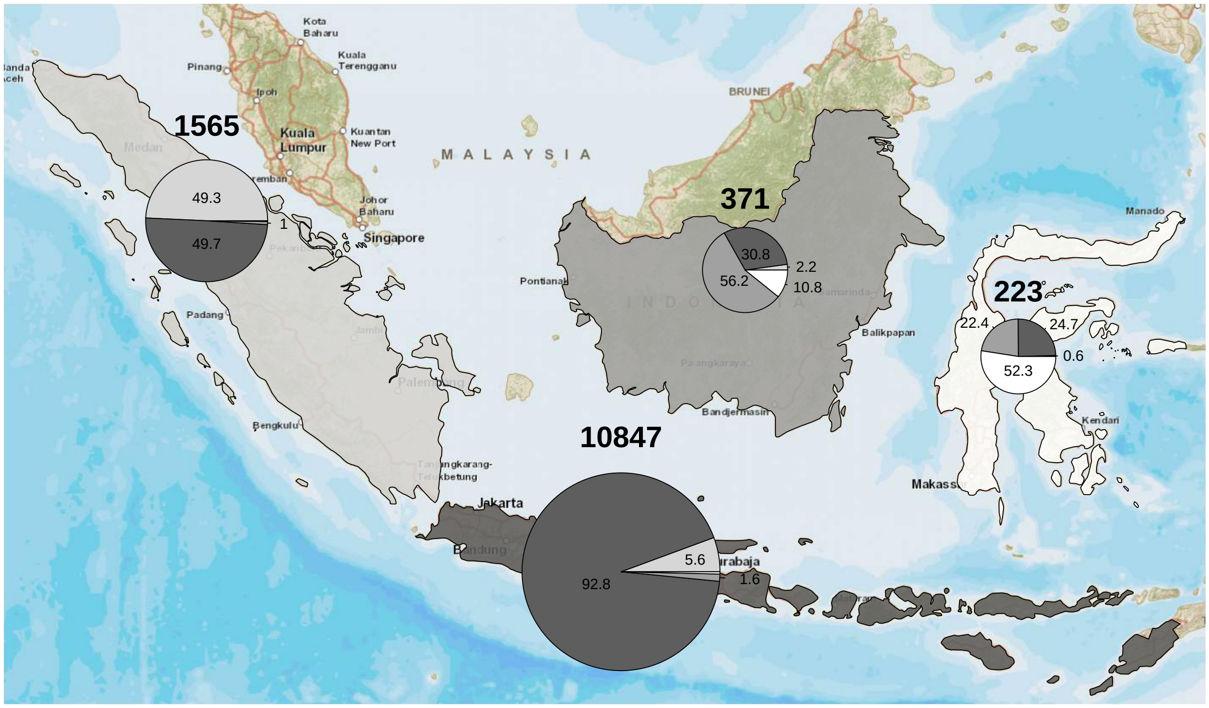


Figure 6: Migration flows between islands

Figure 7: Migration flow into Jakarta

Figure 8: Migration flow into Medan

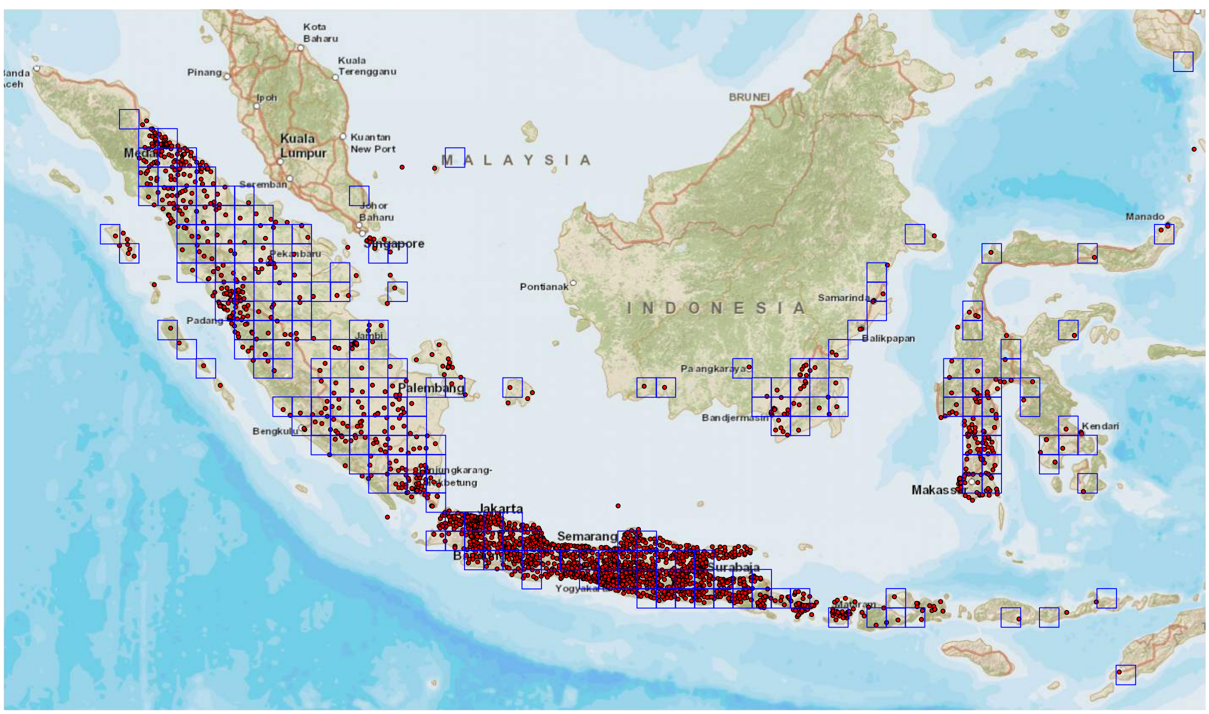


Figure 9: Household locations IFLS with weather data that locations are mapped to

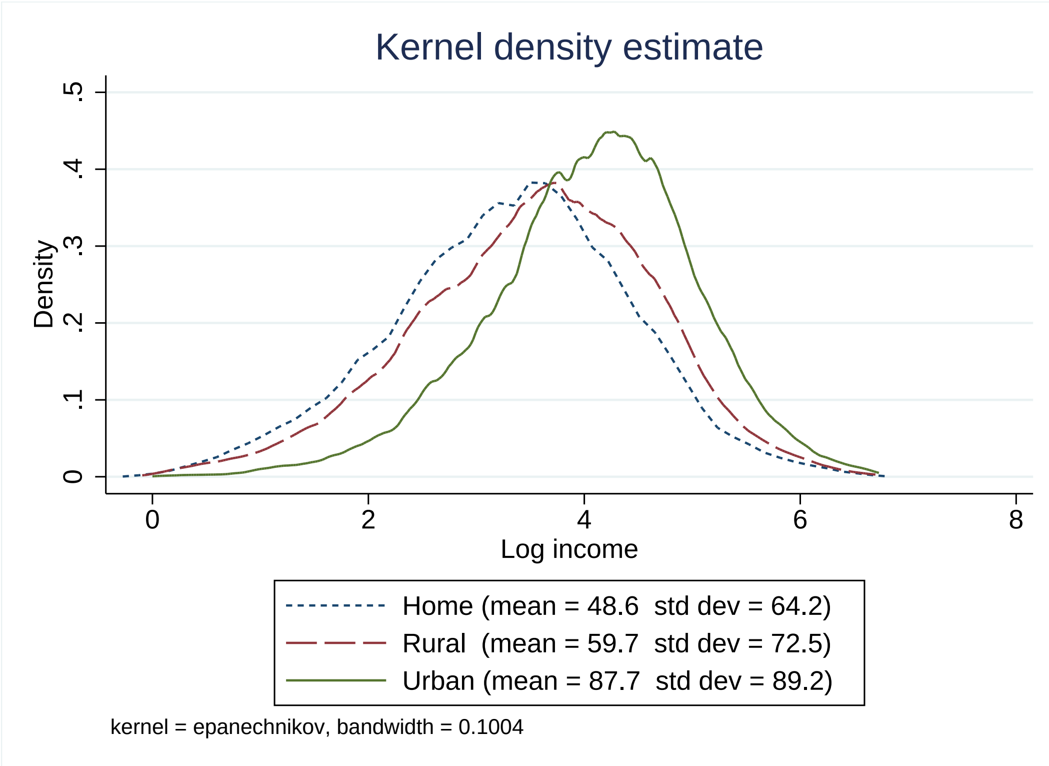


Figure 10: Kernel density wage distribution Home, Rural and Urban

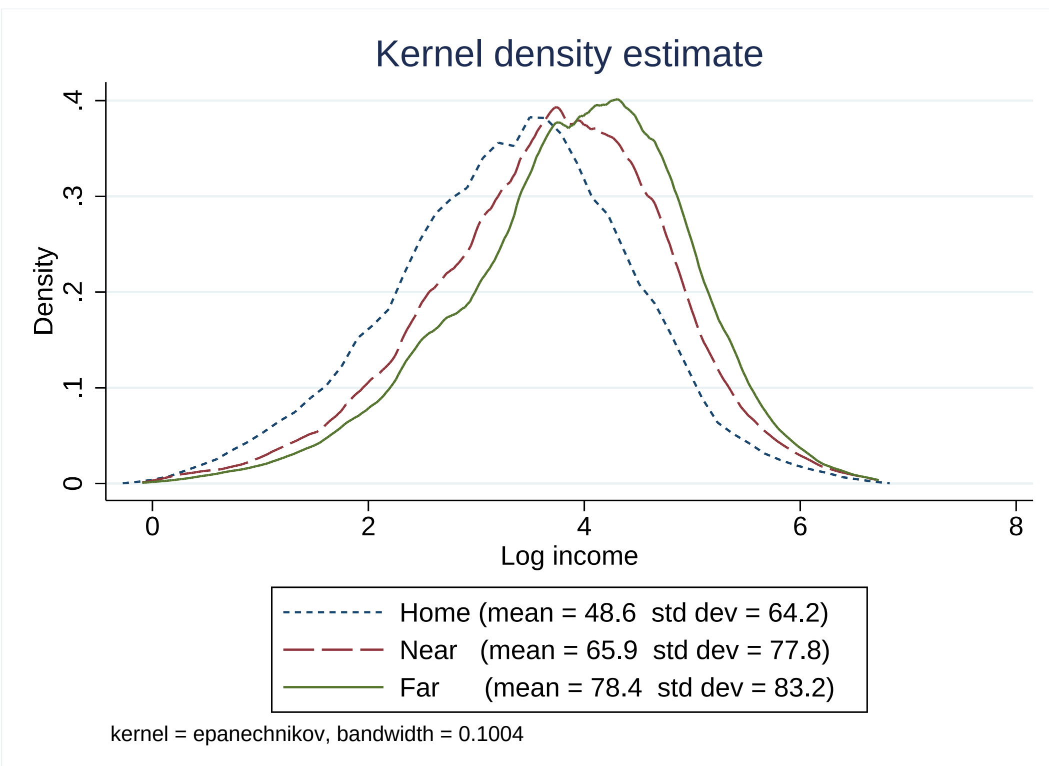


Figure 11: Kernel density wage distribution Home, Near and Far

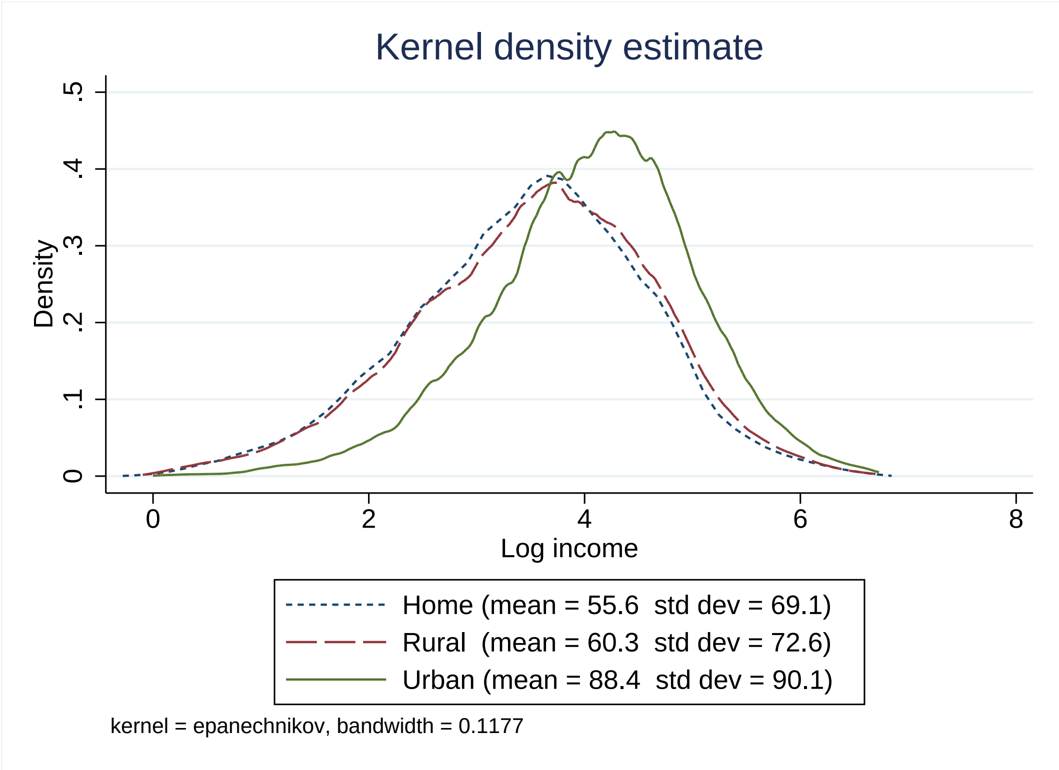


Figure 12: Kernel density wage distribution Home, Rural and Urban - Movers only

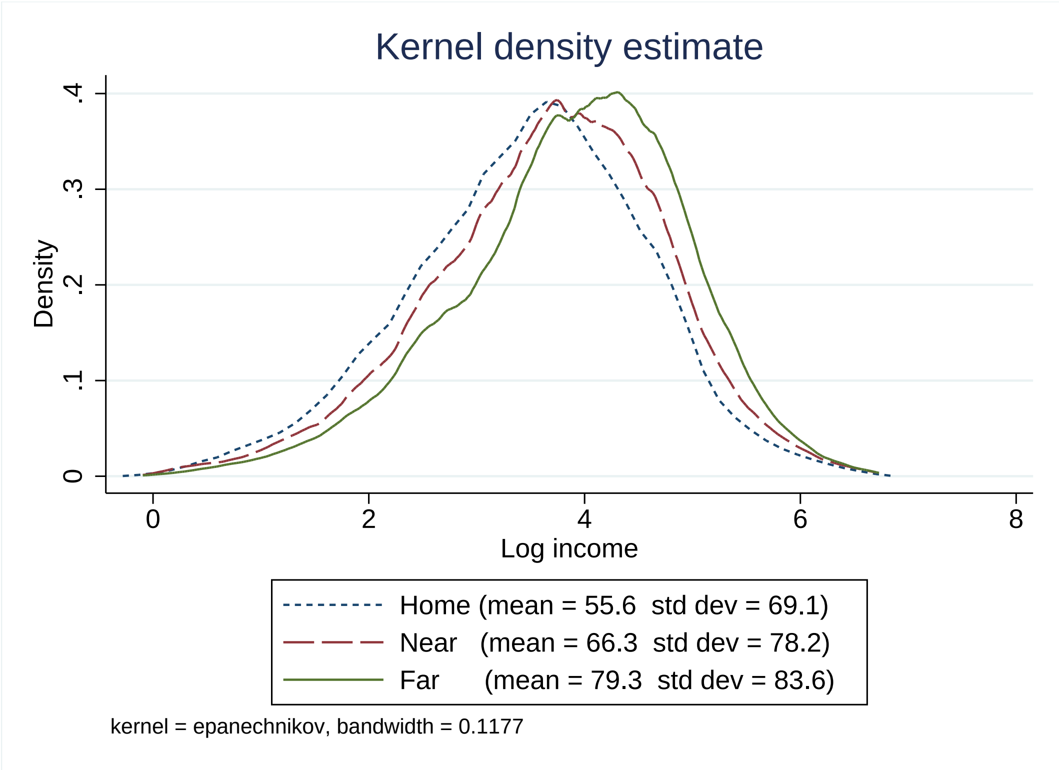


Figure 13: Kernel density wage distribution Home, Near and Far - Movers only

Tables

Table 1: Summary Statistics

Number of Individuals	38914
Number of Individuals-Year Pairs	558425
Migrant Stock (%)	37.74
Percentage of years as migrant	(48.65)
Migrant Flow (%)	4.37
Percentage of years migrating all directions	(20.44)
Duration Median (years away from home)	4.00
Mean	4.30
Standard Deviation	(4.12)
Distance Median (km away from home)	101.3
Mean	199.6
Standard Deviation	(304.4)
Migrating Together (%)	36.06
Of all moves, % with another person	(48.02)
Number of Persons	2.58
Conditional on moving together	(1.76)
Average Wealth	129.37
In 100,000 IDR ≈ 12 USD	(126.75)
Average Annual Income	56.33
In 100,000 IDR ≈ 12 USD	(71.55)

Source: Indonesian Family Life Survey. Means with standard deviations in brackets.

Table 2: Migration Transition Matrix

Percentages		End Location		
		Home	Rural	Urban
Starting Location	Home	62.26	1.25	1.01
	Rural	0.71	17.82	0.23
	Urban	0.55	0.28	15.88
Number of individual-year pairs		End Location		
		Home	Rural	Urban
Starting Location	Home	347675	7003	5623
	Rural	3962	99511	1290
	Urban	3071	1583	88678

Table 3: Wealth Accumulation in Response to Weather Shocks

	Dependent Variable: Individual Wealth			
	(1)	(2)	(3)	(4)
Rainfall at year t	0.642***	0.587***	0.590***	0.533***
	[0.181]	[0.182]	[0.183]	[0.186]
Sum rainfall year t-1 to t-2	1.220***			
	[0.129]			
Sum rainfall year t-1 to t-3		1.010***		
		[0.099]		
Sum rainfall year t-1 to t-4			1.043***	
			[0.086]	
Sum rainfall year t-1 to t-5				0.958***
				[0.075]
Time fixed effects	yes	yes	yes	yes
Individual fixed effects	yes	yes	yes	yes
Observations	99,379	96,416	93,046	89,737
R-squared	0.097	0.099	0.103	0.105
Number of pidlink	36,304	35,712	34,633	33,509

All regressions are clustered at the location level, standard errors in brackets, *** p<0.01, ** p< 0.05, * p<0.1. Rainfall is reported in average meter per month and wealth is reported in 100,000 (2000) Indonesian Rupiah ≈ $ 12 USD.

Table 4: Migration in Response to Weather Shocks

	Dependent Variable: Migrated away from Home Location			
	(1)	(2)	(3)	(4)
Rainfall at year t	-0.150***	-0.142***	-0.129***	-0.107***
	[0.014]	[0.014]	[0.014]	[0.014]
Sum rainfall year t-1 to t-2	0.042***			
	[0.009]			
Sum rainfall year t-1 to t-3		0.034***		
		[0.007]		
Sum rainfall year t-1 to t-4			0.030***	
			[0.006]	
Sum rainfall year t-1 to t-5				0.026***
				[0.005]
Time fixed effects	yes	yes	yes	yes
Individual fixed effects	yes	yes	yes	yes
Observations	354,320	354,320	354,320	354,320
R-squared	0.002	0.002	0.002	0.002
Number of pidlink	35,522	35,522	35,522	35,522

All regressions are clustered at the location level, standard errors in brackets, *** p<0.01, ** p< 0.05, * p<0.1. The dependent variable is a dummy of having migrated away from the individual's rural home location, where he/she lived at age 18; rainfall is reported in average meter per month (in Indonesia, average monthly rainfall equals 150 mm).

Table 5: Migration Strategies by Duration

	Dependent variable: Migrated away from Home Location	
	0 - 4 years	**> 4 years**
	(1)	(2)
Rainfall at year t	-0.096***	-0.046***
	[0.011]	[0.009]
Rainfall at year t-1 to t-3	0.003	0.031***
	[0.005]	[0.005]
Time fixed effects	yes	yes
Individual fixed effects	yes	yes
Observations	354,320	354,320
R-squared	0.003	0.003
Number of pidlink	35,522	35,522

All regressions are clustered at the location level, standard errors in brackets, *** p<0.01, ** p< 0.05, * p<0.1. The dependent variable is a dummy of having migrated away from the individual's rural home location, where he/she lived at age 18; rainfall is reported in average meter per month (in Indonesia, average monthly rainfall equals 150 mm).

Table 6: Migration Strategies by Destination

	Dependent Variable: Migrated away from Home Location	
	Rural Destination	**Urban Destination**
	(1)	(2)
Rainfall at year t	-0.081***	-0.061***
	[0.009]	[0.011]
Rainfall at year t-1 to t-3	-0.002	0.036***
	[0.005]	[0.005]
Time fixed effects	yes	yes
Individual fixed effects	yes	yes
Observations	354,320	354,320
R-squared	0.001	0.002
Number of pidlink	35,522	35,522

All regressions are clustered at the location level, standard errors in brackets, *** p<0.01, ** p< 0.05, * p<0.1. The dependent variable is a dummy of having migrated away from the individual's rural home location, where he/she lived at age 18; rainfall is reported in average meter per month (in Indonesia, average monthly rainfall equals 150 mm).

Table 7: Migration Strategies by Distance

	Dependent variable: Migrated away from Home Location	
	0 - 100 km	**> 100 km**
	(1)	(2)
Rainfall at year t	-0.037***	-0.106***
	[0.009]	[0.011]
Rainfall at year t-1 to t-3	-0.002	0.036***
	[0.005]	[0.005]
Time fixed effects	yes	yes
Individual fixed effects	yes	yes
Observations	354,320	354,320
R-squared	0.001	0.002
Number of pidlink	35,522	35,522

All regressions are clustered at the location level, standard errors in brackets, *** p<0.01, ** p< 0.05, * p<0.1. The dependent variable is a dummy of having migrated away from the individual's rural home location, where he/she lived at age 18; rainfall is reported in average meter per month (in Indonesia, average monthly rainfall equals 150 mm).

Table 8: Migration Strategies by Initial Wealth

	Dependent variable: Migrated away from Home Location
	(1)
Rainfall t * Intial Wealth 0-33%	-0.141***
	[0.036]
Rainfall t * Intial Wealth 34-66%	-0.122***
	[0.033]
Rainfall t * Intial Wealth 67-100%	-0.052*
	[0.035]
Rainfall (t-1 to t-3) * Intial Wealth 0-33%	0.050***
	[0.021]
Rainfall (t-1 to t-3) * Intial Wealth 34-66%	0.071***
	[0.020]
Rainfall (t-1 to t-3) * Intial Wealth 67-100%	0.045**
	[0.020]
Initial Wealth 34-66%	-0.000
	[0.012]
Initial Wealth 67-100%	-0.000
	[0.013]
Time fixed effects	yes
Individual fixed effects	yes
Observations	87,227
R-squared	0.005
Number of pidlink	34,537

All regressions are clustered at the location level, standard errors in brackets, *** p¡0.01, ** p¡ 0.05, * p¡0.1. The dependent variable is a dummy of having migrated away from the individual's rural home location, where he/she lived at age 18; rainfall is reported in average meter per month (in Indonesia, average montly rainfall equals 150 mm); initial wealth is based on the first wealth measurement of the individual and only years after this measurement are included.

Table 9: Wage Model

	Dependent Variable: Income					
	All	**Home**	**Rural**	**Urban**	**Near**	**Far**
	(1)	(2)	(3)	(4)	(5)	(6)
Education Level	16.50***	13.89***	16.32***	22.41***	17.69***	19.78***
	[0.57]	[0.60]	[0.937]	[1.560]	[1.08]	[1.68]
Male	18.95***	18.32***	17.62***	24.04***	20.39***	21.64***
	[0.81]	[0.96]	[1.792]	[2.324]	[1.83]	[2.46]
Age	4.21***	3.55***	3.95***	6.24***	4.58***	5.58***
	[0.18]	[0.19]	[0.374]	[0.429]	[0.40]	[0.48]
Age squared	-0.04***	-0.04***	-0.04***	-0.06***	-0.05***	-0.06***
	[0.002]	[0.002]	[0.005]	[0.005]	[0.005]	[0.006]
Rainfall at year t	1.97***	2.43***	0.65	0.08	1.27	0.63
	[0.43]	[0.50]	[0.84]	[1.15]	[0.81]	[1.10]
Rainfall at year t-1	1.67***	1.56***	1.30	1.26	1.59*	1.41
	[0.43]	[0.45]	[0.93]	[1.01]	[0.83]	[1.03]
Rainfall at year t-2	1.18***	1.58***	-0.84	0.60	0.31	0.39
	[0.45]	[0.48]	[1.01]	[1.05]	[0.89]	[0.98]
Rainfall at year t-3	0.51	0.81**	-2.08**	1.81	-0.86	1.35
	[0.39]	[0.36]	[1.05]	[1.17]	[0.84]	[1.45]
Rainfall at year t-4	0.72	0.79*	-0.85	0.80	0.78	-1.11
	[0.44]	[0.46]	[0.81]	[1.43]	[0.78]	[1.36]
Mean dependent variable	56.33	48.67	59.69	87.67	65.92	78.41
	(71.55)	(64.24)	(72.52)	(82.23)	(77.84)	(83.21)
Time fixed effects	yes	yes	yes	yes	yes	yes
Location fixed effects	yes	yes	yes	yes	yes	yes
Observations	154,179	94,499	35,129	24,551	34,826	23,208
R-squared	0.121	0.104	0.109	0.173	0.123	0.143
Number of locations	2,177	1,189	1,421	556	1,484	1,219

All regressions are clustered at the kecamatan level, standard errors in brackets, *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Income is reported in 100,000 (2000) Indonesian Rupiah ≈ $ 12 USD. 'Home' is the person's location at age 18; 'Rural' refers to rural destinations; 'Urban' refers to urban destinations; 'Near' refers to destinations within 100 km from 'Home'; 'Far refers to destinations farther than 100 km from 'Home'.

Table 10: Model Fit

Observed Data		End Location		
		Home	Rural	Urban
Starting Location	Home	62.26	1.25	1.01
	Rural	0.71	17.82	0.23
	Urban	0.55	0.28	15.88
Model Predicton		End Location		
		Home	Rural	Urban
Starting Location	Home	62.28	1.22	1.03
	Rural	1.36	17.42	0.18
	Urban	0.39	0.00	16.11

Table 11: Mixed Logit Estimation of Migration Costs by Wealth Quartile

		Wealth Quartile			
	All	**0-25%**	**25-50%**	**50-75%**	**75-100%**
	(1)	(2)	(3)	(4)	(5)
Migration cost Rural	11.09	16.87	12.61	12.99	13.22
	(0.34)	(0.41)	(0.52)	(0.51)	(0.58)
Migration cost Urban	72.32	95.95	71.61	56.06	60.24
	(1.26)	(1.64)	(2.07)	(2.03)	(2.34)
Disutility away from home	7.53	7.14	9.06	10.21	11.62
(Consumption Equivalent)	(0.48)	(0.63)	(0.68)	(0.60)	(0.80)
Number of observations	100643	23122	23101	23095	23140
Migration cost Near	15.86	17.21	16.65	10.32	14.68
	(0.29)	(0.63)	(0.67)	(0.60)	(0.64)
Migration cost Far	69.82	92.39	70.84	54.81	59.15
	(1.12)	(1.83)	(2.27)	(2.18)	(2.53)
Disutility away from home	8.74	6.17	9.20	9.33	11.25
(Consumption Equivalent)	(0.49)	(0.60)	(0.79)	(0.63)	(0.86)
Number of observations	99589	22880	22955	22946	23001

Average wealth is 129 (std dev: 127) and average annual income is 56 (std dev: 72). All values are reported in 100,000 (2000) Indonesian Rupiah ≈ $ 12 USD. Standard errors in brackets are not corrected for predicted wages. 'Rural' refers to rural destinations; 'Urban' refers to urban destinations; 'Near' refers to destinations within 100 km from the individual's home location at age 18; 'Far' refers to destinations farther than 100 km from the individual's home location at age 18. All parameters are estimated using mixed logit, which includes estimating standard deviations of all parameter values and of the classical measurement error on wealth. The following parameter values are set exogenously $\beta = .0.95$ $r = 0.03$ $\rho = 2$

Table 12: Mixed Logit Estimation of Migration Costs by Education Level and Age

		Education level		Age		
	All	**Low**	**High**	**25-**	**25-50**	**50+**
	(1)	(2)	(3)	(4)	(5)	(6)
Migration cost Rural	11.09	16.58	13.76	6.44	12.07	20.51
	(0.34)	(0.80)	(0.53)	(0.67)	(0.56)	(0.75)
Migration cost Urban	72.32	127.33	62.06	20.43	73.34	120.27
	(1.26)	(1.12)	(0.58)	(0.92)	(0.95)	(1.18)
Disutility away from home	7.53	6.80	10.52	11.13	9.44	7.70
(Consumption Equivalent)	(0.48)	(0.60)	(0.80)	(0.67)	(0.62)	(0.87)
Number of observations	100643	47585	44779	26737	46231	19425
Migration cost Near	15.86	21.16	13.83	7.40	14.19	23.61
	(0.29)	(0.69)	(0.48)	(0.63)	(0.46)	(0.67)
Migration cost Far	69.82	92.63	63.53	17.13	73.42	116.59
	(1.12)	(1.08)	(0.59)	(0.97)	(0.96)	(2.34)
Disutility away from home	8.74	8.82	11.05	12.17	8.89	7.05
(Consumption Equivalent)	(0.49)	(0.79)	(0.96)	(0.62)	(0.59)	(0.70)
Number of observations	99589	46985	44444	26539	45717	19197

Average wealth is 129 (std dev: 127) and average annual income is 56 (std dev: 72). All values are reported in 100,000 (2000) Indonesian Rupiah $\approx$ \$ 12 USD. Standard errors in brackets are not corrected for predicted wages. 'Rural' refers to rural destinations; 'Urban' refers to urban destinations; 'Near' refers to destinations within 100 km from the individual's home location at age 18; 'Far' refers to destinations farther than 100 km from the individual's home location at age 18. All parameters are estimated using mixed logit, which includes estimating standard deviations of all parameter values and of the classical measurement error on wealth. The following parameter values are set exogenously $\beta = .0.95$ $r = 0.03$ $\rho = 2$

Table 13: Mixed Logit Estimation of Migration Costs by Gender and Marital Status

	All	**Male**	**Female**	**Married**	**Single**	**Married Male**	**Married Female**	**Single Male**	**Single Female**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Migration cost Rural	11.09	16.67	10.65	23.86	7.45	48.94	30.85	13.61	6.05
	(0.34)	(0.46)	(0.72)	(0.66)	(0.69)	(0.83)	(0.74)	(0.93)	(0.65)
Migration cost Urban	72.32	97.78	43.82	114.82	34.07	133.04	115.22	48.45	24.44
	(1.26)	(1.02)	(0.99)	(1.72)	(0.86)	(2.46)	(1.62)	(1.92)	(1.44)
Disutility away from home	7.53	12.46	5.33	8.32	10.00	10.70	-3.25	12.03	12.83
(Consumption Equivalent)	(0.48)	(0.88)	(0.98)	(0.62)	(0.77)	(0.71)	(0.91)	(0.60)	(0.88)
Number of observations	100643	44158	48244	33902	22875	16487	17393	12874	10034
Migration cost Near	15.86	20.44	11.55	24.77	6.95	28.94	19.23	12.65	5.34
	(0.29)	(0.45)	(0.63)	(0.76)	(0.95)	(0.73)	(0.85)	(0.93)	(0.86)
Migration cost Far	69.82	94.69	45.76	110.90	36.87	123.50	79.20	54.68	32.82
	(1.12)	(0.97)	(0.76)	(1.42)	(0.86)	(2.26)	(1.68)	(1.77)	(0.99)
Disutility away from home	8.74	11.71	2.65	5.06	10.77	9.79	-6.35	10.97	11.21
(Consumption Equivalent)	(0.49)	(0.90)	(0.42)	(0.88)	(0.69)	(0.94)	(0.96)	(0.84)	(0.86)
Number of observations	99589	43694	47770	33595	22698	16333	17240	12776	9955

Average wealth is 129 (std dev: 127) and average annual income is 56 (std dev: 72). All values are reported in 100,000 (2000) Indonesian Rupiah $\approx$ \$ 12 USD. Standard errors in brackets are not corrected for predicted wages. 'Rural' refers to rural destinations; 'Urban' refers to urban destinations; 'Near' refers to destinations within 100 km from the individual's home location at age 18; 'Far' refers to destinations farther than 100 km from the individual's home location at age 18. All parameters are estimated using mixed logit, which includes estimating standard deviations of all parameter values and of the classical measurement error on wealth. The following parameter values are set exogenously $\beta = .0.95$ $r = 0.03$ $\rho = 2$

Table 14: Mixed Logit Estimation of Migration Costs by Migration Experience

		Migration Experience				
	All	**None**	**Some**	**Only Rural**	**Only Urban**	**Rural and Urban**
	(1)	(2)	(3)	(4)	(5)	(6)
Migration cost Rural	11.09	15.05	13.39	10.66	24.92	10.73
	(0.34)	(0.84)	(0.70)	(0.74)	(0.87)	(0.95)
Migration cost Urban	72.32	65.47	84.85	109.77	35.57	82.23
	(1.26)	(0.97)	(1.66)	(2.04)	(1.33)	(2.35)
Number of observations	100643	48293	41356	21138	11466	4505
		Migration Experience				
	All	**None**	**Some**	**Only Near**	**Only Far**	**Near and Far**
Migration cost Near	15.86	16.34	14.39	11.59	28.50	11.77
	(0.29)	(0.43)	(0.57)	(0.52)	(0.64)	(0.68)
Migration cost Far	69.82	60.79	89.60	106.05	37.95	87.64
	(1.12)	(1.73)	(2.26)	(2.08)	(2.55)	(2.71)
Number of observations	99589	48293	40450	21138	11466	4505

Average wealth is 129 (std dev: 127) and average annual income is 56 (std dev: 72). All values are reported in 100,000 (2000) Indonesian Rupiah $\approx$ \$ 12 USD. Standard errors in brackets are not corrected for predicted wages. 'Rural' refers to rural destinations; 'Urban' refers to urban destinations; 'Near' refers to destinations within 100 km from the individual's home location at age 18; 'Far' refers to destinations farther than 100 km from the individual's home location at age 18. All parameters are estimated using mixed logit, which includes estimating standard deviations of all parameter values and of the classical measurement error on wealth. The following parameter values are set exogenously $\beta = .0.95$ $r = 0.03$ $\rho = 2$

Table 15: Structural Estimation of Migration Costs by Migrant Network

		Know somebody at destination	
	All	**Yes**	**No**
	(1)	(2)	(3)
Migration cost Rural	11.09	8.50	12.73
	(0.34)	(2.84)	(5.07)
Migration cost Urban	72.32	57.25	72.51
	(1.26)	(9.25)	(8.47)
Number of Observations	100643	2282	9271
		Know somebody at destination	
	All	**Yes**	**No**
Migration cost Rural	15.86	6.84	17.88
	(0.29)	(2.21)	(4.62)
Migration cost Urban	69.82	68.71	74.12
	(1.12)	(9.82)	(6.25)
Number of Observations	99589	2269	9245

Average wealth is 129 (std dev: 127) and average annual income is 56 (std dev: 72). All values are reported in 100,000 (2000) Indonesian Rupiah ≈ \$ 12 USD. Standard errors in brackets are not corrected for predicted wages. 'Rural' refers to rural destinations; 'Urban' refers to urban destinations; 'Near' refers to destinations within 100 km from the individual's home location at age 18; 'Far' refers to destinations farther than 100 km from the individual's home location at age 18. All parameters are estimated using mixed logit, which includes estimating standard deviations of all parameter values and of the classical measurement error on wealth. The following parameter values are set exogenously $\beta = .0.95$ $r = 0.03$ $\rho = 2$

Table 16: Structural Estimation Benefits of Migrating

Change in Individual's Annual Consumption

1 year after migrating		5 years after migrating	
Moving Rural	Moving Urban	Moving Rural	Moving Urban
2.04%	3.62%	5.82%	30.71%

Change in Individual's Annual Wage

1 year after migrating		5 years after migrating	
Moving Rural	Moving Urban	Moving Rural	Moving Urban
8.36%	38.44%	3.36%	43.35%

Consumption changes are predicted by the model and compare consumption levels of those who moved to predicted consumption if the individual had not moved. Wage changes compare the actual wage the person received as observed in the data, to the predicted wage if the person had not moved. All numbers refer to annual averages and predicted changes after five years of moving are unconditional on whether the person is still at the destination.

Table 17: Counterfactual Experiments

	Predicted migration rates (%)		Predicted change average consumption	
	Home to Rural	Home to Urban	1 year	5 year
No Policy Status quo migration rates	1.25	1.01		
Providing credit at 5% interest rate Negative wealth at annual interest rate of 5%	0.87	1.61	6.86%	10.38%
Providing credit at 10% interest rate Negative wealth at annual interest rate of 10%	1.14	1.18	2.91%	4.36%
Credit at 5% conditional on migrating Negative wealth at annual interest rate of 5%	1.13	1.51	4.64%	7.93%
Credit at 10% conditional on migrating Negative wealth at annual interest rate of 10%	1.24	1.15	1.36%	2.49%
Migration 10% more costly 10% increase migration costs to Near and Far	1.20	0.62	1.41%	-6.30%
Migration 10% less costly 10% decrease migration costs to Near and Far	1.35	1.66	-0.73%	12.70%
Estimate Bryan et al (2014) $8.50 USD incentive conditional on migrating	1.31	1.05	-0.01%	0.45%
Restrict migration (Hukou) Migration to Far not allowed	1.46	0.00	1.20%	-21.30%
Increase Shock Intensity 5% increase standard deviation of shocks	1.37 (+ 0.12)	0.89 (- 0.12)	0.21%	-3.19%
Change amongst the poorest 50%	+ 0.23	- 0.21	-0.22%	-6.73%
Increase Shock Intensity 10% increase standard deviation of shocks	1.74 (+ 0.49)	0.74 (- 0.26)	-0.52%	-8.63%
Change amongst the poorest 50%	+ 0.93	- 0.38	-1.49%	-15.43%

Computational Appendix
Numerical Solution Dynamic Migration Choice Model

Infinite time horizon model

The dynamic choice model presented in Section 2 is solved numerically using value function iteration using the following algorithm:

1. Initialize a guess $V_0(A, l)$ for the value function using cubic spline interpolation over a grid of points in continuous A-space, where $A = x + w$ represents total cash on hand, x is an individual's wealth at the beginning of a period and w is the wage draw under consideration. The l-space is a set of discrete locations

2. Begin the iteration loop for $i = 1, 2, max_{iter}$, setting $V_{old} = V_0$ at the outset

   (a) For each combination of state variable values, (A_j, l_k), where A_j is a grid point in discretized A-space and l_k represents location k, calculate the value function $V_{new}(A_j, l_k)$ following equation 4

   (b) Update $V_{old} = V_{new}$

   (c) Repeat steps a and b until $max(V_{old}V_{new}) < tolerance\ level$

   (d) Once converged, save the value of the control variables $(c, l^{'})$ that maximizes the value function $V(A_j, l_k)$

   (e) Repeat steps b - d for all combinations of state variable values, (A_j, l_k). Update the resulting spline interpolation for the function $V(A, l)$

3. In order to derive general predictions of the model, simulate the choices an agent would make given a certain starting value of the state variables, (A, l)

   (a) In each period, the agent receives a random wage draw from his/her current location

   (b) Retrieve each location's value function from the model solution described in step 1

      i. Compute the value of staying at location l and accepting wage draw w_l by evaluating $v_l = V(x + w_l, l)$

      ii. Retrieve the value of moving to each of the other locations based on expected wages at those locations (as the draw draws are still unknown to the agent), that is, $v_{l'} = \int V(x + w_{l'} - m(l, l^{'}), l^{'})dF(w_{l'})$ for $l^{'} \neq l$

      iii. Make migration choice by choosing $max(v_1, v_2, , v_{nLoc})$

   (c) After the migration choice, choose the consumption choice calculated in the model solution described in step 1. If the choice was to stay, then $A = x + w_l$ using the wage draw offered at the beginning of the period. If the choice was to move, then $A = x + w_{l'}$, where $w_{l'}$ is new wage value drawn at random from the wage distribution at the new location $l^{'}$

   (d) Update the values of the state variables to $(x^{'}, l^{'})$ according to equation 2

   (e) Repeat steps a - d for all time periods

   (f) Repeat steps a - e for 10,000 agents

4. In order to derive comparative statics, repeat step 3 for different starting values of the state variables, (A, l), and various model parameters

Finite time horizon model

For the case of a finite time horizon, the model is solved numerically using a backwards induction procedure. The finite time version of the Bellman equation is

$$V_t(x,l) = \max_{c,l'} \left\{ U_t(c,l^{'}) + \beta \int V_{t+1}(x^{'},l^{'}) dF(w_{l'}) \right\} \tag{12}$$

It is assumed that the ending condition for a time horizon consisting of T periods is $V_T(A,l) = 0$. The backwards induction procedure utilizes this fact and is performed as follows:

1. Initialize $V_T(A,l) = 0$
2. For each $t = T-1, T-2, ..., 0$,
   (a) For each combination of state variable values, (A_j, l_k), where A_j is a grid point in discretized A-space and l_k represents location k, calculate the value function $V_t(A_j, l_k)$ according to equation 13, using the known value of the function $V_{t+1}(A_j, l_k)$.
   (b) Using the solution at each grid point, create a spline interpolation for $V_t(A,l)$ as well as for the associated optimal consumption decision (optimal migration decision is assumed to be chosen from a discrete set).

Structural Estimation using Nested Fixed Point Algorithm

I employ a two-loop nested fixed point algorithm for the structural estimation consisting of the following steps:

1. To initiate the outer loop, define a starting vector of model parameter values θ to be structurally estimated
2. In the inner loop, solve the infinite time-horizon model numerically in discrete time for the given set of model parameters θ using value function iteration as described above
3. Given the solution to the infinite time-horizon model, simulate the model solution for each individual-year pair observed in the data.
   (a) For each individual-year pair obtain the values of the state variables, (x,l), and the income shock, w_l, from the data
   (b) From the model solution described in step 2, obtain the values of the control variables $(c,l^{'})$, as predicted by the model
   (c) In the log likelihood function compare the predicted values of $(c,l^{'})$ to those observed in the data
   (d) Accumulate sum of terms in log likelihood function according to equation 10
   (e) Repeat steps a to c for all individual-year pairs in the data
4. In the outer loop, the vector of parameter values θ is varied and the inner loop is repeated to solve the model and simulate the data
5. Using extreme value assumptions on error between observed and predicted control variables, accounting for by the unobserved state variables ϵ, find θ that maximizes the log likelihood function

Chapter 2

Labor Market Changes in Response to Immigration:

Evidence from Internal Migration Driven by Weather Shocks

with Jeremy Magruder[1]

Abstract

We study the labor market impact of internal migration in Indonesia by instrumenting migrant flows with rainfall shocks at the origin area. Estimates reveal that a one percentage point increase in the share of migrants decreases income by 1.22 percent and reduces employment by 0.26 percentage points. These effects are different across sectors: employment reductions are concentrated in the formal sector, while income reduction occurs in the informal sector. Negative consequences are most pronounced for low-skilled natives, even though migrants are systematically highly skilled. We suggest that the two-sector nature of the labor market may explain this pattern.

JEL Classification: J21, J61, O15

---

[1]Magruder: Department of Agricultural and Resource Economics, University of California at Berkeley. We are grateful to Max Auffhammer Michael Anderson, Sam Bazzi, Marshall Burke, David Card, Michael Clemens, Alain de Janvry, Edward Miguel, Sofia Villas-Boas and seminar participants at numerous presentations. We gratefully acknowledge financial support from the U.C. Berkeley Population Center. All errors are our own.

1 Introduction

Public debate often expresses concerns that immigrants take the jobs of natives and increase labor market competition, which causes wages to fall. While this debate is global, the academic literature has concerned itself primarily with immigration to high-income countries, with particular attention given to Mexican immigration to the United States. Even though consensus within this literature remains somewhat elusive, we know even less about the labor market impacts of internal migration in developing countries.

We might anticipate these labor market impacts to be quite different for several reasons. First, the costs of migrating internally are much lower than the costs of international migration, which may allow migration to respond more quickly to favorable labor market conditions and affect the number and characteristics of migrants. Secondly, labor markets in developing countries are structurally quite different from the United States. The conventional characterization of developing country labor markets features a heavily regulated formal sector which coexists with an uncovered informal sector, exhibiting lower wages and productivity (e.g. Harris and Todaro, 1970). The effects of an increase in labor supply on working conditions may be quite different as labor supply puts pressure on both of these sectors, which could change both wages within a sector as well as the availability of labor market opportunities across sectors. Finally, relatively thin markets may limit the firm's capacity to adapt to a surge (or reduction) in labor supply by relocating or entering new markets, and thereby potentially increase the magnitude of labor market responses.

Despite the difference in potential mechanisms for labor market effects of immigration, estimating the effects of internal migration in developing countries retains the primary econometric concern that has challenged estimates of Mexico-U.S. migration. That is, regressing labor market outcomes on immigrant stocks may be confounded by the tendency of migrants to be attracted to areas with better labor market opportunities, often referred to as the 'moving to opportunity bias'. OLS estimates of labor market impacts of migration are therefore likely to be biased in the positive direction. This paper uses an instrumental variable approach to address this issue. Using the Indonesia Family Life Survey, we document the migration decisions of almost 29,000 individuals within Indonesia over 13 years. We use these empirical migration patterns to form catchment areas of origins that send migrants to each destination district. We then generate exogenous variation in the number of migrants in each district using rainfall shocks in these catchment areas, following Munshi (2003).

We find that a one percentage point increase in the share of migrants decreases average income per hour by 1.22 percent and reduces the employment rate by 0.26 percentage points. We show that, as expected, the negative effects using IV-2SLS are larger than OLS estimates. The negative income effects are concentrated in the informal sector, with a 2.45 percent decrease of informal sector income, and the employment effects are largest in the formal sector at 0.32 percentage points. This distinction between sectors provides direct evidence in support of the conventional characterization of how a two-sector labor market responds to an increase in labor supply, and is consistent with other evidence on the importance of binding wage floors in the formal sector in 1990s Indonesia (e.g. Alatas and Cameron 2008, Magruder 2013).

Furthermore, while there are negative labor market effects of immigration for all natives, we find that these negative consequences are most pronounced for those with lower levels of education. This finding is consistent with earlier studies on the U.S. that have attributed this pattern to increased substitutability between low-skilled immigrants and low-skilled natives. Unlike in those studies, however, the pool of internal migrants in Indonesia is relatively high-skilled, at least in terms of education levels. We discuss a number of reasons why poorly educated Indonesians may be disproportionately affected by an influx of highly educated migrants. We find little evidence that this result could be explained by differences in the returns to skill between migrants and natives, such that highly-educated migrants are more substitutable to low-educated natives in terms of skill level. We also do not find support for the hypothesis that this result is driven by differences between average treatment effects of migrants in general and local average treatment effects of weather-induced migrants that we estimate. Instead we propose that this result may be understood as another consequence of the two-sector labor market with a wage floor in the formal sector where the less-skilled group faces chances of disemployment or employment in the informal sector.

The identification assumption underlying our estimation strategy is that precipitation in the migration origin areas does not affect local labor market at migrant destination areas once precipitation at the destination itself is controlled for. A concern that may arise is that rainfall measures at the origin areas are correlated with wages or employment in destination areas through channels other than migration. This would constitute a violation of the exclusion restriction and may occur due to local trade, for example of agricultural products, which could affect labor market conditions at the destination. We test for the exclusion restriction by restricting our analysis to migration over longer distances. If the effects were driven by local trade channels, we would expect the magnitudes of our results to reduce as the intensity of trade and economic linkages decreases with distance. In contrast, our results in Section 6 document that labor market impacts are stronger using only longer distance migrants. In this section we also use simulations to test if serial correlation within the catchment area could drive our results and show that the exact migration patterns we observe - and not any other patterns, for example those created by local trade - are responsible for our results.

A large number of studies have estimated labor market impacts of immigration in OECD countries, especially in the case of migration from Mexico to the United States. An overview of the literature is provided by the survey articles Okkerse (2008) and Pekkala Kerr and Kerr (2011). While the literature on high-income countries is vast, fewer related studies have been carried out in a developing country context. Two exceptions are Bryant and Rukumnuaykit (2007), who found that immigration in Thailand reduced wages but did not adversely impact employment, and Strobl and Valfort (2014) who find adverse employment effects in Uganda, especially in areas with low capital mobility[2]. Despite the large focus on OECD countries, fears that immigrants increase unemployment and lower wages are expressed not only in high-income countries, but also in less developed countries. Comparing 46 countries with varying levels of income per capita, Kleemans and Klugman (2009) find that negative attitudes towards immigrants are most pronounced in middle income countries. Indonesia ranks second in terms of negative attitudes towards immigrants, with only Malaysia being ranked higher.

While much earlier work studies international migration, this study focuses on internal

[2] Strobl and Valfort (2014) was developed contemporaneously to this paper and also uses weather shocks as a source of exogenous variation. It differs from this paper in its context, its focus on infrastructure and capital stocks, and in abstracting away from frictions in a two-sector labor market.

migration[3]. The absence of crossing an international border is likely to affect the number, type and possibly the labor market impact of immigrants. Despite frequent discussions about migrants from developing countries entering developed countries, the overwhelming majority of migrants move within developing countries. Bell and Muhidin (2013) estimate that worldwide there are approximately 763 internal migrants compared to 230 international migrants, and Deb and Seck (2009) estimate that one out of four Indonesians live parts of their lives in a district different from their place of birth, which translates to over 60 million internal migrants.

Indonesia - like other developing countries - also distinguishes itself from the OECD countries considered in prior work through the coexistence of a large informal sector and a formal sector characterized by strict labor market regulation. In the 1990s, minimum wages in Indonesia tripled in nominal terms and doubled in real terms to reach the 38^{th} percentile of wages for full-time workers in the IFLS data in 1997. Other work, such as Alatas and Cameron (2008) and Magruder (2013), document that these minimum wages were binding on at least part of the labor force. As we find in this paper, the two-sector labor market in Indonesia results in different effects of large scale migration on labor market outcomes for natives.

We open this paper by motivating the empirical strategy in Section 2 and describing the data in Section 3. The main results are presented in Section 4 and Section 5 explores heterogeneous labor market effects by skill level. Robustness checks are presented in Section 6, and Section 7 concludes.

2 Empirical Strategy

This paper uses weather in the migrant's origin area as an instrument to get exogenous variation in the number of migrants entering a destination area. Our approach follows Munshi (2003) who studies network effects amongst Mexican immigrants in the U.S. This instrument may be successful if economic outcomes depend on rainfall, which is the case in Indonesia as many Indonesians depend on rainfed agriculture and various studies have shown that higher rainfall raises agricultural productivity, income and wealth (Levine and Yang, 2006, Kishore et al., 2000 and Kleemans, 2015).

Intuitively, the first stage is meant to capture the following process: If a particular destination area d hosts immigrants from origin area o, then we expect that a negative rainfall shock in origin area o will drive people to destination area d. This gives exogenous variation in the number of immigrants in a destination area. In the estimation, we use individual-year pairs as units of observation. In equation form, the percentage of people who are migrants in each destination area at time t, $migrant_{dt}$, is regressed on rainfall in the origin areas. Each destination area d, hosts migrants from a number of origin areas, which we refer to as the 'catchment area' of that destination, defined as $C(d)$. The first stage can

[3] In doing so, it builds on work by Boustan et al. (2010), who examine labor market effects of internal migration in the U.S. during the Great Depression.

be expressed by the following equation:

$$migrants_{dt} = b \sum_{o \in C(d)} (w_o rainfall_{o,t-1}) + X_{dt}c + d_t + a_d + e_{dt} \tag{1}$$

Unlike Munshi (2003), we take rainfall in the entire catchment area of each destination into account. This is captured through the summation in equation (1), where w_o is the weight of origin area o, which is proportional to the share of migrants from origin area o in destination area d. Weights are determined by the number of migrants from that origin area over all years in the sample and are fixed over time. Section 6 shows that the results are robust to various alternative definitions of the origin area weights w_o, including using weights based on census data. X_{dt} are control variables including dummies for gender, age group and education level. d_t and a_d represent time and destination fixed effects and e_{dt} is the error term. By using only time-invariant migration patterns, we ensure that whichever labor market characteristics affect those patterns will be absorbed by the destination fixed effects a_d. In addition to destination fixed effects, we run robustness checks including individual fixed effects. In all specifications, X_{dt} includes destination rainfall measures as control variables to account for possible correlation between origin and destination rainfall and the direct impact of destination level rainfall on the labor market in the second stage. While equation (1) shows the first stage when using rainfall in year $t-1$ as an instrument, we experiment with several lagged variables and various measures of rainfall as discussed in Section 6.

Given that negative rainfall shocks in the origin area drive people to a destination area, the second stage asks whether this changes individual labor market outcomes in the destination area at time t. The second stage is given by

$$Y_{it} = \beta_1 \widehat{migrants}_{dt} + X_{dt}\gamma + d_t + a_d + \varepsilon_{dt} \tag{2}$$

Y_{it} refers to the individual-level labor market variables of interest. We look at the effect on income and employment, overall as well as in the formal and informal sector. $\widehat{migrants}_{dt}$ are the predicted values from the first stage. X_{dt} contains the same set of control variables as in the first stage, d_t and a_d are time and destination fixed effects and ε_{dt} represents the error term.

The main assumption underlying this approach is the exclusion restriction, which states that the only channel through which rainfall in the origin area affects labor market outcomes in the destination area is through increased numbers of immigrants. Given that we have controlled for destination area rainfall and deviations from historical rainfall patterns are hard to predict, this restriction amounts to assuming that local rainfall is a sufficient statistic for the direct effects of global rainfall patterns on labor market outcomes. We test the robustness of this assumption by considering only long-distance migration and alternate summary measures of local rainfall as a control variable in section 6, below. Furthermore, as with any instrumental variables approach, the estimated effects will be local average treatment effects (LATE). This means that the labor market impacts are estimated for those immigrants that are induced by weather shocks. Section 5 explores how weather-induced migrants compare to average migrants.

3 Data

3.1 Migration and labor market data

Limited data availability has prevented earlier studies from obtaining empirical evidence on migration in developing countries as such studies require data collection across regions and across time. In most panel datasets, migrating individuals attrite from the panel, which hinders inference. This study uses the Indonesia Family Life Survey (IFLS) as the main data source, a panel dataset that is known for relatively low rates of attrition. This longitudinal survey is representative of about 83 percent of the Indonesian population and contains individuals living in 13 of the 27 provinces of the country (Strauss et al., 2004). The analyses are based on the first three waves of the IFLS, which covers the period from 1988 to 2000[4].

Attrition is low in the IFLS: between the first and second wave, attrition is less than six percent, and the cumulative attrition of households between the first and third wave is five percent (Thomas et al., 2001). Low attrition and intensive efforts to track respondents from the original sample makes the IFLS particularly suitable for migration analysis. In addition to migration data, the dataset contains extensive information on the respondents' labor market outcomes, education, and other characteristics.

Using the migration modules of the IFLS, a dataset is obtained of 28,841 individuals, who recorded when and where they migrated after the age of 12. In addition to migration data that is based on recall between waves, the dataset contains information on where respondents were born and where they lived at age 12. This information is transformed into a panel dataset that reports the person's location in each year. This results in a panel dataset of individual location decisions and labor market outcomes over 13 years with a total of 192,522 individual-year observations. Table 1 provides summary statistics of this dataset.

Education is defined on a scale from 0 to 4, ranging from no education (0) to university education (4). While imperfect, education is used to determine a person's skill level: if a person has obtained at least some high school education (education value of 2), he or she will be defined as 'high-skilled', while those with no or only primary education are referred to as 'low-skilled'. This cut-off is chosen because it leads to the most even split between low and highly-educated individuals. The main labor market variables in the current study are income and employment, overall and in the formal and informal sectors. All these variables are defined at the individual level and observed for each year of the panel. Income is recorded as log income per hour in Indonesian Rupiah and the employment variable indicates whether a person is working, as opposed to housekeeping, going to school, being unemployed or retired etc. The overall employment rate is 79 percent and 52 percent of the sample is self-employed. While imperfect, this variable is used to characterize the informal sector as self-employed individuals are more likely than wage workers to work informally. Note that an individual may simultaneously have a job in the formal and another job in the informal sector, in which case they count as being employed formally and informally at the same time.

Throughout this study, a migrant is defined as a person who does not live at his or her place of birth, as opposed to natives who still live where they were born. Although other

[4] The fourth and last wave, carried out in 2007 and 2008, is not included in the sample because no information was collected on annual income and the answer categories of sector of employment changed and became incompatible with earlier waves.

definitions have been explored, this is the most commonly used definition in the literature (UNDP, 2009). For each destination we count the number of migrants in each year and call this the migrant stock. This number is divided by the total population in that destination to get the migrant share of the population, which is used as the main migration variable in this study. Location information is available for three geographical levels. The largest level is the province of which there are 34 in Indonesia. These are further divided into districts (Kabupaten) and sub-districts (Kecamatan). This study defines districts (Kabupaten) as separate geographical units, meaning that a migrant is someone who is not born in the district they live in. While sub-districts could have been chosen instead, these are often small geographical units of which there are more than 6,500 in Indonesia. This would mechanically create a large number of migrants, some of whom only move over a short distance and may not consider themselves migrants. The final dataset contains 205 districts hosting 28,841 individuals. Even at this level of aggregation, several districts host only few sampled individuals. As a robustness check, districts that host only one or two individuals are dropped, which does not alter the results (see Section 6).

Comparing natives and migrants in Table 1 reveals that internal migrants in Indonesia are systematically higher skilled as measured by education level than most natives. Their hourly income is 23 percent higher and they work more hours per week. While overall employment rates of migrants and natives are comparable at 79 percent, migrants are 10 percentage points more likely than natives to work in the formal sector (47 percent and 37 percent, respectively).

Measurement of overall labor market impacts is challenged by the fact that the pool of employed people is changing as immigrants arrive and leave. Migrants who recently arrived may still be looking for work or may initially have to settle for a lesser-paying job, which would mechanically push coefficients in the negative direction. In order to deal with this potential bias, we estimate impacts for natives only.

3.2 Weather data

Weather data are obtained from the Center for Climatic Research of the University of Delaware (Matsuura and Willmott, 2009). Monthly estimates of precipitation and temperature are available for grids of 0.5 by 0.5 degree, which is approximately 50 by 50 kilometers. These data are based on interpolated weather station data and are matched to IFLS household locations using GIS data. Figure 1 shows the survey locations of the IFLS on a map of Indonesia as red dots and the blue grids represent the weather data that the locations are mapped to.

While this study explores various weather measures, precipitation z-score is used as the main weather variable. Z-scores are obtained by subtracting the precipitation mean and dividing by the standard error. This is in line with Maccini and Yang (2009) who argue that rainfall is the most important source of weather variation in Indonesia. Figure 2 shows how average precipitation varies across years. Temperature shows less variation over time due to Indonesia's equatorial location. Lagged weather variables are used to allow for a lagged response to bad weather shocks. Instead of using annual data from each calendar year, all measures are created from July until June in the year after to reflect the growing seasons in Indonesia. All analyses are repeated using calendar years, which does not significantly change the results (results not shown).

In addition to precipitation z-scores, this study carries out robustness checks with various other weather variables. Precipitation levels and temperature are used, as well as precipitation squared and cubed to allow for nonlinear effects. To capture unusual weather patterns, deviations from the mean and growth are used. Finally, variables for extreme weather events are created. Droughts are defined as seasons in which precipitation is less than a standard deviation below the mean, and floods as seasons in which precipitation is more than a standard deviation above the mean. The next section describes the results using precipitation z-scores and Section 6 discusses the robustness of these results when using a range of alternative weather variables.

4 Results

4.1 Migrants' responsiveness to weather shocks

The first stage analysis examines whether people are more likely to leave the place they live after negative weather shocks. Table 2 shows the first stage results using various sets of rainfall lags. Origin level precipitation z-scores, shown in the upper part of the table, are summed up over the catchment area of each destination according to equation (1). As explained in Section 2, all specifications include the same number of lags of destination rainfall, shown in the lower part of the table, to control for possible correlation between origin and destination weather measures and a direct impact of destination level rainfall on the labor market. All regressions include socio-economic control variables as well as time and destination fixed effects, and standard errors are clustered at the destination level[5].

The results consistently show a significant negative coefficient on origin area rainfall measures, indicating that people are more likely to migrate in response to lower rainfall. The coefficients on rainfall at time $t-1$ are largest and highly significant, which is in line with the basic intuition that people respond to bad weather shocks with a slight lag. Note that, as discussed in the previous section, rainfall in year $t-1$ starts in July of year $t-1$ and ends in June of year t, so as expected this creates the largest migratory response in year t. The F-statistic of joint significance when using only lagged rainfall in the second column is sufficiently high, at 21.04. Therefore, lagged precipitation will be used as the main instrument.

Destination rainfall measures are only used as control variables rather than instruments to avoid violating the exclusion restriction, as there are likely other channels through which weather shocks at the destination affect labor markets at the same location. The coefficients on rainfall at the destination are positive and significant, and slightly smaller in magnitude

[5] Since rainfall shocks across the catchment area are used in calculating the independent variable of interest, it would be desirable to use a Conley (1999) cluster to allow for correlations across the catchment area. This correction becomes computationally infeasible on our full dataset with 192,522 observations at the individual-year level. To test whether the potential correlations in the instrument could be biasing our standard error estimates, we reran the primary analysis at the destination level using the Conley errors. Results for the first stage and primary second stage are available in Appendix Table A5. This approach demonstrates that the Conley errors are smaller than the destination clustered errors, suggesting that p-values presented in this paper are conservative.

than the coefficients on origin rainfall. This may suggest that positive weather shocks at the destination are a pull factor to migrants, to a slightly lesser extent than bad weather shocks at the origin are a push factor. Destination rainfall measures are used as control variables in all remaining analyses.

Table A1 in the Appendix compares our main instrument (column 1) to alternative specifications of the first stage. Results are broadly similar when using precipitation levels instead of z-scores (columns 2 and 4) and individual fixed effects instead of destination fixed effects (columns 3 and 4). Table A6 shows results using longer lags of rainfall and reveals that rainfall remains significant for four years but that the F-statistic of joint significance reduces. The overall labor market estimates in columns 3 and 4 are similar but preference is given to specifications with stronger predictive power of the first stage relationship.

4.2 Labor market response to immigration

Using exogenous variation in the number of migrants entering a destination area caused by weather shocks in the origin areas, this section investigates whether increased labor market competition affects income and employment.

First, OLS regressions are carried out of the reduced-form relation between rainfall at the origin areas and labor market conditions at the destination. If increased numbers of immigrants, induced by negative weather shocks, reduces income and employment, then we expect this reduced-form relationship to be positive. The first column of Table 3 shows the reduced-form relationship between precipitation and average log income per hour, and the second shows the relation between precipitation and employment. Both columns confirm the prediction of a positive reduced-form relationship.

The first stage has established that negative rainfall shocks in origin areas increase the likelihood of migrating. This exogenous variation in migrant stock is used to study labor market responses in the second stage. Table 4 shows the main second stage results and compares them to OLS regressions in the first and third column. The first column shows that the coefficient on log income per hour is indistinguishable from zero when using OLS. This contradicts economic theory that predicts negative effects, but this result is likely due to the fact that a simple OLS regression is unable to isolate causal effects. As discussed in Section 1, OLS regressions may be biased in the positive direction due to the 'moving to opportunity' bias. Comparing column 1 to the IV-2SLS specifications in column 2 suggests that the moving to opportunity bias in the OLS is large enough to cancel out the negative causal estimates that are revealed using the preferred IV-2SLS specifications. Column 2 indicates that an increase in the migrant share of the population by 1 percentage point reduces average income by 1.22 percent. In the IFLS data, the average share of migrants at a destination is 16 percent so this corresponds to a 6 percent increase of the current migrant share.

It is worth noting that the difference between OLS and IV-2SLS estimates is larger in this study than the difference found in the literature on Mexican immigrants entering the U.S. One interpretation is that the moving to opportunity bias is larger in our study, which may be caused by the fact that internal migrants face fewer social and physical barriers to migrate, including no international border to cross.

The last two columns of Table 4 show that an increased proportion of immigrants reduces employment. Comparing the OLS specification in column 3 to the IV-2SLS regression in column 4 again suggests the existence of a moving to opportunity bias in the OLS specification. Column 4 shows that increasing the share of immigrants by 1 percentage point reduces the employment rate by 0.26 percentage points.

4.3 Labor market effects across sectors

Indonesia's labor market in the 1990s can be characterized by a competitive informal sector and a formal sector with high and binding minimum wages, as described in more detail in the next section. If wages are determined competitively in the informal sector and affected by a wage floor in the formal sector, we would anticipate different effects of migration in these two sectors. We test for this possibility in Table 5 and confirm that the effects differ across sectors: a one percentage point increase in the migrant share reduces formal sector employment by 0.32 percentage points while no employment effects are found in the informal sector. Conversely, a percentage point increase in the migrant share reduces income in the informal sector by 2.45 percent, while no adverse income effects are found in the formal sector. Appendix Tables 2, 3 and 4 repeat this analysis using precipitation levels instead of precipitation z-scores as instruments, and using individual fixed effects instead of destination fixed effects. The results are broadly similar to those in Table 5. These results provide support for the two-sector characterization, and for the hypothesis that wage minima bind and affect the number of jobs: when labor supply increases, we observe some workers get crowded out of the formal sector. We similarly observe workers in the informal sector receive lower wages, consistent with the hypothesis that those wages are set competitively.

5 Heterogeneity by Skill Level

In the previous section, we have established that native workers are on average negatively impacted by an influx of migrants. We may anticipate that these negative impacts will be borne heterogeneously by workers of different skill levels. A motivation for this insight is given in the labor market model developed by Card and Lemieux (2001) and Borjas (2003), which proposes a single-sector, competitive labor market. We begin by developing their model.

Suppose aggregate output in the economy, y, is given by the following economy-wide production function:

$$y = K^{1-\alpha} L^{\alpha}$$

The aggregate labor supply L incorporates contributions from n types of labor with substitutability parameter ν, so that

$$L = \left(\sum_{k=1}^{n} \theta_k L_k^{\nu} \right)^{1/\nu}$$

where $0 < \nu < 1$. Substituting in gives

$$y = K^{1-\alpha} \left(\sum_{k=1}^{n} \theta_k L_k^{\nu} \right)^{\alpha/\nu}$$

Wage of group j is determined by the marginal product of labor:

$$w_j = \frac{\partial y}{\partial L_j} = \alpha \theta_j K^{1-\alpha} L_j^{\nu-1} \left(\sum_{k=1}^{n} \theta_k L_k^{\nu} \right)^{(\alpha-\nu)/\nu}$$

In this set-up, we can examine how wage of group j changes as the amount of laborers of group g increases. Two types of comparisons here are particularly interesting: how wages change with an influx of workers of the same type and how wages change with an influx of workers of a different type. First, consider wage responses to an influx of own-type workers:

$$\frac{\partial w_j}{\partial L_j} = \alpha(\alpha-\nu)\theta_j^2 K^{1-\alpha} L_j^{2\nu-2} \left(\sum_{k=1}^{n} \theta_k L_k^{\nu} \right)^{(\alpha-2\nu)/\nu} - (1-\nu)\alpha\theta_j K^{1-\alpha} L_j^{\nu-2} \left(\sum_{k=1}^{n} \theta_k L_k^{\nu} \right)^{(\alpha-\nu)/\nu}$$

Alternatively, consider the wage response of group j to an influx of group g workers:

$$\frac{\partial w_j}{\partial L_g} = \frac{\alpha(\alpha-\nu)\theta_j \theta_g K^{1-\alpha}}{L_j^{1-\nu} L_g^{1-\nu}} \left(\sum_{k=1}^{n} \theta_k L_k^{\nu} \right)^{(\alpha-2\nu)/\nu}$$

The following statements summarize the relative wage responses due to an influx of group j and group g workers when labor markets are characterized by competitive wages and a single sector. Please refer to the Appendix for proofs of these statements.

$$1 : \frac{\partial w_j}{\partial L_j} < 0$$

$$2 : \alpha < \nu \Rightarrow \frac{\partial w_j}{\partial L_g} < 0$$

$$3 : \frac{\partial w_j}{\partial L_j} < \frac{\partial w_j}{\partial L_g}$$

The last statement conveys the basic intuition that groups that are similar to incoming migrants, and therefore demonstrate a larger degree of substitutability, face increased competition and in this model will be affected to a larger extent than groups that are more dissimilar from migrants.

Extrapolating to the Indonesian context, we have documented that internal migrants in Indonesia are more educated than native non-migrants. If labor markets in Indonesia respond similarly as the competitive markets in more developed countries, and if education is an effective proxy for skill, we would anticipate that highly educated natives are most negatively impacted by migrants.

5.1 Empirical estimates by skill level

This section examines how the labor market effects of internal migration in Indonesia differ by skill level. Table 6 compares labor market effects of those with primary school education or less (uneven columns) to the effects on those who received higher levels of education (even columns), and performs these analyses across sectors (columns 1 and 2), for the formal sector only (columns 3 and 4) and for the informal sector only (columns 5 and 6). In contrast to model predictions, negative labor market effects are most pronounced for those with lower levels of education: increased competition in the formal sector drives low-skilled workers out of this sector, making them 0.28 percentage points less likely to be formally employed after a one percentage point increase in the migrant share. Formal employment impact for high-skilled individuals is smaller and insignificant. Similarly, a percentage point increase in migration reduces informal sector income by 3.39 percent for the low-skilled, while no adverse income effect is detected for high-skilled individuals.

In some ways, these results recall estimates from the literature on Mexico-U.S. migration, where the lowest-skilled natives are also disproportionately affected. In that context this finding is usually interpreted as a substitution effect, because immigrants are similarly low-skilled. In the Indonesian context, however, migrants achieve higher levels of education than natives. This seems at odds with the results derived from the single sector, competitive labor market model. In the subsequent sections, we discuss several hypotheses that could explain our finding that low-skilled natives still face the largest negative labor market consequences from high-skilled migration.

5.2 Measuring skill level

First, it is possible that education is not the important dimension of skill in this setting, or is differentially important for natives and migrants, so that highly educated migrants are most substitutable with less educated natives. If this hypothesis were true, we might expect the returns to education to be lower for migrants than for natives. Table 7 shows Mincer-style regressions, using both the coarse educational categories we have used throughout in this paper (in columns 1-3) and a more conventional specification using a linear years of education variable (in columns 4-6). Focusing on columns 4-6, we see that an extra year of education is worth about 8% more in earnings, which accords with similar analyses in other contexts. Moreover, whichever educational measure we use, we see that the returns to education are qualitatively similar for migrants and for natives (and in fact, larger for migrants). We conclude that education has a similar relationship to skill level for migrants and for natives, as the difference in earnings between more educated and less educated migrants is at least as large as that for natives. Given that Table 1 already showed us that migrants earn on average higher incomes than natives, it seems implausible that migrants have on average lower skill than natives. We thus find little support for the hypothesis that education is not a relevant dimension of skill.

5.3 Weather-induced migrants

Second, it is possible that local average treatment effects are different from average treatment effects in this context. Our estimates capture the effects of the type of migrants who respond to negative rainfall shocks, who may be different from the average migrants in our summary skill measures. Our first stage showed that individuals are more likely to engage in internal migration in response to bad rainfall; poorer individuals may respond this way due to an insurance motive (Kleemans, 2015), which could lead to our empirical results.

Alternatively, richer people may respond to poor rainfall due to lower opportunity costs at home (Bazzi, 2015), which would render the patterns here even more striking. In Table 8, we test whether contemporaneous rainfall shocks also create a pool of migrants who are less educated compared to most migrants by restricting our sample to migrants and regressing an indicator variable for high education on origin rainfall measures. If the people who migrate when rainfall is poor have less education than migrants on average, we should see a negative and significant relationship between origin precipitation and migrant education. The point estimates on both contemporaneous and lagged precipitation are positive and mostly insignificant, suggesting that the LATE-complying migrants are not different on average from other migrants in terms of education.

5.4 Dual-sector model

Alternatively, it is possible that the differential labor market structure leads to different substitution patterns. In the beginning of this section, we maintained the assumption of competitive wage-setting in a single sector to replicate the Card and Lemieux (2001) and Borjas (2003) result that similar natives would be most negatively affected by migrants, a pattern which was rejected in our data. Here, we weaken the assumption to allow for a two-sector labor market.

The classical characterization of a two-sector labor market features a formal sector where wages are subject to binding labor regulation, resulting in a shortage of formal sector jobs (e.g. Harris and Todaro, 1970). This labor market institution is one of the main motivations for the presence of a large and competitive informal sector. To assess these features, suppose that in the formal sector $w_j \geq \hat{w}, \forall j$, where $\hat{w}$ represents a wage minima. Excess workers work in the informal sector, where workers of all types are homogeneously productive, and the production function is given by $\frac{1}{\gamma}L_I^\gamma,\ \gamma \leq 1$. They therefore earn $I \equiv L_I^{\gamma-1}$, with $I(L_I) < \hat{w}$ This case seems particularly relevant to 1990s Indonesia, where minimum wages were high and quickly growing, and there is a large and vibrant informal sector[6]. Here, we demonstrate that the key result of the Card and Lemieux (2001) and Borjas (2003) model - that individuals of the same skill group are most affected by immigration - no longer necessarily holds in labor markets with these features. Returning to the model, consider group g who is constrained, so that

$$w_g = \hat{w} = \frac{\alpha\theta_j K^{1-\alpha}}{\hat{L}_g^{1-\nu}} \left(\sum_{k=1}^{n} \theta_k L_k^\nu\right)^{(\alpha-\nu)/\nu}$$

$$\hat{L}_g = \left[\frac{\alpha\theta_j K^{1-\alpha}}{\hat{w}} \left(\sum_{k=1}^{n} \theta_k L_k^\nu\right)^{(\alpha-\nu)/\nu}\right]^{1/(1-\nu)}$$

and mean wages for group g are

$$\bar{w}_g = \frac{I(L_g - \hat{L}_g) + \hat{w}\hat{L}_g}{L_g}$$

[6] For a description of wage minima in 1990s Indonesia see, e.g., Alatas and Cameron (2008) or Magruder (2013).

Note that now, the changes in wages for group g in response to an influx of group j is

$$\frac{\partial \hat{w}_g}{\partial L_j} = \frac{\hat{w} - I}{L_g}\frac{\partial \hat{L}_g}{\partial L_j} + \frac{(L_g - \hat{L}_g}{L_g})(\gamma - 1)\frac{I}{L_I}\frac{\partial L_I}{\partial L_j}$$

There are therefore two effects of an influx of group j on constrained group g. First, the fraction of group g workers in the formal sector may change. Since formal sector wages are higher, this changes the average wage of group g. Second, the wage rate in the informal sector may change due to a change in the labor supply to that sector. This observation leads to the following predictions, all focused on the case where $\alpha < \nu$. Proofs can be found in the Appendix.

1. An increase in immigration from any unconstrained group will decrease formal employment for the constrained group g $\left(\frac{\partial \hat{L}_g}{\partial L_{g'}} < 0\right)$. Since group g is constrained, formal sector wages will stay constant.

2. An increase in immigration from any unconstrained group will (weakly) decrease wages in the informal sector.

3. For some parameterizations, the mean effect on immigration of unconstrained group j on wages of group g will be larger than the effect on own-group wages. If labor has a constant marginal product in the informal sector ($\gamma = 1$), then effects on group g will be larger if

$$w_j(1-\nu) < \frac{(\alpha - \nu)\theta_j L_j^{\nu}}{(\sum_{k=1}^{n} \theta_k L_k^{\nu})}\left[w_j - \left(\frac{\hat{w} - I}{1-\nu}\right)\frac{\hat{L}_g}{L_g}\right] \tag{3}$$

With a declining marginal product in the informal sector, this bound is conservative, that is, the range of parameters which produce larger out-group effects is larger.

Inequality 3 is guaranteed to be satisfied if $\nu \to 1$, and for any given set of parameter values where $\alpha < \nu$, there exist values of ν, strictly less than 1, which satisfy this equation. Moreover, we note that as $\theta_j L_j$ increases relative to the overall size of the effective labor force, the set of values of ν that cause a greater change in wages for lower-skilled groups grows. This may be the case in our empirical analysis, as we compare low and high-skilled workers, each of which constitute about half of the total labor force. Low-skilled natives resemble group g in the model, which is adversely impacted by the minimum wage in the formal sector, while employment opportunities for high-skilled natives in the formal sector are unaffected. Summary statistics on formal and informal employment in Indonesia indeed reveal that workers with high levels of education are much more likely to work in the formal sector (72 percent) than those with lower levels of education (48 percent). The two-sector model put forward in this section predicts that if there is significant substitutability between high and low-skilled workers, low-skilled natives may face larger negative effects due to an influx of high-skilled migrants than high-skilled natives, which may explain our empirical results.

6 Robustness

In this section we perform various robustness checks to assess how sensitive our results are to changes in the specifications. Our necessary identification assumption is that local labor markets at the destination are not impacted by rainfall in the migration catchment area after controlling for the precipitation that the destination actually experiences. This assumption would be violated if, for example, increased incomes in the catchment area increase labor demand at the destination, perhaps due to trade. This would be a problem for our analysis if trade patterns are correlated with the migration patterns in the catchment area. First we will test this assumption by looking at the labor market impact of long-distance migration only. Then, we will test whether the migration patterns themselves appear relevant or if serial correlation within the catchment area would yield similar estimates for many correlated effects.

6.1 Long-distance migration

The exclusion restriction would be violated if rainfall at the origin affects labor market conditions at the destination via channels other than migration. This would happen if the areas are economically connected through the movements of goods rather of people. Specially, good rainfall conditions at the origin could increase the supply and affordability of agricultural trade into urban areas, which could stimulate local labor markets. Table 9 tests this alternative channel by comparing our main results to those obtained when only considering migration that is at least 100 km in distance. After revealing a strong first stage relationship in column 2, columns 4 and 6 show that labor market effects are in fact larger for long-distance migration. The coefficient of column 4 shows that a one percentage point increase in the share of migrants reduces income by 2.32 percent, compared to 1.18 percent in our main specification, and that employment decreases by 0.62 instead of 0.23 percentage point[7]. Results are robust to alternative distance cut-offs (results not shown). The increased labor market impact of long-distance migration may result from differential sorting of distinct types of migration across distance[8]. If our results were driven by local trade, there would be no reason to believe that these effects would be stronger over longer distances.

6.2 Serial correlation across the catchment area

In our analysis so far, migration patterns showed up as a weighting on catchment area rainfall: origins for a potential destination received higher weight if a larger number of migrants came from this origin. If our exclusion restriction is invalid, and trade patterns (or any other relationship between destination labor markets and catchment area rainfall) are not

[7] The migration rate is naturally lower for long-distance migration at 7.4 percent compared to 15.9 percent, so a one percentage point increase corresponds to a 13.5 percent increase in the share of long-distance migrants.

[8] Using the same data source, Kleemans (2015) finds that those migrating over longer distances are positively selected in the sense that they are higher skilled and wealthier.

coincident with migration patterns, we may expect different weighting schemes on catchment area rainfall to produce similar estimates. We test this hypothesis by bootstrapping precipitation weights. Our approach is as follows: for each destination, we fix the bootstrap catchment area to be the empirical migration catchment area we observe. We then bootstrap the weighted origin rainfall measure by drawing a set of weights for the districts within the catchment area from a uniform distribution[9]. If it is the case that our migration rainfall measure is simply proxying for some correlated activity that takes place within the catchment area, we may expect many of these alternate weighting schemes to produce a similar relationship between catchment area rainfall, migration, and labor market outcomes.

Figure 3 demonstrates the F-statistic for migration responses to catchment area rainfall using 10000 bootstrapped weights. While the empirical F-statistic using actual migration patterns shown in 2 equals 21.04, the largest of the 10000 bootstrapped F-statistics is under 3. In other words, while migration patterns are strongly related to rainfall at the origins weighted by the places that migrants actually come from, it is not strongly related to rainfall at alternate weighting schemes within the catchment area. The migration weights seem critical for migration patterns, which is reassuring. Figure 4 presents the reduced-form coefficients from destination wages on catchment area rainfall. When using the actual migration weights, this coefficient is 0.077. The distribution of coefficients using the bootstrapped weights is nearly always small and positive, suggesting that any effects of local rainfall within the catchment area are positive. However, in 10000 bootstrapped replications, they are never as large as the coefficient using actual migration patterns.

From this analysis, we infer that the correlation between origin area rainfall and destination labor market effects are largest for the parts of the catchment area which send the most migrants to the destination in response to rainfall shocks. This analysis rules out the possibility that destination labor markets are affected by origin-area rainfall if these labor market effects follow a different pattern than migration.

6.3 Alternative origin area weights

The first stage analysis uses weights w_o according to equation (1) to indicate the relative importance of an origin area to the destination area under consideration. For each destination, the weights are calculated by dividing the number of migrants from a certain origin area o who reside at destination area d, by all migrants at destination area d. As a robustness check, we calculate weights using auxiliary data from the Intercensal Population Surveys (SUPAS), which is carried out in the mid-period between two population censuses. We use the 1985 SUPAS which is complete just before the start of our panel in 1988, and that reports the current and birth location of 605,858 individuals. The sampling frame of the SUPAS is unfortunately different from that of the IFLS and as a result only 74 percent of locations can be matched. While this reduces power, we nonetheless show in 10 that the second stage results are consistent with our main findings in Table 5.

[9] We impose that the total catchment area weights sum to 1. This is necessary to preserve the magnitude of the independent variable. Since this normalization is necessary, it is not possible to preserve the distribution of underlying weights, which motivates this methodological choice.

6.4 Alternative weather measures

So far, the analyses have used precipitation z-score as weather variable and have thereby implicitly assumed a linear relation between rainfall z-scores and economic outcomes. Robustness checks show that results are robust to using precipitation levels, deviations from the mean, adding squared precipitation, and adding temperature (results not shown). If dummies are used for extreme weather events like droughts and floods, the results become less precise but are still broadly consistent with the findings of this paper. One concern is the possibility that a few extreme weather events play a strong role in inference. In particular, Indonesia was impacted by an extreme drought which was coincident with the financial crisis of 1997. To ensure that extreme weather events in this year are not driving the results, we repeated all analyses excluding 1997 from the sample and excluding 1997 and 1998. The results remain broadly unchanged and increases slightly in magnitude[10].

6.5 Sample definition

Various robustness checks have been carried out which alternative sample definitions, none of which significantly affect the results. Instead of carrying out all analysis at the individual level, the dataset can be collapsed to the destination level in order to study the effect of migration on average income and employment at the destination. This does not significantly alter the results. Several districts in our sample were not part of the original IFLS sample, but were added as respondents were tracked over time. The results do not change when we only use the original IFLS sample. Some of the new districts only host a few IFLS respondents in a given year. When leaving out districts with less than 3 respondents in a year, the results do not change either.

7 Conclusion

This paper employs an instrumental variable approach to study the labor market response to immigration in Indonesia. Exogenous variation in the number of immigrants arriving at a destination is obtained from rainfall shocks at their areas of origin. This paper finds a strong and robust first stage relationship, indicating that people are more likely to leave areas after experiencing a bad weather shock. The second stage confirms predictions from economic theory that increased immigration tends to lower income and employment. Point estimates from this study indicate that a one percentage point increase in the share of migrants at the destination decreases average income per hour by 1.22 percent and reduces the percentage of people employed by 0.26 percentage points. The analysis shows that the negative income effects are concentrated in the informal sector with a 2.45 percent decrease in informal sector income, while employment effects are strongest in the formal sector at 0.32 percentages point, both following a one percentage point increase in the migrant share. These effects are what we would expect in Indonesia's two-sector labor market, as employment is

[10] When excluding 1997, a one percentage point increase in the share of migrants decreases income by 1.29 percent (s.e. 0.598) and reduces employment by 0.33 percentage points (s.e. 0.157). When excluding 1997 and 1998, income reduces by 1.65 percent (s.e. 0.678) and employment decreases 0.40 percentage point (s.e. 0.175).

the primary mechanism for adjustment in the heavily-regulated formal sector, while wages should adjust in the more competitive informal sector.

Exploring heterogeneous labor market effects reveals that the negative consequences are not evenly distributed across subgroups of the population, but are most pronounced for low-skilled individuals Previous studies have attributed disproportional negative impacts on low-skilled natives to a high degree of substitutability with incoming migrants. This argument does not hold for our sample, as migrants have higher levels of education than natives. In Section 5, we find little evidence that this result could be explained by differences in the returns to skill for migrants or by distinctions in migrant characteristics among LATE compliers. Instead we suggest that this result may be understood as another consequence of the two-sector labor market with a wage floor in the formal sector where the disadvantaged group faces changes of disemployment or employment to the informal sector. Short of a conclusive test for this explanation, we suggest that a fruitful avenue for further research is a continued investigation of whether migration more adversely affects similar natives or the most disadvantaged individuals in developing country labor markets.

Finally, it is important to note that this paper considers short-term impacts only. If labor demand can be approximated to be fixed in the short run, then increased labor supply will drive down wages and employment. In the long run, however, labor markets may adjust to the migration-induced increase of labor by expanding production or adjusting the production input mix, which may mitigate or even cancel out short-term economic losses. Without a suitable long-run migration instrument, we can only speculate as to these dynamic patterns.

References

Alatas, Vivi and Lisa Cameron (2008) "The Impact of Minimum Wages on Employment in a Low-Income Country: A Quasi-Natural Experiment in Indonesia," *Industrial and Labor Relations Review*, Vol. 61, pp. 201–223.

Angrist, Joshua and Adriana Kugler (2001) "Protective or counter-productive? Labor market institutions and the effect of immigration on EU natives," Economics Working Papers 587, Department of Economics and Business, Universitat Pompeu Fabra.

Aydemir, Abdurrahman and George Borjas (2010) "Attenuation Bias in Measuring the Wage Impact of Immigration," NBER Working Papers 16229, National Bureau of Economic Research, Inc.

Bauer, Thomas (1998) "Do Immigrants Reduce Natives' Wages? Evidence from Germany," Departmental Working Papers 199802, Rutgers University, Department of Economics.

Bazzi, Sam (2014) "Wealth Heterogeneity and the Income Elasticity of Migration," *working paper*.

Bell, Martin and Salut Muhidin (2009) "Cross-National Comparisons of Internal Migration," Human Development Research Papers (2009 to present) HDRP-2009-30, Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

Borjas, George (1987) "Immigrants, minorities, and labor market competition," *Industrial and Labor Relations Review*, Vol. 40, pp. 382–392.

——— (1995) "The Economic Benefits from Immigration," *Journal of Economic Perspectives*, Vol. 9, pp. 3–22.

——— (1999) "The economic analysis of immigration," in O. Ashenfelter and D. Card eds. *Handbook of Labor Economics*, Vol. 3 of Handbook of Labor Economics: Elsevier, Chap. 28, pp. 1697–1760.

——— (2003) "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," NBER Working Papers 9755, National Bureau of Economic Research, Inc.

Borjas, George, Richard Freeman, and Lawrence Katz (1991) "On the Labor Market Effects of Immigration and Trade," NBER Working Papers 3761, National Bureau of Economic Research, Inc.

Boustan, Leah, Price Fishback, and Shawn Kantor (2010) "The Effect of Internal Migration on Local Labor Markets:American Cities during the Great Depression," *Journal of Labor Economics*, Vol. 28, pp. 719–746, URL: https://ideas.repec.org/a/ucp/jlabec/v28y2010i4p719-746.html.

Bryant, John and Pungpond Rukumnuaykit (2007) "Labor Migration in the Greater Mekong Sub-region: Does Migration to Thailand Reduce the Wages of Thai Workers?" *Washington DC: World Bank*.

Card, David (1990) "The impact of the Mariel boatlift on the Miami labor market," *Industrial and Labor Relations Review*, Vol. 43, pp. 245–257.

——— (2001) "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, Vol. 19, pp. 22–64.

Card, David and John DiNardo (2000) "Do Immigrant Inflows Lead to Native Outflows?," *American Economic Review*, Vol. 90, pp. 360–367.

Card, David and Thomas Lemieux (2001) "Can Falling Supply Explain The Rising Return To College For Younger Men? A Cohort-Based Analysis," *The Quarterly Journal of Economics*, Vol. 116, pp. 705–746.

Card, David and Ethan Lewis (2005) "The Diffusion of Mexican Immigrants During the 1990s: Explanations and Impacts," Working Paper 11552, National Bureau of Economic Research.

Conley, Timothy (1999) "GMM estimation with cross sectional dependence," *Journal of Econometrics*, Vol. 92, pp. 1–45.

Deb, Partha and Papa Seck (2009) "Internal Migration, Selection Bias and Human Development: Evidence from Indonesia and Mexico," Human Development Research Papers (2009 to present) HDRP-2009-31, Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

Frankenberg, Elizabeth and Lynn Karoly (1995) "The 1993 Indonesian Family Life Survey: Overview and Field Report," *RAND, Santa Monica, CA.*

Frankenberg, Elizabeth and Duncan Thomas (2000) "The Indonesia Family Life Survey (IFLS): Study Design and Results from Waves 1 and 2," *RAND, Santa Monica, CA. DRU-2238/1-NIA/NICHD.*

Friedberg, Rachel (2001) "The Impact Of Mass Migration On The Israeli Labor Market," *The Quarterly Journal of Economics*, Vol. 116, pp. 1373–1408.

Grossman, Jean (1982) "The Substitutability of Natives and Immigrants in Production," *The Review of Economics and Statistics*, Vol. 64, pp. 596–603.

Harris, John and Michael Todaro (1970) "Migration, Unemployment & Development: A Two-Sector Analysis," *American Economic Review*, Vol. 60, pp. 126–42.

Hicks, Joan Hamory, Marieke Kleemans, and Edward Miguel (2014) "Individual Ability and Selection into Migration in Kenya,"Technical Report 19228.

Kerr, Sari Pekkala and William Kerr (2011) "Economic Impacts of Immigration: A Survey," *Finnish Economic Papers*, Vol. 24, pp. 1–32.

King, Allan, Lindsay Lowell, and Frank Bean (1986) "The effects of Hispanic immigration on the earnings of native Hispanic Americans," December.

.

Kishore, Kamal, A.R. Subbiah, Tien Sribimawati, Sri Diharto, Sutarto Alimoeso, Peter Rogers, and Adang Setiana (2000) *Indonesia Country Study*.

Kleemans, Marieke (2015) "Migration Choice under Risk and Liquidity Constraints."

Kleemans, Marieke and Jeni Klugman (2009) "Understanding Attitudes Towards Migrants: A Broader Perspective," *Human Development Research Papers*.

Kugler, Adriana and Mutlu Yuksel (2008) "Effects of Low-Skilled Immigration on U.S. Natives: Evidence from Hurricane Mitch," NBER Working Papers 14293, National Bureau of Economic Research, Inc.

Levine, David and Dean Yang (2006) "The Impact of Rainfall on Rice Output in Indonesian Districts," *Mimeo,University of California, Berkeley and University of Michigan*.

Longhi, S., P. Nijkamp, and J. Poot (2010) "Joint impacts of immigration on wages and employment: review and meta-analysis," *Journal of Geographical Systems*, Vol. 12, pp. 355–387.

Maccini, Sharon and Dean Yang (2009) "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall," *American Economic Review*, Vol. 99, pp. 1006–26.

Magruder, Jeremy (2013) "Can minimum wages cause a big push? Evidence from Indonesia," *Journal of Development Economics*, Vol. 100, pp. 48–62.

Matsuura, Kenji and Cort Willmott (2009) "Terrestrial Air Temperature and Precipitation: Monthly Climatologies," *Center for Climatic Research, Department of Geography, University of Delaware*.

Munshi, Kaivan (2003) "Networks In The Modern Economy: Mexican Migrants In The U.S. Labor Market," *The Quarterly Journal of Economics*, Vol. 118, pp. 549–599.

Okkerse, Liesbet (2008) "How To Measure Labour Market Effects Of Immigration: A Review," *Journal of Economic Surveys*, Vol. 22, pp. 1–30.

Olken, Benjamin (2009) "Merged Indonesia Kecamatam Codes," *NBER electronic data set*.

Peri, Giovanni (2007) "Immigrants' Complementarities and Native Wages: Evidence from California," NBER Working Papers 12956, National Bureau of Economic Research, Inc.

Programme, United Nations Development (2009a) *HDR 2009 - Overcoming barriers: Human mobility and development* in , Human Development Report (1990 to present), No. hdr2009: Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

——— (2009b) *HDR 2009 - Overcoming barriers: Human mobility and development* in , Human Development Report (1990 to present), No. hdr2009: Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

Pugatch, Todd and Dean Yang (2011) "The Impact of Mexican Immigration on U.S. Labor Markets: Evidence from Migrant Flows Driven by Rainfall Shocks."

Strauss, John, Kathleen Beegle, Bondan Sikoki, Agus Dwiyanto, Yulia Herawati, and Firman Witoelar (2004) "The Third Wave of the Indonesia Family Life Survey (IFLS): Overview and Field Report," *WR-144/1-NIA/NICHD*.

Strobl, Eric and Marie-Anne Valfort (2013) "The effect of weather-induced internal migration on local labor markets : evidence from Uganda," Policy Research Working Paper Series 6600, The World Bank.

Thomas, Duncan, Elizabeth Frankenberg, and James Smith (2001) "Lost but Not Forgotten: Attrition and Follow-up in the Indonesia Family Life Survey," *Journal of Human Resources*, Vol. 36, pp. 556–592.

Tables and Figures

Table 1: Summary Statistics

	All	Natives	Migrants	t-statistic of mean comparison test
Number of individuals	28,841	-	-	
Migrant stock at destination (%)	15.90 (17.71)	-	-	
Precipitation (mm per month)	172.61 (48.92)	-	-	
Male	0.52 0.50	0.51 -0.50	0.54 -0.50	-8.41
Education level	1.64 (1.21)	1.57 (1.18)	1.93 (1.26)	-73.25
Highly-educated dummy	0.47 (0.49)	0.45 (0.49)	0.56 (0.49)	-62.66
Age	36.97 (15.11)	37.05 (15.46)	36.69 (13.83)	4.34
Log hourly income (Rp)	6.90 (1.18)	6.84 (1.18)	7.06 (1.14)	-28.74
Log hourly income formal sector (Rp)	7.08 (1.11)	7.03 (1.12)	7.23 (1.05)	-22.28
Log hourly income informal sector (Rp)	6.90 (1.32)	6.88 (1.32)	7.00 (1.31)	-10.31
Employment rate (%)	78.71 (40.93)	78.67 (40.96)	78.88 (40.82)	-0.96
Formal employment rate (%)	39.74 (48.94)	37.46 (48.40)	47.49 (49.94)	-37.80
Informal employment rate (%)	51.89 (49.96)	54.32 (49.81)	43.66 (49.60)	39.34
Hours per week	37.06 (21.48)	35.93 (21.51)	40.88 (20.95)	-38.18
Number of individual-year pairs	192,522	148,708	43,751	

Dataset consists of individual-year observations, each of which is counted seperately, based on the information of 28,841 individuals who may change migrant status over time. Mean is shown with standard deviation between brackets. Values for education level are: 0 = no education, 1 = primary school, 2 = junior high school, 3 = senior high school, 4 = university. A person is classified as highly-educated if he or she has at least some high school education (education level 2 or higher). Log income per hour is measured in Indonesian Rupiah and employment rate is the percentage of individuals employed. All employment is included and some individuals have a job in the formal sector as well as a job in the informal sector.

Table 2: Migrants' Responsiveness to Weather Shocks (First Stage)

	Dependent variable: Migrant Share of the Population				
	(1)	(2)	(3)	(4)	(5)
Precipitation	-0.050***		-0.037***		-0.041***
	[0.015]		[0.011]		[0.012]
Precipitation lagged		-0.069***	-0.062***	-0.064***	-0.055***
		[0.015]	[0.013]	[0.014]	[0.012]
Precipitation lagged twice				-0.024***	-0.029***
				[0.007]	[0.007]
Precipitation at destination	0.038***		0.028***		0.031***
	[0.013]		[0.010]		[0.010]
Precipitation at destination lagged		0.051***	0.045***	0.046***	0.039***
		[0.013]	[0.011]	[0.012]	[0.010]
Precipitation at destination lagged twice				0.022***	0.026***
				[0.005]	[0.005]
F statistic of joint significance	**11.66**	**21.04**	**11.39**	**12.13**	**8.78**
Time fixed effects	yes	yes	yes	yes	yes
Destination fixed effects	yes	yes	yes	yes	yes
Socio-Economic control variables	yes	yes	yes	yes	yes
R-squared	0.29	0.31	0.31	0.31	0.32
Observations	148,565	148,565	148,565	148,565	148,565
Number of destinations	205	205	205	205	205

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. Unless noted otherwise, precipitation variables refer to origin area precipitation z-scores and are summed over the entire catchment area of each destination according to equation (1). Socio-economic control variables include dummies for gender, education level and age group.

Table 3: Rainfall and Labor Market Effects (Reduced Form)

	Dependent variable:	
	Log income per hour	**Employment**
	(1)	(2)
Precipitation lagged	0.077***	0.018**
	[0.028]	[0.008]
Precipitation at destination lagged	-0.02	-0.020***
	[0.016]	[0.005]
Time fixed effects	yes	yes
Destination fixed effects	yes	yes
Socio-Economic control variables	yes	yes
R-squared	0.43	0.19
Observations	100,616	148,300
Number of destinations	205	205

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. Unless noted otherwise, precipitation variables refer to origin area precipitation z-scores and are summed over the entire catchment area of each destination according to equation (1). Log income per hour is measured in Indonesian Rupiah and employment is the share of individuals employed. The number of observations is lower for income because this is only measured if a person is employed. Socio-economic control variables include dummies for gender, education level and age group.

Table 4: Labor Market Response to Immigration (Second Stage)

	Dependent variable:			
	Log income per hour		**Employment**	
	OLS	IV-2SLS	OLS	IV-2SLS
	(1)	(2)	(3)	(4)
Migrant share predicted from the first stage	0.02	-1.22**	-0.03	-0.26**
	[0.091]	[0.530]	[0.026]	[0.124]
Precipitation destination lagged	0.03**	0.04**	-0.01***	-0.01**
	[0.011]	[0.015]	[0.003]	[0.004]
F statistic of first stage relationship		20.60		20.98
Time fixed effects	yes	yes	yes	yes
Destination fixed effects	yes	yes	yes	yes
Socio-Economic control variables	yes	yes	yes	yes
R-squared	0.43	0.43	0.19	0.19
Observations	100,616	100,616	148,300	148,300
Number of destinations	205	205	205	205

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. (1) and (3) use OLS; (2) and (4) use IV-2SLS with lagged precipitation z-scores as instrument for migrant share. Log income per hour is measured in Indonesian Rupiah and employment is the share of individuals employed. The number of observations is lower for income because this is only measured if a person is employed. Socio-economic control variables include dummies for gender, education level and age group.

Table 5: Labor Market Response in Formal and Information Sector

Panel A. Dependent variable: Log income per hour

	Overall	Formal sector	Informal sector
	(1)	(2)	(3)
Migrant share predicted from the first stage	-1.23**	-0.39	-2.45**
	[0.531]	[0.417]	[0.970]
Precipitation destination lagged	0.04**	0.02	0.05**
	[0.015]	[0.014]	[0.022]
F statistic of first stage relationship	20.48	19.34	19.78
Time fixed effects	yes	yes	yes
Destination fixed effects	yes	yes	yes
Socio-Economic control variables	yes	yes	yes
R-squared	0.43	0.45	0.37
Observations	100,549	53,979	60,858
Number of destinations	190	190	190

Panel B. Dependent variable: Employment

	Overall	Formal sector	Informal sector
	(1)	(2)	(3)
Migrant share predicted from the first stage	-0.26**	-0.32***	-0.02
	[0.124]	[0.107]	[0.131]
Precipitation destination lagged	-0.01**	0	-0.01
	[0.004]	[0.004]	[0.005]
F statistic of first stage relationship	20.92	20.92	20.92
Time fixed effects	yes	yes	yes
Destination fixed effects	yes	yes	yes
Socio-Economic control variables	yes	yes	yes
R-squared	0.19	0.10	0.12
Observations	148,188	148,161	148,161
Number of destinations	190	190	190

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. All specifications use IV-2SLS with lagged precipitation z-scores as instrument for migrant share. Log income per hour is measured in Indonesian Rupiah and employment is the share of individuals employed. The number of observations is slightly smaller than in Table 4 in order to ensure that all destinations have sufficient observations across sectors. The number of observations is lower for income because this is only measured if a person is employed. Socio-economic control variables include dummies for gender, education level and age group.

Table 6: Heterogeneous Labor Market Impacts by Sector

Panel A. Dependent variable: Log income per hour

	All		Formal sector		Informal sector	
	Low Education	High Education	Low Education	High Education	Low Education	High Education
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant share predicted from the first stage	-2.33**	-0.25	-1.23	-0.08	-3.39***	-1.04
	[0.906]	[0.432]	[0.877]	[0.373]	[1.192]	[1.140]
Precipitation destination lagged	0.03	0.02	0.01	0.01	0.03	0.06
	[0.017]	[0.019]	[0.019]	[0.019]	[0.023]	[0.036]
F statistic of first stage relationship	16.76	22.37	11.37	22.52	17.00	20.18
Time fixed effects	yes	yes	yes	yes	yes	yes
Destination fixed effects	yes	yes	yes	yes	yes	yes
Socio-Economic control variables	yes	yes	yes	yes	yes	yes
R-squared	0.38	0.44	0.40	0.48	0.34	0.36
Observations	67,372	31,902	31,272	22,182	45,936	14,043
Number of destinations	173	173	173	173	173	173

Panel B. Dependent variable: Employment

	All		Formal sector		Informal sector	
	Low Education	High Education	Low Education	High Education	Low Education	High Education
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant share predicted from the first stage	-0.36	-0.04	-0.28*	-0.17	-0.14	0.11
	[0.230]	[0.121]	[0.159]	[0.138]	[0.212]	[0.140]
Precipitation destination lagged	-0.01***	0	0	0	-0.01	-0.01
	[0.004]	[0.006]	[0.005]	[0.006]	[0.005]	[0.008]
F statistic of first stage relationship	15.71	22.90	15.71	22.93	15.71	22.93
Time fixed effects	yes	yes	yes	yes	yes	yes
Destination fixed effects	yes	yes	yes	yes	yes	yes
Socio-Economic control variables	yes	yes	yes	yes	yes	yes
R-squared	0.13	0.26	0.09	0.15	0.09	0.09
Observations	95,523	50,529	95,508	50,517	95,508	50,517
Number of destinations	173	173	173	173	173	173

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. All specifications use IV-2SLS with lagged precipitation z-scores as instrument for migrant share. Log income per hour is measured in Indonesian Rupiah and employment is the share of individuals employed. Low-skilled individuals are defined as those with no or only primary education and high-skilled individuals are those with at least some secondary education. The number of observations is smaller than in Table 5 in order to ensure that all destinations have sufficient low and high-skilled individuals in each sector. The number of observations is lower for income because this is only measured if a person is employed. Socio-economic control variables include dummies for gender, education level and age group.

Table 7: Mincer Regressions

	Dependent variable: Log income per hour					
	(1)	(2)	(3)	(4)	(5)	(6)
	All	**Natives**	**Migrants**	**All**	**Natives**	**Migrants**
Education level	0.297***	0.282***	0.317***			
	[0.009]	[0.011]	[0.014]			
Years of education				0.078***	0.073***	0.087***
				[0.002]	[0.003]	[0.004]
Male	0.274***	0.288***	0.228***	0.267***	0.286***	0.208***
	[0.019]	[0.024]	[0.028]	[0.020]	[0.024]	[0.027]
Age	0.045***	0.040***	0.064***	0.047***	0.041***	0.064***
	[0.003]	[0.003]	[0.006]	[0.003]	[0.003]	[0.006]
Age squared	-0.000***	-0.000***	-0.001***	-0.000***	-0.000***	-0.001***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Time fixed effects	yes	yes	yes	yes	yes	yes
Destination fixed effects	yes	yes	yes	yes	yes	yes
R-squared	0.44	0.43	0.43	0.44	0.43	0.44
Observations	131,953	100,616	31,295	131,902	100,588	31,279
Number of destinations	205	205	205	205	205	205

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. Log income per hour is measured in Indonesian Rupiah and education values stand for: 0 = no education, 1 = primary school, 2 = junior high school, 3 = senior high school, 4 = university.

Table 8: Quantifying the Local Average Treatment Effect

	Sample: Those who arrived at a destination as a migrant Dependent variable: Being highly-educated				
	(1)	(2)	(3)	(4)	(5)
Precipitation		0.034	0.034		0.046**
		[0.023]	[0.023]		[0.023]
Precipitation lagged	0.012		0.008	0.001	-0.01
	[0.034]		[0.034]	[0.033]	[0.033]
Precipitation lagged twice				0.03	0.044
				[0.026]	[0.027]
Time fixed effects	yes	yes	yes	yes	yes
Destination fixed effects	yes	yes	yes	yes	yes
R-squared	0.02	0.02	0.02	0.02	0.02
Observations	3,858	3,858	3,858	3,858	3,858
Number of destinations	198	198	198	198	198

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. Precipitation variables refer to origin area precipitation z-scores and are summed over the entire catchment area of each destination according to equation (1).

Table 9: Long Distance Migration

	Migrants' responsiveness to weather shocks		Labor market response to immigration			
Dependent variable:	Migrant share		Log income per hour		Employment	
	All migration	**Migration > 100 km**	**All migration**	**Migration > 100 km**	**All migration**	**Migration > 100 km**
	(1)	(2)	(3)	(4)	(5)	(6)
Precipitation lagged	-0.069***	-0.022***				
	[0.015]	[0.005]				
Migrant share predicted from the first stage			-1.18**	-2.32**	-0.23*	-0.62**
			[0.529]	[1.165]	[0.122]	[0.288]
Precipitation at destination lagged	0.051***	0.018***	0.03**	0.05**	-0.01**	0
	[0.013]	[0.005]	[0.015]	[0.020]	[0.004]	[0.005]
Mean dependent variable	15.9%	7.4%	6.17	6.17	78.7%	78.7%
F statistic of first stage relationship	20.82	16.72	20.37	18.74	20.76	16.69
Time fixed effects	yes	yes	yes	yes	yes	yes
Destination fixed effects	yes	yes	yes	yes	yes	yes
Socio-Economic control variables	yes	yes	yes	yes	yes	yes
R-squared	0.31	0.20	0.43	0.428	0.19	0.187
Observations	147,098	147,098	99,636	99,636	146,835	146,835
Number of destinations	198	198	198	198	198	198

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. Unless noted otherwise, precipitation variables refer to origin area precipitation z-scores and are summed over the entire catchment area of each destination according to equation (1). Columns (2), (4) and (6) only consider migration over at least 100 km distance. Columns (3), (4), (5) and (6) use IV-2SLS with lagged precipitation z-scores as instrument for migrant share. Log income per hour is measured in Indonesian Rupiah and employment is the share of individuals employed. The number of observations is lower for income because this is only measured if a person is employed. Socio-economic control variables include dummies for gender, education level and age group.

Table 10: Labor Market Response using Origin Area Weights from the Intercensal Population Survey

Panel A. Dependent variable: Log income per hour

	Overall	Formal sector	Informal sector
	(1)	(2)	(3)
Migrant share predicted from the first stage	-0.76*	-0.01	-1.68**
	[0.395]	[0.344]	[0.697]
Precipitation destination lagged	0.03**	0.01	0.04**
	[0.013]	[0.014]	[0.019]
F statistic of first stage relationship	23.9	21.87	21.98
Time fixed effects	yes	yes	yes
Destination fixed effects	yes	yes	yes
Socio-Economic control variables	yes	yes	yes
R-squared	0.43	0.45	0.38
Observations	99,520	53,399	60,282
Number of destinations	182	182	182

Panel B. Dependent variable: Employment

	Overall	Formal sector	Informal sector
	(1)	(2)	(3)
Migrant share predicted from the first stage	-0.11	-0.25***	0.07
	[0.100]	[0.094]	[0.115]
Precipitation destination lagged	-0.01**	0	-0.01*
	[0.003]	[0.004]	[0.005]
F statistic of first stage relationship	25.1	25.08	25.08
Time fixed effects	yes	yes	yes
Destination fixed effects	yes	yes	yes
Socio-Economic control variables	yes	yes	yes
R-squared	0.19	0.11	0.12
Observations	146,597	146,570	146,570
Number of destinations	182	182	182

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. All specifications use IV-2SLS with lagged precipitation z-scores as instrument for migrant share. Log income per hour is measured in Indonesian Rupiah and employment is the share of individuals employed. The number of observations is slightly smaller than in Table 4, to ensure that all destinations have sufficient observations across sectors. The number of observations is lower for income because this is only measured if a person is employed. Socio-economic control variables include dummies for gender, education level and age group.

Figure 1: Household locations IFLS with weather data that locations are mapped to

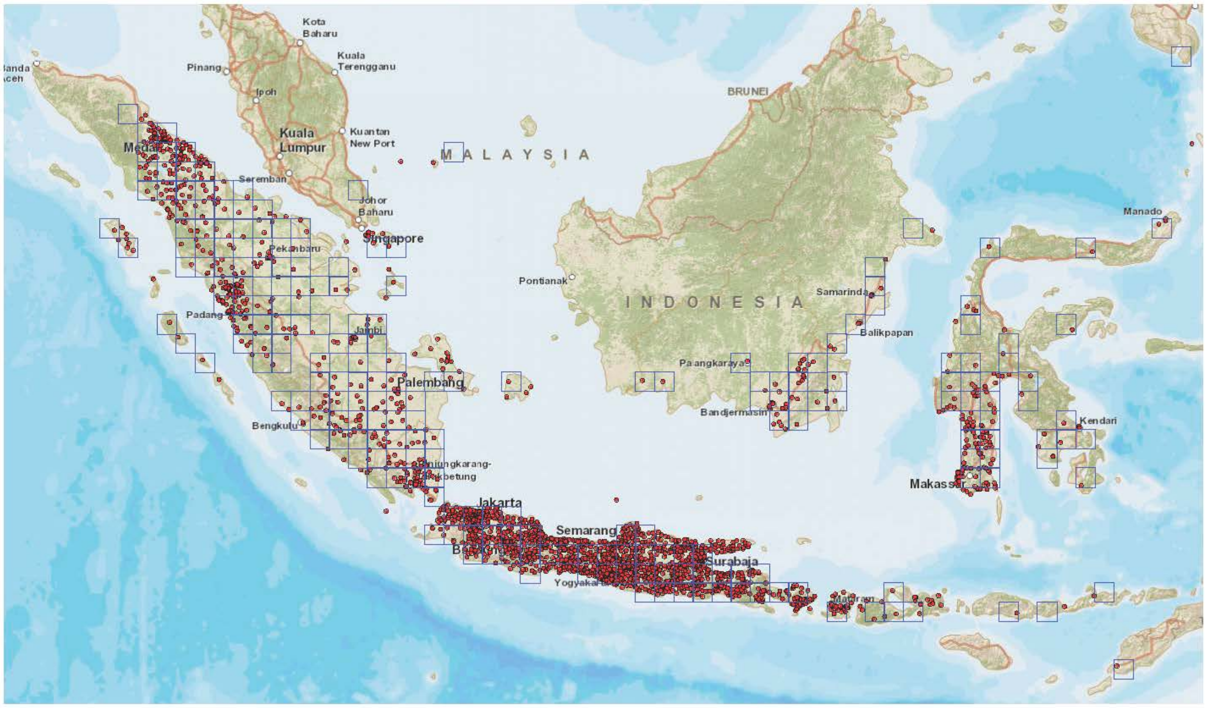


Figure 2: Precipitation (mm per month)

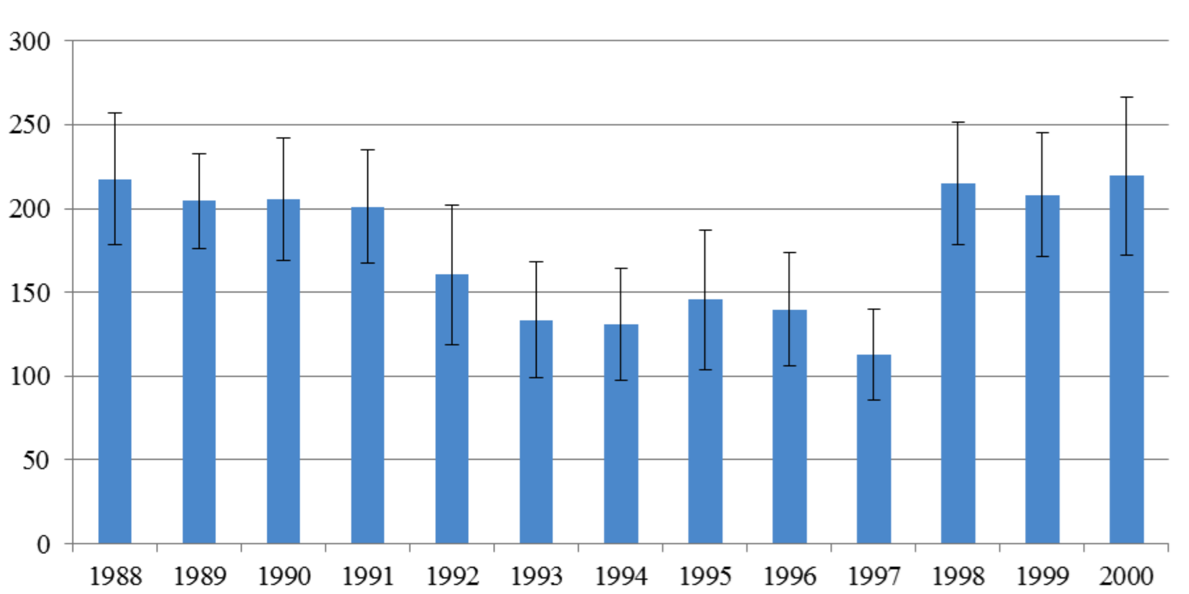


Figure 3: Bootstrapped first stage F-statistics using random rainfall weights

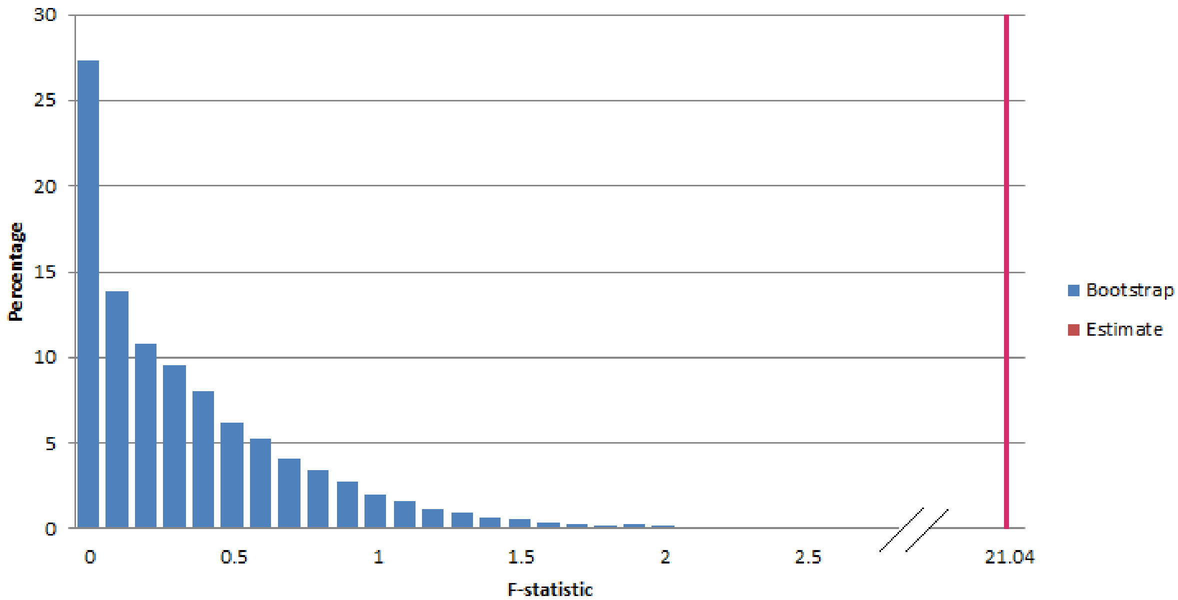
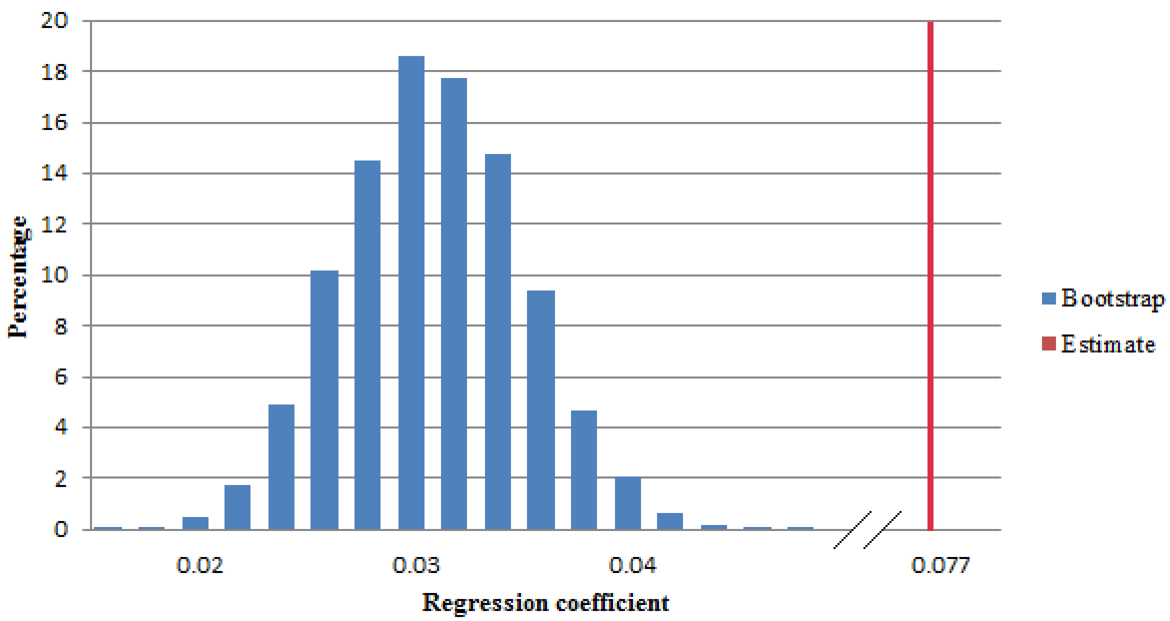


Figure 4: Bootstrapped reduced-form coefficients using random rainfall weights



Appendix

Proof of Statement 1 (Page 15):

$$\frac{\partial w_j}{\partial L_j} < 0$$

$$\frac{\partial w_j}{\partial L_j} = \alpha(\alpha-\nu)\theta_j^2 K^{1-\alpha} L_j^{2\nu-2}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{(\alpha-2\nu)/\nu} - (1-\nu)\alpha\theta_j K^{1-\alpha} L_j^{\nu-2}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{(\alpha-\nu)/\nu} < 0$$

$$(\alpha-\nu)(\theta_j L_j^{\nu}) < (1-\nu)\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)$$

We know that $-1 < (\alpha-\nu) < 0$ and $0 < 1-\nu < 1$ so it must be the case that $(\alpha-\nu) < 1-\nu$. Furthermore, we know that $(\theta_j L_j^{\nu}) < (\sum_{k=1}^{n}\theta_k L_k^{\nu})$. Therefore, we can conclude $\frac{\partial w_j}{\partial L_j} < 0$

Proof of Statement 2 (Page 15):

$$\frac{\partial w_j}{\partial L_g} < 0$$

Note that terms are positive except $(\alpha-\nu)$, which is negative for $\alpha < \nu$. Hence it follows that if $\alpha < \nu \Rightarrow \partial w_g/\partial L_g < 0$.

Proof of Statement 3 (Page 15):

$$\frac{\partial w_j}{\partial L_j} < \frac{\partial w_j}{\partial L_g}$$

$$\alpha(\alpha-\nu)\theta_j^2 K^{1-\alpha} L_j^{2\nu-2}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{(\alpha-2\nu)/\nu} - (1-\nu)\alpha\theta_j K^{1-\alpha} L_j^{\nu-2}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{(\alpha-\nu)/\nu}$$

$$< \frac{\alpha(\alpha-\nu)\theta_j\theta_g K^{1-\alpha}}{L_j^{1-\nu} L_g^{1-\nu}}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{(\alpha-2\nu)/\nu}$$

$$(\alpha-\nu)(\theta_j L_j^{\nu} - \theta_g L_j L_g^{\nu-1}) < (1-\nu)\sum_{k=1}^{n}\theta_k L_k^{\nu}$$

We know that $(\theta_j L_j^{\nu} - \theta_g L_j L_g^{\nu-1}) < \theta_j L_j^{\nu} < \sum_{k=1}^{n}\theta_k L_k^{\nu}$ and we also showed earlier that $(\alpha-\nu) < (1-\nu)$ so the inequality holds.

Derivation of conditions under which $\partial \bar{w}_g/\partial L_j < \partial w_j/\partial L_j$ in the two-sector model (page 19):

The cross derivative is

$$\bar{w}_g = \frac{L_I^{\gamma-1}(L_g - \hat{L}_g) + \hat{w}\hat{L}_g}{L_g}$$

Define $I = L_I^{\gamma-1}$. Then

$$\frac{\partial \bar{w}_g}{\partial L_j} = \frac{\hat{w} - I}{L_g}\frac{\partial \hat{L}_g}{\partial L_j} + \frac{L_g - \hat{L}_g}{L_g}(\gamma - 1)L_I^{\gamma-2}\frac{\partial L_I}{\partial L_j}$$

where

$$\hat{L}_g = \left[\frac{\alpha\theta_g K^{1-\alpha}}{\hat{w}}\right]^{\frac{1}{1-\nu}} \left(\sum_{k=1}^{n} \theta_k L_k^{\nu}\right)^{\frac{(\alpha-\nu)}{\nu(1-\nu)}}$$

since

$$\frac{\partial \hat{L}_g'}{\partial L_j} = \left[\frac{\alpha\theta_{g'} K^{1-\alpha}}{\hat{w}}\right]^{\frac{1}{1-\nu}} \left(\frac{\alpha-\nu}{1-\nu}\right) \frac{\theta_j}{L_j^{1-\nu}} \left(\sum_{k=1}^{n} \theta_k L_k^{\nu}\right)^{\frac{a-\nu-\nu(1-\nu)}{\nu(1-\nu)}} \qquad \forall g'$$

And all components are necessarily positive save $\alpha - \nu$, we have immediately that $\frac{\partial \hat{L}_{g'}}{\partial L_j}$ has the same sign as $\alpha - \nu$

Moreover, since

$$\frac{\partial L_I}{\partial L_j} = -\sum_{g'} \frac{\partial \hat{L}_{g'}}{\partial L_j}$$

We have $\alpha < \nu \Rightarrow \frac{\partial L_I}{\partial L_J} > 0$. Define

$$\xi_j^I = \frac{(L_g - \hat{L}_g)}{L_g}(\gamma - 1)L_I^{\gamma-2}\frac{\partial L_I}{\partial L_j}$$

Which has the opposite sign from $\frac{\partial L_I}{\partial L_j}$.

Now, note that

$$\frac{\partial \bar{w}_g}{\partial L_j} = \frac{\hat{w} - I}{L_g}\frac{\partial \hat{L}_g}{\partial L_j} = \frac{\hat{w} - I}{L_g}\left[\frac{\alpha\theta_g K^{1-\alpha}}{\hat{w}}\right]^{\frac{1}{1-\nu}} \left(\frac{\alpha-\nu}{1-\nu}\right) \frac{\theta_j}{L_j^{1-\nu}} \left(\sum_{k=1}^{n} \theta_k L_k^{\nu}\right)^{\frac{a-\nu-\nu(1-\nu)}{\nu(1-\nu)}} + \xi_j^I$$

We will now derive a bound for the conditions under which

$$\frac{\partial \bar{w}_g}{\partial L_j} < \frac{\partial w_j}{\partial L_j}$$

$$\frac{\hat{w}-I}{L_g}\left[\frac{\alpha\theta_g K^{1-\alpha}}{\hat{w}}\right]^{\frac{1}{1-\nu}}\left(\frac{\alpha-\nu}{1-\nu}\right)\frac{\theta_j}{L_j^{1-\nu}}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{\frac{a-\nu-\nu(1-\nu)}{\nu(1-\nu)}}+\xi_j^I$$

$$<\alpha(\alpha-\nu)\theta_j^2 K^{1-\alpha}L_j^{2\nu-2}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{(a-\nu)/2\nu}$$

$$-(1-\nu)\alpha\theta_j K^{1-\alpha}L_j^{\nu-2}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{(a-\nu)/\nu}$$

Recall that

$$w_j=\alpha\theta_j K^{1-\alpha}L_j^{\nu-1}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{(a-\nu)/\nu}$$

We need to show that

$$\frac{\hat{w}-I}{L_g}\left[\frac{\alpha\theta_g K^{1-\alpha}}{\hat{w}}\right]^{\frac{1}{1-\nu}}\left(\frac{\alpha-\nu}{1-\nu}\right)\frac{\theta_j}{L_j^{1-\nu}}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{\frac{a-\nu-\nu(1-\nu)}{\nu(1-\nu)}}+\xi_j^I<w_j\left[\frac{(\alpha-\nu)\theta_j}{L_j^{1-\nu}\sum_{k=1}^{n}\theta_k L_k^{\nu}}-\frac{1-\nu}{L_j}\right]$$

As shown earlier

$$\hat{w}=\frac{\alpha\theta_j K^{1-\alpha}}{\hat{L}_g^{1-\nu}}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{(\alpha-\nu)/\nu}$$

so that

$$\hat{w}^{1/(1-\nu)}\hat{L}_g=(\alpha\theta_g K^{1-\alpha})^{1/(1-\nu)}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{(\alpha-\nu)/\nu(1-\nu)}$$

Substituting this in gives

$$\frac{\hat{w}-I}{L_g}\left[\frac{\alpha\theta_g K^{1-\alpha}}{\hat{w}}\right]^{\frac{1}{1-\nu}}\left(\frac{\alpha-\nu}{1-\nu}\right)\frac{\theta_j}{L_j^{1-\nu}}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{\frac{a-\nu-\nu(1-\nu)}{\nu(1-\nu)}}+\xi_j^I<\frac{w_j}{L_j}\left[\frac{(\alpha-\nu)\theta_j L_j^{\nu}}{\sum_{k=1}^{n}\theta_k L_k^{\nu}}-(1-\nu)\right]$$

$$\frac{\hat{w}-I}{L_g}\left[\frac{1}{\hat{w}}\right]^{1/(1-\nu)}\left(\frac{\alpha-\nu}{1-\nu}\right)\frac{\theta_j}{L_j^{1-\nu}}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{-1}\hat{L}_g\hat{w}^{1/(1-\nu)}+\xi_j^I<\frac{w_j}{L_j}\left[\frac{(\alpha-\nu)\theta_j L_j^{\nu}}{\sum_{k=1}^{n}\theta_k L_k^{\nu}}-(1-\nu)\right]$$

$$\frac{\hat{w}-I}{L_g}\left(\frac{\alpha-\nu}{1-\nu}\right)\frac{\theta_j}{L_j^{1-\nu}}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{-1}\hat{L}_g+\xi_j^I<\frac{w_j}{L_j}\left[\frac{(\alpha-\nu)\theta_j L_j^{\nu}}{\sum_{k=1}^{n}\theta_k L_k^{\nu}}-(1-\nu)\right]$$

$$\frac{\hat{w}-I}{L_g}\left(\frac{\alpha-\nu}{1-\nu}\right)\theta_j L_j^{\nu}\left(\sum_{k=1}^{n}\theta_k L_k^{\nu}\right)^{-1}\hat{L}_g+L_j\xi_j^I<w_j\left[\frac{(\alpha-\nu)\theta_j L_j^{\nu}}{\sum_{k=1}^{n}\theta_k L_k^{\nu}}-(1-\nu)\right]$$

The inequality holds if:

$$w_j(1-\nu) < w_j \frac{(\alpha-\nu)\theta_j L_j^{\nu}}{\sum_{k=1}^{n} \theta_k L_k^{\nu}} - \frac{\hat{w}-I}{L_g} \left(\frac{\alpha-\nu}{1-\nu}\right) \frac{\hat{L}_g L_j^{\nu} \theta_j}{(\sum_{k=1}^{n} \theta_k L_k^{\nu})} - L_j \xi_j^I$$

or if

$$w_j(1-\nu) < \frac{(\alpha-\nu)\theta_j L_j^{\nu}}{\sum_{k=1}^{n} \theta_k L_k^{\nu}} \left[w_j - \left(\frac{\hat{w}-I}{1-\nu}\right) \frac{\hat{L}_g}{L_g}\right] - L_j \xi_j^I$$

$\gamma = 1 \Rightarrow \xi_j^I = 0$ and $\gamma < 1 \Rightarrow \xi_j^I < 0$ which gives the result in the paper.

Table A1: Migrants' Responsiveness to Weather Shocks

	Dependent variable: Migrant Share of the Population			
	(1)		(2)	(3)
Precipitation-zscores lagged	-0.069***		-0.060***	
	[0.015]		[0.014]	
Precipitation-levels lagged		-0.159***		-0.135***
		[0.039]		[0.035]
Precipitation-zscores at destination lagged	0.051***		0.046***	
	[0.013]		[0.012]	
Precipitation-levels at destination lagged		0.116***		0.102***
		[0.032]		[0.029]
F statistic of joint significance	**21.04**	**17.07**	**19.25**	**15.10**
Time fixed effects	yes	yes	yes	yes
Destination fixed effects	yes	yes	no	no
Individual fixed effects	no	no	yes	yes
Socio-Economic control variables	yes	yes	no	no
R-squared	0.31	0.31	0.34	0.34
Observations	148,565	148,565	148,607	148,607
Number of destinations	205	205		
Number of individuals			24,164	24,164

All regressions are clustered at the destination level, *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Unless noted otherwise, precipitation variables refer to origin area precipitation and are summed over the entire catchment area of each destination according to equation (1). Precipitation-levels are reported in decimeter of precipitation per month. Socio-economic control variables include dummies for gender, education level and age group.

Table A2: Labor Market Response in Formal and Informal Sector with Precipitation Levels

Panel A. Dependent variable: Log income per hour

	Overall	Formal sector	Informal sector
	(1)	(2)	(3)
Migrant share predicted from the first stage	-1.30**	-0.36	-2.72***
	[0.518]	[0.368]	[1.051]
Precipitation-levels lagged at destination	0.08**	0.05*	0.09**
	[0.032]	[0.032]	[0.047]
F statistic of first stage relationship	15.90	15.80	14.48
Time fixed effects	yes	yes	yes
Destination fixed effects	yes	yes	yes
Socio-Economic control variables	yes	yes	yes
R-squared	0.43	0.45	0.37
Observations	100,549	53,979	60,858
Number of destinations	190	190	190

Panel B. Dependent variable: Employment

	Overall	Formal sector	Informal sector
	(1)	(2)	(3)
Migrant share predicted from the first stage	-0.24**	-0.29**	0.01
	[0.114]	[0.117]	[0.127]
Precipitation-levels lagged at destination	-0.01	0	-0.01
	[0.008]	[0.009]	[0.009]
F statistic of first stage relationship	16.92	16.92	16.92
Time fixed effects	yes	yes	yes
Destination fixed effects	yes	yes	yes
Socio-Economic control variables	yes	yes	yes
R-squared	0.19	0.11	0.12
Observations	148,188	148,161	148,161
Number of destinations	190	190	190

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. All specifications use IV-2SLS with lagged precipitation levels as instrument for migrant share. Log income per hour is measured in Indonesian Rupiah and employment is the share of individuals employed. The number of observations is lower for income because this is only measured if a person is employed. Socio-economic control variables include dummies for gender, education level and age group.

Table A3: Labor Market Response in Formal and Informal Sector with Individual Fixed Effects

Panel A. Dependent variable: Log income per hour

	Overall	Formal sector	Informal sector
	(1)	(2)	(3)
Migrant share predicted from the first stage	-1.36**	-0.15	-2.97***
	[0.598]	[0.368]	[1.106]
Precipitation-zscores at destination lagged	0.03**	0.01	0.04*
	[0.013]	[0.011]	[0.020]
F statistic of first stage relationship	18.79	16.66	19.19
Time fixed effects	yes	yes	yes
Individual fixed effects	yes	yes	yes
Socio-Economic control variables	no	no	no
R-squared	0.52	0.56	0.45
Observations	98,951	52,420	60,012
Number of individuals	14,420	8,938	8,566

Panel B. Dependent variable: Employment

	Overall	Formal sector	Informal sector
	(1)	(2)	(3)
Migrant share predicted from the first stage	-0.16	-0.19*	-0.01
	[0.117]	[0.115]	[0.118]
Precipitation-zscores at destination lagged	-0.01***	0	-0.01**
	[0.003]	[0.003]	[0.004]
F statistic of first stage relationship	19.17	19.16	19.16
Time fixed effects	yes	yes	yes
Individual fixed effects	yes	yes	yes
Socio-Economic control variables	no	no	no
R-squared	0.02	0.01	0.01
Observations	144,672	144,643	144,643
Number of individuals	20,563	20,561	20,561

All regressions are clustered at the destination level, *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All specifications use IV-2SLS with lagged precipitation z-scores as instrument for migrant share. Log income per hour is measured in Indonesian Rupiah and employment is the share of individuals employed. The number of observations is lower for income because this is only measured if a person is employed. Socio-economic control variables include dummies for gender, education level and age group.

Table A4: Labor Market Response in Formal and Informal Sector with Precipitation Levels and Individual Fixed Effects

Panel A. Dependent variable: Log income per hour

	Overall	Formal sector	Informal sector
	(1)	(2)	(3)
Migrant share predicted from the first stage	-1.40**	-0.14	-3.25**
	[0.630]	[0.358]	[1.280]
Precipitation-levels lagged at destination	0.07**	0.02	0.07*
	[0.030]	[0.022]	[0.044]
F statistic of first stage relationship	13.87	12.62	13.25
Time fixed effects	yes	yes	yes
Individual fixed effects	yes	yes	yes
Socio-Economic control variables	no	no	no
R-squared	0.52	0.56	0.45
Observations	98,951	52,420	60,012
Number of individuals	14,420	8,938	8,566

Panel B. Dependent variable: Employment

	Overall	Formal sector	Informal sector
	(1)	(2)	(3)
Migrant share predicted from the first stage	-0.22*	-0.28**	-0.02
	[0.113]	[0.119]	[0.112]
Precipitation-levels lagged at destination	-0.02**	-0.01	-0.01*
	[0.007]	[0.006]	[0.007]
F statistic of first stage relationship	14.99	14.98	14.98
Time fixed effects	yes	yes	yes
Individual fixed effects	yes	yes	yes
Socio-Economic control variables	no	no	no
R-squared	0.02	0.01	0.01
Observations	144,672	144,643	144,643
Number of individuals	20,563	20,561	20,561

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. All specifications use IV-2SLS with lagged precipitation levels as instrument for migrant share. Log income per hour is measured in Indonesian Rupiah and employment is the share of individuals employed. The number of observations is lower for income because this is only measured if a person is employed. Socio-economic control variables include dummies for gender, education level and age group.

Table A5: Labor Market Response to Immigration in Destination-level Dataset

	Destination-level clustering		Spatial clusters across origin areas	
Dependent variable:	**Log income per hour**	**Employment**	**Log income per hour**	**Employment**
	(2)	(2)	(2)	(2)
Migrant share predicted from the first stage	-1.69**	-0.66***	-1.69***	-0.66***
	[0.728]	[0.254]	[0.601]	[0.205]
Precipitation at destination lagged	0.02	-0.01*	0.02	-0.01***
	[0.016]	[0.006]	[0.014]	[0.0037]
Time fixed effects	yes	yes	yes	yes
Destination fixed effects	yes	yes	yes	yes
Socio-Economic control variables	yes	yes	yes	yes
Observations	2,226	2,242	2,226	2,242
R-squared	0.879	0.047	0.879	0.047
Number of destinations	191	191	191	191

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. This table uses data that is collapsed to the destination-year level. The standard errors in columns (1) and (2) are clustered at the destination-level and columns (3) and (4) use Conley (1999) spatial clusters. Unless noted otherwise, precipitation variables refer to origin area precipitation z-scores and are summed over the entire catchment area of each destination according to equation (1). Socio-economic control variables include dummies for gender, education level and age group.

Table A6: Labor Market Effects using Additional Years of Lagged Precipitation

Dependent variable:	**Migrant Share of the Population**		**Log income per hour**	**Employment**
	(1)	(2)	(3)	(4)
Migrant share predicted from the first stage			-1.33**	-0.19**
			[0.551]	[0.094]
Precipitation	-0.040***	-0.042***		
	[0.011]	[0.010]		
Precipitation lagged 1 year	-0.054***	-0.054***		
	[0.011]	[0.011]		
Precipitation lagged 2 years	-0.037***	-0.034***		
	[0.008]	[0.008]		
Precipitation lagged 3 years	-0.013**	-0.016**		
	[0.006]	[0.006]		
Precipitation lagged 4 years	-0.033***	-0.032***		
	[0.009]	[0.009]		
Precipitation lagged 5 years		-0.007		
		[0.008]		
Precipitation at destination	0.031***	0.034***	0.02**	0.01**
	[0.009]	[0.010]	[0.012]	[0.003]
Precipitation at destination lagged 1 year	0.039***	0.038***	0.03**	-0.01***
	[0.010]	[0.009]	[0.013]	[0.003]
Precipitation at destination lagged 2 years	0.032***	0.032***	0.02	0
	[0.007]	[0.007]	[0.012]	[0.003]
Precipitation at destination lagged 3 years	0.014***	0.017***	0.03**	0.01*
	[0.005]	[0.005]	[0.011]	[0.003]
Precipitation at destination lagged 4 years	0.034***	0.033***	0.04***	0
	[0.008]	[0.008]	[0.013]	[0.003]
Precipitation at destination lagged 5 years		0.013**		
		[0.006]		
F statistic of joint significance	6.04	5.17	5.69	6.02
Time fixed effects	yes	yes	yes	yes
Destination fixed effects	yes	yes	yes	yes
Socio-econ control variables	yes	yes	yes	yes
R-squared	0.34	0.34	0.43	0.19
Observations	148,565	148,565	100,615	148,300
Number of destinations	205	205	205	205

All regressions are clustered at the destination level, *** p<0.01, ** p<0.05, * p<0.1. Unless noted otherwise, precipitation variables refer to origin area precipitation z-scores and are summed over the entire catchment area of each destination according to equation (1). Columns (3) and (4) use IV-2SLS and the instruments for migrant share are precipitation z-scores lagged 1 to 4 years. Log income per hour is measured in Indonesian Rupiah and employment is the share of individuals employed. The number of observations is lower for income because this is only measured if a person is employed. Socio-economic control variables include dummies for gender, education level and age group.

Chapter 3
Individual Ability and Selection into Migration in Kenya

with Joan Hamory Hicks and Edward Miguel[1]

Abstract

This study exploits a new longitudinal dataset to examine selective migration among 1,500 Kenyan youth originally living in rural areas. More than one-third of individuals report moving to an urban area during the study period. Understanding how this migration differs for people with different ability levels is important for correctly estimating urban-rural wage gaps, and for characterizing the process of "structural transformation" out of agriculture. We examine whether migration rates are related to individual "ability", broadly defined to include cognitive aptitude as well as health, and then use these estimates to determine how much of the urban-rural wage gap in Kenya is due to selection versus actual productivity differences. Whereas previous empirical work has focused on schooling attainment as a proxy for cognitive ability, we employ an arguably preferable measure, a pre-migration primary school academic test score. Pre-migration randomized assignment to a deworming treatment program provides variation in health status. We find a positive relationship between both measures of human capital (cognitive ability and deworming) and subsequent migration, though only the former is robust at standard statistical significance levels. Specifically, an increase of two standard deviations in academic test score increases the likelihood of rural-urban migration by 17%. Results are robust to conditioning on household demographic and socioeconomic measures that might capture some aspect of credit constraints or household bargaining. In an interesting contrast with the existing literature, schooling attainment is not significantly associated with urban migration once cognitive ability is accounted for. In contrast, academic test score performance is not correlated with international migration to neighboring Uganda. Accounting for migration selection due to both cognitive ability and schooling attainment does not explain more than a small fraction of the sizeable urban-rural wage gap in Kenya, suggesting that productivity differences across sectors remain large.

[1]Hamory Hicks: Center of Evaluation for Global Action, University of California at Berkeley; Miguel: Department of Economics, University of California at Berkeley. We thank Sarah Baird, David Evans, Matthew Jukes, and Michael Kremer, our collaborators on the broader KLPS project. Francisco Rodriguez, Duncan Thomas, Chris Woodruff and seminar audiences at UCLA, U.C. Berkeley, and the 2008 ASSA Meetings provided useful comments. We are grateful for financial support from the UNDP. All errors are our own.

1 Introduction

Migration is a central issue in the study of labor markets in less developed countries. While the issue of selection into migration has been widely studied in the context of Mexico-U.S. migration (Chiquiar and Hanson 2005), there is little rigorous evidence on patterns of selective rural-urban migration in less developed countries, in large part due to the scarcity of panel datasets that track individuals over time as they make migration decisions (Rosenzweig 1988). Understanding the nature of selection into urban migration as a function of individual ability can help shed light on urban-rural wage gaps, in particular, how much of the gap is due to real productivity differences across sectors versus unobserved differences in average worker ability. Characterizing rural-urban migration is also fundamental to understanding the "structural transformation" out of agriculture that is central to the process of economic development.

We explore selection into rural-urban migration and estimation of the urban-rural wage gap using a new panel data set of Kenyan youth. The Kenyan Life Panel Survey (KLPS) is a unique database, tracking over time 7,500 children who attended primary school in Busia, a rural district of western Kenya, in 1998. In Round 1 of this survey, enumerated during 2003-2005 and referred to hereafter as KLPS-1, longitudinal information was collected for more than 5,200 of these individuals on a wide range of outcomes, including all past residential locations. Round 2 of the KLPS (abbreviated hereafter as KLPS-2), a follow-up survey administered to these same individuals, is currently in the field. Prior to the launch of KLPS-2 enumeration, individuals to be interviewed were randomly divided into two groups (waves), the first to be tracked during 2007/2008, and the second to be tracked during 2008/2009. At the close of Wave 1 in November 2008 nearly 2,500 individuals had been surveyed. This study employs information from these survey respondents, a fully representative subsample of the KLPS population. A main strength of our analysis is the use of this exceptional data source.

The individuals in our analysis were surveyed in 1998, 2003/2005 and 2007/2008, and the latter two surveys collected retrospective migration histories over the intervening periods. As a result, we are able to both measure migration intensity as a series of events (employing the panel aspect of our data), as well as a transition (between survey enumeration rounds). Following Bell and Muhidin (2008), we construct transition measures as descriptive tables early in our paper, measuring migration as a change in "usual home" from the residence in Busia District during the 1998 baseline survey to residence at the time of KLPS-2 survey enumeration in 2007/2008. In the main econometric analysis of selection into urban migration, we then employ the retrospective panel data on all residential moves to capture the full extent of urban migration among rural Kenyan youth.

We focus our analysis on a restricted sample of KLPS-2 respondents with a rich set of pre-migration data on academic test scores, child and household characteristics. Individuals in this age group, primarily 18-26 years old at the time of KLPS-2 tracking, are extremely mobile. During 1998-2008, more than two-thirds of adolescents report migrating from their 1998 residence for a period of at least four months, and 41% report having lived outside of western Kenya and the neighboring parts of Uganda. The vast majority of relocation outside of these local areas is to urban centers elsewhere in Kenya. According to self-reports, schooling, employment search, and lengthy family "visits" are the three most popular reasons given for these moves.

Given this high level of mobility, sample attrition in the KLPS-2 is a natural concern. One of the unique aspects of this survey project is its commitment to locate individuals regardless of where they might have moved: survey enumerators traveled all over Kenya and neighboring Uganda in multiple rounds of long-distance tracking. As a result, 82% of target respondents were interviewed, a remarkably high tracking rate for young adults in a less developed country context. We provide a detailed analysis of tracking patterns to alleviate attrition bias concerns, and fortunately find little evidence that key explanatory variables are systematically related to attrition.

Our main empirical emphasis is two-fold. First, we examine the relationship between individual ability and subsequent migration. Such a relationship can be thought of in the context of a Roy (1951) selection model, as formulated in Borjas (1987). Previous empirical work has used schooling attainment as a proxy for ability (see appendix table 1 for a summary of main results). Resulting evidence is mixed, with most studies finding a positive association between attainment and later migration (Chiquiar and Hanson 2005, McKenzie et al 2006, Grogger and Hanson 2007), but some finding no relationship or even a negative relationship (Ibarraran and Lubotsky 2007). Hunt (2004) finds that long-distance migrants within Germany tend to be high-skilled. The evidence on the relationship between ability and migration in Africa and other low-income regions generally suggests that urban migrants are positively selected. Hoddinott (1994) examines one rural sub-location in western Kenya, and finds a positive relationship between years of schooling and urban migration. Lanzona (1998) similarly finds a positive relationship between years of schooling and migration out of rural areas in Philippines. Zhao (1999) examines migration among inhabitants of China's rural Sichuan province in 1994-5, but finds a small and only weakly positive relationship between years of schooling and migration.

Most of the empirical work on selective migration focuses on a single measure of ability, schooling attainment. We explore a broader definition, including cognitive ability as well as health status. We employ a pre-migration primary school academic test score as a proxy for cognitive aptitude, which to our knowledge is the first use of measure of this kind in a migration selection study. We also exploit pre-migration randomized assignment to a primary school deworming treatment program as a source of exogenous variation in health status, another component of human capital, and thus can more credibly identify the impact of improved health on later migration decisions.

We find only one of these ability measures to be significantly and robustly related to subsequent rural-urban migration, cognitive test scores. This suggests that cognitive aptitude is valued in the urban labor market and physical robustness perhaps less so on average. Specifically, we find that an increase of two standard deviations in 1998 academic test score increases the likelihood of subsequent migration to a city by 17%. Results are robust to several different specifications, including conditioning on measures of parent education and household asset ownership. We conclude young adults with higher cognitive ability are more likely to migrate to urban areas in Kenya. In an interesting contrast with the existing literature, schooling attainment is not associated with urban migration once cognitive ability is accounted for.

Given the high level of migration into Uganda among individuals in our sample, we extend this analysis further to explore selection into international migration. We find no relationship between our multiple measures of individual ability and subsequent international migration, likely because most adolescents moving from Busia, Kenya settle just across the border in similarly rural areas of Uganda, where cognitive and other skills are apparently not as highly valued as they are in urban labor markets.

In the second part of the analysis, we use these improved ability measures to provide more credible estimates of the urban-rural wage gap in Kenya. Specifically, we estimate how much of the massive observed Kenyan urban wage premium - urban wages in our sample are nearly twice as large as rural wages - falls when cognitive and other ability terms are included as controls in the analysis. Cognitive ability and schooling attainment are both meaningful predictors of higher wages, particularly for men. However, accounting for both individual cognitive ability and schooling attainment can explain only a small fraction of the urban-rural wage gap in our sample of Kenyan youth. This suggests that the large urban-rural wage gap in Kenya is driven by large productivity differences, or perhaps by some measures of individual ability not well captured in the variables we employ in our analysis (e.g., personality traits).

The paper proceeds as follows: section 2 describes the data, section 3 lays out a Roy selection framework, section 4 provides the main empirical evidence on selective migration, section 5 estimates the selection-corrected urban-rural wage gap in Kenya, and the final section concludes.

2 Data

In 1998, the Primary School Deworming Project (PSDP), an intestinal helminth treatment program, was launched in Busia, a rural district in western Kenya. Under this program, a local non-governmental organization (NGO) provided deworming treatment to over 30,000 primary school children aged 6-18. In order to evaluate the effects of this health intervention, baseline data was collected on individual school participation, academic performance, health and household characteristics[1]. Five years later a follow-up survey known as the Kenyan Life Panel Survey Round 1 (KLPS-1) was launched. Between 2003 and 2005, this survey tracked a representative sample of 7,500 of these adolescents who were confirmed enrolled in primary school grades 2-7 in Busia District in 1998[2]. Survey data on a wide range of outcomes was successfully collected for over 5,200 of these young adults, including panel information on all residences inhabited for a period of at least four months between 1998 and 2005. In mid-2007, a second round of the Kenyan Life Panel Survey (KLPS-2) went to the field. All sample individuals were randomly divided into two groups, to be tracked in two separate waves of data collection, both of which are fully representative of the main sample. Wave 1 of the KLPS-2 was completed in November 2008, and contains survey information for nearly

[1] Miguel and Kremer (2004) provide more background information on the PSDP.

[2] Note that this population is still fairly representative of the adolescent population in western Kenya: according to a Kenya Demographic and Health Survey, 85 percent of children in Western Province aged 6-15 were enrolled in school in 1998. However it should be noted that eighteen schools in Busia District were excluded from the sample, mainly because they were either economically and geographically unrepresentative of schools in the district or they had already received health and worm treatments under prior programs prior to the PSDP.

2,500 individuals that form the core of the analysis in this paper.[3]

In the current analysis, we employ both the baseline PSDP and the follow-up Wave 1 KLPS-2 data. We focus on a restricted sample of 1,518 individuals with detailed baseline academic test score, school participation and survey data in addition to the KLPS panel residential location information. Baseline academic test score and survey data exist for individuals who were present in school on the pre-announced day the test or survey was administered, and includes only students in grades 3 through 7 in 1998.

A key strength of the KLPS is its respondent tracking methodology. In addition to interviewing individuals still living in Busia District, survey enumerators scoured Kenya and Uganda to interview those who had moved out of local areas. Information was collected on each location inhabited since 1998 for a period of four months or more, as well as reasons for the move and any known contacts in the new location. This endeavor results in a dataset well-suited to the study of migration. Furthermore, the KLPS-2 collects detailed information on the employment and wage history of respondents, providing a rare opportunity to explore labor market outcomes among a group of highly mobile African youth.

In addition to the panel information on residential location, employment and wages, we focus on two unique variables contained in the baseline PSDP data: a pre-migration academic test score and an exogenously assigned proxy for pre-migration individual health. The baseline academic test score data comes from an exam administered to primary school students in grades 3-8 as part of the initial PSDP evaluation. The test was based on standard Kenya Ministry of Education exams, and covered three subjects - English, Math, and Science/Agriculture. Each grade level was administered a separate exam.[4] Students present in school on the day the test was administered are included in the sample. In addition, a small sample of students who had dropped out of school during 1998 were tracked to their homes and also asked to complete the exam, and we use this latter group for robustness checks in our analysis.

Our measure of pre-migration health is based on the randomized deworming treatment provided to primary school children in Busia District under the PSDP. A parasitological survey conducted by the Kenya Ministry of Health, Division of Vector Borne Diseases in early 1998 suggested that this district is characterized by an extremely high intestinal worm infection rate, on the order of 92% among sampled children in grades 3 through 8 (Miguel and Kremer 2004). Intestinal helminth infections, especially more severe cases, lead to a broad range of negative health outcomes, including abdominal pain, anemia, malnutrition, stunting, wasting, and lethargy. Since intestinal worms have life spans of just one to three years and do not replicate in the human host, periodic deworming treatment can greatly reduce infection.

Under the PSDP, a local NGO provided deworming treatment to individuals in seventy-five schools in Busia District. Due to administrative and financial constraints, the program was phased in over a four-year period. Schools were randomly divided into three groups, with Group 1 schools receiving treatment starting in 1998, Group 2 schools receiving treatment

[3] Wave 2 of KLPS-2 data collection is currently underway, and will be completed in late 2009 and included in future analyses.

[4] We implicitly assume that normalized test score at different ages captures ability to the same extent. This is plausible given our data - each year only 2-8% of students stop attending school between the grades of 3 and 7, suggesting only a second-order ability bias in higher grade levels.

starting in 1999, and Group 3 schools receiving treatment starting in 2001. Thus, Group 3 children received three fewer years of treatment than Group 1, and Group 3 children initially in grades 6 or 7 received no treatment at all.[5]

Below, we examine the relationship between the randomized deworming treatment and subsequent migration. Evidence on the link between the intervention and individual health status has been established elsewhere. Miguel and Kremer (2004) evaluate the short-run impacts of the PSDP, and find significant self-reported health and height-for-age gains after just one year of treatment. Such improvements could be associated with greater strength and labor productivity. The authors also found a drop in school absenteeism by one quarter in treatment schools, although no early academic or cognitive test impacts were found; they suggest this lack of an academic performance effect could be due in part to increased classroom congestion.

Miguel, Baird and Kremer (2007) examine the longer-run impacts of the program, using the KLPS-1 follow-up survey. The authors find long-term height and weight gains for those in lower grades in 1998, females, and for those that live in particularly high infection areas. Recognizing the difficulty in disentangling particular health impacts from each other, a mean effects approach is also used to determine the overall impact of the deworming intervention, and the authors report a positive impact of the treatment on height, weight and general health. Together, these studies suggest that deworming treatment has significant positive impacts on individual health. Such effects could continue to work through later life health, strength, and cognitive ability. We will not attempt to disentangle these effects here, but instead we focus on the randomized deworming intervention as a proxy for pre-migration individual health status.[6]

3 A Model of Selective Migration

The Roy (1951) selection model provides a useful framework for considering rural-urban migration in less developed countries, as further developed in Borjas' (1987) work. Consider an economy with two sectors, one urban and one rural. Wages in both sectors, denoted w_U and w_R , depend on individual ability, h_i. Further, there is some individual cost to migration, c_i. The Roy model suggests that individuals move to exploit wage differences across different sectors or regions. The migration decision can be characterized as:

$$Migrate\ if \quad w_U(h_i) - w_R(h_i) - c_i \geq 0 \tag{1}$$

It is natural to consider positive returns to ability in both sectors, $w_U^{'}(h) > 0,\ w_R^{'}(h) > 0$. There are many ways to think about individual ability. Traditionally, this trait has been modeled in terms of school attainment. However, ability can be thought of as a multidimensional variable, also including cognitive aptitude and health.

[5] Although only children who were in school on the day of the drug administration received treatment, compliance rates were high, on the order of 70% (Miguel and Kremer 2004).

[6] We recognize that the measure we use would be more easily interpretable if it were linked more concretely to a particular health outcome. However, as shown by these previous studies on the wide-ranging effects of the deworming treatment, choosing a single health outcome such as height or weight is restrictive.

Migration costs can be modeled more explicitly as a function of observed (X_i) and unobserved (e_i) individual and household characteristics. For instance, if credit constraints matter, then costs could be related to household income or wealth. This leads to a natural specification for a cost function:

$$c_i = -X_i'b - e_i \tag{2}$$

Allowing the urban-rural wage gap to be defined as

$$g(h_i) = w_U(h_i) - w_R(h_i) \tag{3}$$

then it follows that the migration decision can be rewritten in a standard discrete choice framework:

$$Migrate_i = 1\{g(h_u) + X_i'b + e_i \geq 0\} \tag{4}$$

Such a formulation leads to a probit specification in which individuals choose to migrate as long as the return from doing so is greater than the cost. Here, higher ability people are more likely to migrate if there are greater returns to ability in the urban sector, $w_U'(h_i) > w_R'(h_i)$, conditional on any migration costs. This is quite plausible, for instance if cognitive ability matters more in factory or office work than it does on the farm.

4 Empirical results

4.1 Attrition

Searching for individuals in rural Kenya is an onerous task, and migration of target respondents is particularly problematic in the absence of information such as forwarding addresses or phone numbers. This difficulty is especially salient for the KLPS, which follows young adults in their teens and early twenties. This age group is likely to be extremely mobile due to marriage, schooling, and labor market opportunities. Thus, it is essential to carefully examine survey attrition. If our key explanatory variables are related to attrition, then any resulting estimation will likely be biased.

Table 1, Panel A provides a summary of tracking outcomes for the individuals we study. Nearly 86% of adolescents were located by the field team, such that 82% were surveyed and 4% refused participation, were found but unable to survey, or were found to be deceased.[7] Tables 1 and table 2 break out these statistics by PSDP deworming treatment group, gender and 1998 age group. These figures suggest that tracking rates are fairly similar across treatment groups, though they are somewhat higher for males than females, and decrease monotonically with age.

[7] The 7,500 individuals sampled for KLPS-2 participation were divided in half, to be tracked in two separate waves. KLPS-2 Wave 1 tracking launched in Fall 2007 and ended in November 2008. During the first several months of Wave 1, all sampled individuals were tracked. In August 2008, a random subsample containing approximately one-quarter of the remaining unfound focus respondents was drawn. Those sampled were tracked "intensively" for the remaining months, while those not sampled were no longer tracked. We reweight those chosen for the "intensive" sample by their added importance. As a result, all figures reported here are "effective" rates - calculated as a fraction of those found, or not found but searched for during intensive tracking, with weights adjusted properly. For a detailed explanation of this methodology, see Orr et. al (2003).

We have detailed information on where all surveyed respondents were living at the time of KLPS-2 tracking. Table 1, Panel B and Table 2, Panel B summarize this information. These statistics suggest a great deal of migration in the cross-section: the crude migration intensity capturing moves outside of Busia District from 1998 until the KLPS-2 survey is 28%. Since individuals we did not find, and did not obtain residential information for, are even more likely to have moved away, these figures almost certainly understate true migration rates.[8]

More than 7% of individuals had moved to neighboring districts, including just across the border into the districts of Busia and Bugiri, Uganda. Over 20% of those with location information were living further afield, with nearly 80% of these individuals inhabiting the five major urban areas in Kenya - Nairobi, Mombasa, Kisumu, Nakuru and Eldoret.[9] Five percent of individuals had moved outside of Kenya, nearly all into the neighboring country of Uganda.

Migration rates are fairly similar across deworming treatment groups, with a slightly higher proportion of Group 3 individuals located outside of Busia and nearby districts. Females appear to have somewhat higher migration rates than males, primarily to regions neighboring Busia. This can likely be explained by high female mobility due to marriage. We also see strong evidence of migration rates increasing with age, particularly with regard to migration outside of Busia and its environs, as well as outside of Kenya as a whole.

Table 3 provides a more formal analysis of survey attrition, with focus on two key measures of individual ability, the 1998 academic test score and years assigned deworming treatment during 1998-2003, in probit specifications. The first column contains the deworming measure by itself, along with a set of controls for gender and 1998 grade, as well as baseline individual and household characteristics (whose descriptive statistics are presented in Table 4. Column (2) adds individual test score to this base specification, and column (3) further includes a control for average baseline school participation. Column (4) includes interactions of both ability measures with each other, gender and age, and columns (5) and (6) repeat earlier specifications using a linear probability model including school fixed effects. We find no evidence that years assigned deworming is systematically related to whether or not an individual was surveyed, and only weak evidence that higher pupil test scores contribute to survey attrition. This latter result is consistent with our findings below, namely that individuals with higher test scores are also more likely to migrate, and thus are generally more difficult to find. Together, this indicates that biases related to differential sample attrition in our main analysis are unlikely to be severe, but indeed likely work against our finding a selection effect: we may actually be slightly understating the relationship between cognitive ability and migration if more high ability migrants are lost from the analysis, as seems plausible given the results in Table 3.

[8] This figure is roughly comparable to Bell and Muhidin's (2008) estimate of lifetime migration intensity, 20% using IPUMS data, though we study migration from 1998 origin rather than birthplace. Our rate is higher, likely in part due to the younger age and rural origin of our focus population.

[9] We define urban areas as those with populations of greater than 150,000. Our measure of location is imperfect in that we observe districts of residence rather than cities. However, the 1999 Kenyan Census indicates that 100% of Nairobi and Mombasa districts - our respondents' main destinations - are urban, with lesser fractions for Kisumu, Nakuru and Uasin Gishu.

4.2 Migration in the KLPS-2

Over 28% of young adults were no longer living in Busia District at the time of KLPS-2 enumeration. This cross-sectional figure understates total migration among this age group, however. Panel residential location information for the period 1998-2008 among surveyed individuals suggests that 55% of adolescents migrated outside of Busia District at some point for a period of at least four months. This is perhaps not surprising: most individuals in the study group are in their early twenties at the time of KLPS-2 tracking, a period in their lives of tremendous flux as they embark on marriage, job searches or higher education.

Figure 1 displays locations of residence for individuals in our data during 1998-2008.[10] Nearly all adolescents report living in Busia district at some point or in the neighboring areas of Kenya's Western Province and the bordering districts of Uganda. The most popular residential destination by far outside of these local areas is the capital city of Nairobi. Comparatively large fractions of individuals also lived in Rift Valley Province (which houses the major urban areas of Nakuru and Eldoret, and is also an important tea-growing region with large plantations providing relatively well-paid employment), Coastal Province (home to Mombasa), and Nyanza Province (home to Kisumu). In this analysis, we characterize urban migration as residence in cities in Kenya with populations of over 150,000, as well as foreign cities (e.g, Kampala). More than one-third of individuals report living in such locations at some point during the study period. Finally, migration outside of Kenya is substantial: nearly 13% of individuals lived in Uganda at some point. More than 80% of this international migration, however, entailed a move just across the heavily trafficked and porous border between the two countries into neighboring rural districts. Migration to the Ugandan cities of Kampala or Jinja remains comparatively rare.[11]

Table 5 provides a simple comparison between individuals who have migrated to a city and those who have not, over a range of individual and household characteristics. Females and older individuals are much more likely to have lived in an urban area. Children who received more years of deworming treatment are actually less likely to live in urban areas, a result which may in part reflect that these individuals tend to be younger (and hence were able to participate in the primary school treatment program longer), and that younger individuals are less likely to have migrated. Individuals with higher baseline body weight are more likely to have migrated, a finding that again may reflect the positive association between urban migration and age instead of a nutrition effect per se. These patterns call for a more rigorous multivariate regression analysis, which we provide below. Interestingly, in the cross-section urban migration is associated with both higher baseline test scores and more years of education attained. This finding goes to the heart of our interest in the measurement of cognitive ability, and we disentangle these two measures in later regressions. Mother's educational attainment is higher for the sample of migrants, though father's attainment does

[10] Note that since many individuals lived in more than one location over the eleven-year period, these figures sum to greater than 100%. Further, these figures are not re-weighted to maintain initial population proportions.

[11] Indeed, the authors of this study themselves once unwittingly found out just how porous the Kenya-Uganda border can be. They crossed into Uganda while walking around what they thought was the outskirts of Busia Town in Kenya, and actually strolled for some time in Busia, Uganda before being stopped (and sent back to Kenya with a warning) by a plainclothes Ugandan policeman who noticed the two apparently suspicious-looking economists.

not seem to matter. Finally, urban migrants have more elder siblings on average, a finding perhaps related to family social networks that ease the information and financial costs of migration.

Table 6 displays this same set of simple comparisons, this time for individuals who migrated outside of Kenya (to Uganda) at some point during the survey period versus others. These results differ greatly from the rural-urban migration patterns. First, there is no significant difference in gender between international migrants and non-migrants, although in general older individuals are still more likely to have moved. There also does not appear to be any association between baseline test scores and later migration, and those with higher educational attainment are actually somewhat less likely to have moved outside of Kenya. Finally, migrants are more likely to come from households without a latrine, and thus perhaps come from homes of lower socio-economic status, and have fewer siblings. Together, Tables 5 and 6 demonstrate that rural-urban migration and international migration patterns differ sharply in the Kenyan context, consistent with the finding that an overwhelming proportion of migrants to Uganda settle in rural districts near the Kenyan border. We explore the differences between these migration patterns with further descriptive statistics and a more detailed regression analysis below.

Table 7 provides descriptive information on the frequency of moves and length of stay among these rural Kenyan youth, for both urban and international migrants. Panel A focuses on the former group. As previously noted, over one-third of adolescents report living in a city at some point during the 1998-2008 period, and rates are slightly higher for females and older individuals. Individuals who report rural-urban migration moved on average 2.38 times during 1998-2008, and the average length of stay in a city among these movers is 2.25 years.[12] Though older females are more likely to have ever lived in an urban area, it is older males who tend to stay longer. This may be due to the activities undertaken in the new location - as shown below, women who move to the city tend to work in domestic service jobs as temporary or casual laborers, while men are more likely to obtain permanent positions in an industrial sector.

Panel B of Table 7 explores these same figures for international migrants, and again patterns are quite different. Individuals who have lived outside of Kenya tend to be older and male, while it is the older females who stay abroad longer. Again, this appears to be related to the migrants' activities: a large share of female migration into Uganda is due to marriage, which is typically a long-term proposition.

Table 8 breaks down the stated reasons for migration. The three most popular motivations for urban migration are visiting friends or relatives, schooling/training and employment search, although marriage is also a leading factor in female migration. The former reasons fit well with the temporal pattern of moves. As Figure 2 suggests, most urban migration occurs in December and January, at the close of the calendar school year and when one might move to begin a new course of schooling, to look for a new job, or for an extended holiday with friends or relatives.

Panel B of Table 8 suggests a similar set of broad motivations for international migration. One key difference here is that few women migrate abroad to look for work, and instead most

[12] Many of these stays were censored, i.e., were still ongoing at the time of enumeration, so this is an underestimate.

move for marriage. However, the temporal pattern of international migration remains quite similar to that of urban migration, with most moves occurring in January (not shown).

Thus far we have discussed when and why young Kenyan adults move out of their rural homes into urban areas, or to international locations (which are almost entirely rural districts of Uganda). Table 9 presents individual characteristics at the time of survey enumeration for those living in rural versus urban locations. Compared to their counterparts, young adults living in a city are slightly less likely to ever have been married or pregnant, and this effect is largely driven by younger males. While over 25% of young adults living in rural areas are still attending school, this is true for only 14% of individuals who have migrated. In contrast, urban migrants are much more likely to be in a vocational training program, both men and women alike. Inhabitants of rural areas are apt to run their own business (almost entirely in the informal sector), while those in urban areas are more likely to be employed in formal sector jobs.[13] Unemployment rates are high in both the rural and urban samples, and are similar across age and gender among those living in a city.

4.3 The Kenyan Demographic and Economic Climate

Our study focuses on young adults in Kenya. This age group, composing nearly a quarter of the Kenyan population, is extremely important in shaping both current and future economic outcomes. In order to better understand the migration decisions and labor market activities of these individuals, a brief discussion of the Kenyan demographic and economic setting is useful.

The Kenyan population has increased rapidly since independence, with urban areas experiencing the fastest growth (Republic of Kenya 2002a). Nairobi in particular has grown much faster than any other province, with population increasing by more than 60% each decade. In fact, Nairobi and the Rift Valley province have shown consistent increases in their share of the national population over this period, while shares in other provinces have stagnated or decreased (Republic of Kenya 2001).

This urban population expansion has been fueled in large part by internal migration. Tabulations from the 1999 Kenyan Census suggest that nearly 70% of individuals living in Nairobi at the time of enumeration were born elsewhere, and similarly 57% in Mombasa, 48% in Nakuru, 39% in the district containing Eldoret and 34% in Kisumu. In contrast, only 13% of inhabitants of Busia District (our baseline study district) had migrated there. Further, net migration figures show large influxes of migrants to four of the five main urban areas (Kisumu being the exception), with the numbers of migrants increasing each decade since 1979. Statistics describe a net increase in migrants aged 10-29 for females and males in these four urban centers, the age group we study here (Republic of Kenya 2002b).

The Kenyan economy has also undergone dramatic changes in the post-independence period. Average annual GDP growth was highest in the 1970s, and has slowed since. Indeed, the second half of the 1990s saw shrinking per capita income. Annual GDP growth rates more recently have been extremely volatile, ranging during 1998-2007 from 0.5% to nearly 7% depending on the year (World Bank 2007). In addition, the sectoral composition of national

[13] Employment in the KLPS-2 is defined as working for pay, volunteering, or interning, and does not include most home agricultural activities.

income has shifted considerably. National accounts data (presented in Figure 3) demonstrate a growing importance of the services sector since the late 1970s, now accounting for over half of value-added, while agriculture has waned and industry stagnated. Focusing more specifically on 1998-2007, the share of agriculture in value-added fell from 32% to 23%, and industry's share increased slightly from 18% to 19%, while value-added in services increased sharply from 50% to 58% (World Bank 2007).[14]

One recent survey finds that nearly half of all Kenyans are unable to meet daily minimum food and non-food requirements (World Bank 2008). Consumption growth is quite uneven across Kenyan provinces, and poverty is especially salient in rural areas. Indeed, the Kenya Poverty and Inequality Assessment (World Bank 2008) finds that mean household consumption grew 24% in urban areas during 1997-2006, while only growing 1.5% in rural areas over the same period. However, it is interesting to note that this same study suggests that poverty rates are lower in households with a migrant. This is perhaps because better-off or more able people migrate, or that migration opens up more opportunities for income creation. We seek to partly disentangle these possibilities in the main analysis that follows (in sections 4.4 and 5).

Despite macroeconomic volatility, the labor force has continued to grow. Census data reveals a nearly seven percentage point increase in the labor force participation rate between 1989 and 1999, with faster growth for females than males.[15] Unsurprising for a country with a high fertility rate, the majority of the labor force remains young, with the largest proportion of individuals between the ages of 20 and 29. Educational attainment among the economically active has also improved dramatically in recent years: the proportion of Kenyan workers with no formal education declined from one-third to one sixth during 1989-1999, though the majority of workers have still attained no more than a primary school education (Republic of Kenya 2002c).

A snapshot of the labor force in a 1998/99 national survey finds more than three-quarters of Kenyans economically active: 66% working and 11% unemployed. Just over half of individuals in the 15-24 age group are labor force participants-38% are employed and 14% unemployed-while many of the inactive individuals are still undoubtedly pursuing their education. More recent figures suggest youth unemployment is now over 20% (World Bank 2008). Labor force participation rates are higher for women than men in this group, but higher for men in older cohorts. Labor market participation rates among individuals aged 15-64 are substantially higher in urban areas than rural ones, as are unemployment rates, with young women having the most severe unemployment. The national data further suggests that unemployed men generally seek paid work in both rural and urban areas, while unemployed women focus their search in urban areas (Republic of Kenya 2003).

Figures from the KLPS-2 provide a similar snapshot for 2007/2008. Among the KLPS population nearly 60% of adolescents are active in the labor force, with more than one-third employed or self-employed, and approximately one-quarter unemployed. Labor force participation is higher in urban areas (76%) than in rural ones (55%), and unemployment is also higher in cities. One key divergence from the national figures is that young adults in

[14] It should be noted that these figures may not fully account for growth in the increasingly important informal sector in Kenya, as enterprises in this sector are generally not officially registered. For a discussion of national accounts source data and the poor quality of data on the informal sector, see IMF (2005).

[15] The following discussion of the labor force focuses on individuals aged 15-64 unless otherwise noted.

the KLPS-2 sample show higher participation rates among men (67%) than women (48%).

According to nationally representative data, small-scale agriculture is the dominant sector of employment in rural areas, while urban workers tend to be employed in the modern (formal) and informal sectors. The 1998/99 Integrated Labour Force Survey (ILFS) reports that 51% of urban employees work in the modern sector, while 39% work in the informal sector. Employment in both sectors has increased in recent years (Republic of Kenya 2005; Republic of Kenya, various years).

Table 10 utilizes the KLPS-2 data to outline the industrial breakdown of working adolescents in urban versus rural locations. Note that agriculture for own use, which is the primary activity for rural individuals, is not included in our definition of employment and hence is left out. Among those working for pay or family gain, or self-employed, most rural inhabitants work in retail or other unclassified industries. In contrast, urban migrants primarily work in manufacturing, domestic service, retail and other service industries. The first and last of these are dominated by male migrants, while female migrants are much more likely to work in retail and domestic service.

Employment questions in the KLPS-2 survey attempt to, but cannot always, distinguish perfectly between formal and informal sector employment. However, it is likely that most of our respondents work in informal sector jobs. Table 10 shows that urban female migrants are most often employed as "house girls" (domestic servants), the quintessential female informal sector job. Furthermore, individuals' employment status presented in Table 11 suggests that most positions are temporary or casual, for rural and urban workers alike, again implying largely informal sector employment. Finally, the types of industries in which most KLPS respondents work (restaurants, domestic service and other service industries) line up closely with employment in the informal sector (World Bank 2008).

Modern sector real average wages per employee in Kenya have generally increased over the past two decades, with notable exceptions in the early-to-mid 1990s. Between 2000 and 2005, wage growth was fastest in the private sector industries of transport and communications; finance, insurance, real estate and business services; and community, social and personal services. The fastest growing wages in the public sector were in transport and communications, as well as in trade, restaurants and hotels. Although wage growth was slow in some private sector industries over this period-especially in commercial agriculture-public sector wage growth was actually negative in mining and quarrying, and in manufacturing (Republic of Kenya, various years). [16]

The last panel of Table 11 presents figures on average monthly wages from paid employment, generated using the KLPS-2 sample. Cash salaries and in-kind payments taken together are twice as high in urban areas than rural areas.[17] Among those living in a city, remuneration is nearly twice as high for men than for women. Recall that large shares of

[16] The Kenyan government has outlined a minimum wage policy since Kenyan independence in 1963, and guidelines are adjusted on nearly an annual basis. However, this policy does not apply to formal public sector employment (in which wages are determined by service and periodic performance reviews) or to informal sector employment (due to legal weak enforcement), and thus does not constrain wages for most employees. Even replacing the cash salaries of those who report being unemployed with zero, the gap is similarly large, at Ksh 1,061 in rural versus Ksh 2,799 in urban areas.

[17] Even replacing the cash salaries of those who report being unemployed with zero, the gap is similarly large, at Ksh 1,061 in rural versus Ksh 2,799 in urban areas.

KLPS-2 urban women work in generally low-paying domestic service jobs.

This description of the Kenyan demographic and economic climate has highlighted several key differences between urban and rural regions. Migration rates are largest to urban areas, where average wages are much higher and jobs in manufacturing and service sectors are concentrated. There is also evidence that families with migrants tend to have lower poverty rates. We now proceed into our main analysis, examining which individuals migrate and whether such selection can explain the large observed urban-rural wage gap in Kenya.

4.4 Empirical Evidence on Selection into Urban Migration in Kenya

Table 12 presents the main empirical results on the migration selection analysis. Column (1) displays results using a linear probability model, including one of the two key variables of interest, years of assigned deworming treatment, as well as individual and household control variables. Although the point estimate on deworming is positive and of moderate magnitude, it is not statistically significant at traditional confidence levels. This is true across all specifications the table. It may be that health status is not valued more highly in urban sector jobs than it is on the farm. (We will reevaluate this relationship in future analysis featuring both the Wave 1 and Wave 2 KLPS-2 subsamples.)

The 1998 academic test score is positively and significantly related to subsequent urban migration (column 2), and this holds robustly across specifications in this and ensuing tables. Note that none of the other individual characteristics or proxies for household socioeconomic status are robustly related to migration, with the exception of mother's educational attainment, which is also positively correlated with urban migration. The finding that household assets and other socio-economic characteristics do not predict migration argues weakly against the hypothesis that credit constraints are a major impediment to rural-urban migration in this context. A probit model produces similar results (column 3), and suggests that a two standard deviation increase in academic test score results in 17% increase in the likelihood of rural-urban migration. Disaggregating the 1998 test score measure by subject (English, mathematics, science/ agriculture) does not reveal that a single subject drives the results (not shown).

Results are robust to the inclusion of additional regression controls. Column (4) includes a measure of individual school attendance in 1998. The size and significance of the main results are unchanged, suggesting that, above and beyond how frequently an individual attended school, cognitive ability has a positive relationship with later migration. Column (5) includes an interaction between the two main ability variables of interest, as well as their interactions with gender and age, but these interaction results are not large in magnitude nor significant.

Figure 4 displays the relationship between the individual test score and migration using a cubic polynomial fit for the full sample (a variety of polynomial controls or nonparametric methods produce visually similar relationships). The strong positive association between test score and migration at higher scores is apparent especially for those with scores greater than one standard deviation above the mean, although we cannot reject a linear relationship. Splitting the sample by gender produces similarly positive relationships (not shown).

Columns 6 and 7 of Table 12 include school fixed effects, and produce similar results, although standard errors increase somewhat, not surprisingly. The school fixed effects might better capture local socio-economic status measures or transport costs not adequately picked up in the earlier regressions, hence this is an important robustness check. Here we focus on the academic test score results; the deworming treatment was randomized at the school level, and so there is not sufficient within-school variation to estimate impacts (any variation comes from differences across initial grade level).[18] The test results support the earlier findings, with an almost identical positive relationship between pupil test score and subsequent rural-urban migration (column 6) and weak interaction effects (column 7).[19]

The results in Table 13 examine the role of schooling attainment in urban migration, and provides an interesting contrast to existing studies of selective migration. We consider the relationship between urban-rural migration and schooling attainment - the almost universal measure of individual ability in the literature - in column 1, and find it to be positive, of moderate magnitude, and highly statistically significant. A three year increase in schooling increases the likelihood of migration by more than 5 percentage points, or roughly 16%. However, the magnitude of this coefficient is cut nearly in half, and loses statistical significance at traditional confidence levels, when controls for parent education are added to the specification (column 2). Mother's education is particularly influential, as in Table 12. When the test control for individual cognitive ability is also included (column 3), we continue to find a strong positive relationship between pre-migration test score and subsequent urban migration, nearly unchanged from Table 12, while the coefficient on schooling attainment falls close to zero. These results provide evidence that cognitive ability is an arguably preferable measure in the study of selection into urban migration in the Kenyan context, and that results of existing studies might be revised if authors had had access to detailed test score data, such as that in the current study.[20]

Table 14 provides results on international migration into Uganda. These results reinforce the earlier descriptive findings of a sharp contrast with the urban migration results. Neither cognitive test scores, nor deworming, nor educational attainment significantly predict international migration in our sample, nor does mother's education (though the latter actually has a small and weak negative relationship). There do appear to be some socioeconomic correlates of international migration but these are inconsistent in sign and difficult to interpret: years of father's schooling is positively linked to migration to Uganda, but those from households with latrines (who tend to be better off households in our setting) are less likely to move.

Overall, there is no evidence that any dimension of ability is related to international migration in our sample. This stands in sharp contrast to the large literature on Mexico-U.S. international migration discussed in the introduction, but of course an important difference

[18] Years assigned deworming treatment is still included as a control in Table 12, columns 6 and 7, nonetheless.

[19] The test score information utilized in the forgoing analysis was only available for individuals present on the day the test was administered. To provide robustness checks on these results, we include additional test score information obtained from a sample of students who had dropped out of school during 1998, but were tracked to their homes and asked to complete the exam. This increases the sample size slightly to 1531 individuals. As before, there is a strong relationship between pupil test scores and subsequent urban-rural migration (not shown).

[20] Given the high rate of enrollment in Kenya during early primary school, exam scores - especially for students in lower grades in 1998 - are more likely to reflect raw academic and cognitive ability rather than household factors which could influence school enrollment and attendance (such as wealth, results not shown).

between the two settings are the relative living standards in each pair of countries: the U.S. is much wealthier than Mexico, while Kenya and Uganda are at broadly similar levels of economic development. Further the vast majority of international migrants in our sample move just across the border into in the rural districts of eastern Uganda, settings where ex ante few would expect migration to be strongly selected on individual ability.

5 Estimating the Urban Wage Premium in the Presence of Selective Migration

In this section, we use cognitive test score data as an improved measure of individual ability in order to provide more credible estimates of the urban-rural wage gap in Kenya. There is a massive urban wage premium in this setting: conditioning on all of the household and school controls in the previous tables, except for the cognitive test score and schooling attainment, average urban wages in our sample remain twice as large as rural wages (Table 15, column 1).[21] This premium is much larger for men than women in magnitude, at 2648 Kenya shillings per month for men and only 1113 shillings for women (not shown in the table), although the proportional urban wage premium is more similar given men's much higher average earnings.

As expected, both schooling attainment and higher cognitive test score performance are associated with much higher wages, although the test score effect is only marginally statistically significant. A three year increase in schooling is associated with 22% higher wages in the whole sample, while an increase of two standard deviations in the 1998 cognitive test is associated with a roughly 17% wage gain in our sample conditional on other covariates (column 2), and both effects are almost entirely driven by male workers (not shown).

The question we ask is whether the observed urban wage premium continues to hold when these observed ability measures are taken into account, given the strong link between cognitive tests and urban migration documented in Tables 12 and 13. We estimate this in column 3, including controls for the test score, schooling attainment and interactions of each with urban location, to assess whether there are differential returns to skill in urban areas.[22]

We find in column 3 that the large Kenyan urban wage premium is largely robust to including these controls, and running these regressions separately with the two ability measures yields largely similar results (not shown). Both the test score and schooling attainment measures in this table are demeaned, and thus the urban wage premium is 1933.3 Shillings per month. The overall average urban wage premium (in column 1) is 2111.1, which implies that considering observed schooling attainment reduces the urban wage premium by only 8.4%.[23]

Figures 5 and 6 show this graphically. The urban versus rural returns to cognitive test scores and schooling attainment (both conditional on other household and school character-

[21] Wages are measured here as the cash salary from primary employment among individuals who are employed. The analysis was repeated includes zero salary for individuals who are unemployed (no current job but are looking for work), and the results are substantively the same (see below).

[22] This is conceptually related to the Blinder-Oaxaca decomposition.

[23] These are all nominal wage differences. In future work, we will consider urban-rural prices differences and thus real wage differences across sectors. Nonetheless, the main conclusion that ability measures cannot explain the urban wage premium will remain largely unchanged.

istics) are presented in these two figures, respectively. The relationships are strongly upward sloping, indicating that higher skilled individuals earn higher wages, and there remains a large urban rural wage gap in both cases. Together, Table 15 and Figures 5 and 6 provide evidence that the urban-rural wage gap in our sample of Kenyan youths is largely robust to observed schooling attainment and cognitive test score differences between urban and rural residents, due to large inherent productivity differences across sectors, or perhaps due to some measures of individual ability not well captured in the variables we employ in our analysis (e.g., personality traits), rather than due to migration selection along individual ability.

6 Conclusions and Future Work

We conclude from this analysis that high ability young adults are more likely to migrate out of rural Kenya and into cities, and the magnitude of these effects is quite large. While perhaps not surprising in and of itself, given the number of recent studies that also find positive selection into migration, our use of a true panel dataset of young adults over a decade and novel measures of ability - including both pre-migration cognitive aptitude and health status - sets this work apart from previous studies. Our ability to exploit exogenous variation in health status induced by randomized assignment to deworming treatment is also a strength. Future work will extend the analysis by considering the KLPS-2 Wave 2 sample, which will roughly double the sample size.

In addition to building on the results of previous selective migration studies, the novel cognitive test score data allows us to make further progress on the classic issue of determining how much of the urban wage premium is due to actual productivity differences rather than selection on unobserved ability. We find that including controls for both schooling attainment and cognitive performance does not appreciably diminish the very large observed urban-rural wage gap observed in our sample of Kenyan youths, in which urban jobs appear to pay roughly twice as much as rural employment. At least in this population, there appear to be very large productivity differences across sectors - perhaps due to agglomeration externalities or other characteristics of the urban environment - beyond what can be explained by selective urban migration.

Our analysis focuses on a population of young adults born in rural areas, and as such not all findings will likely generalize to older workers or those born in urban areas of Kenya. In particular, a study by the World Bank (2008) notes that in general youth unemployment rates are twice those for adults and their wages are much lower. Despite these caveats regarding generalizeability, rural youths remain a key and arguably understudied population, and one which composes a large fraction of the population of many African societies.

Another important issue is whether these findings generalize beyond Kenya. If migration depends on relative returns to skill across sectors, then the extent of technological sophistication in agriculture and the types of urban sector jobs will be critical in determining relative returns to skill. Kenya has relatively unsophisticated agriculture and plentiful formal and informal sector jobs in Nairobi - East Africa's largest city - and such opportunities continue to improve in Kenyan cities given the country's recent economic growth. This is exactly the type of setting in which we would expect to see a great deal of selective urban migration for skilled young adults. It is possible, however, that different patterns would prevail in other countries where cities are smaller and skilled employment opportunities less abundant. We leave this for future research.

References

Bell, Martin, and Salut Muhidin. (2008). "Cross-National Comparison of Internal Migration". Report prepared for the United Nations Human Development Report Office for the 2009 Human Development Report.

Borjas, George J. (1987). "Self-Selection and the Earnings of Immigrants." The American Economic Review, 77(4), 531-553.

Chiquiar, Daniel, and Gordon Hanson. (2005). "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States." Journal of Political Economy, 113(2), 239-281.

Grogger, Jeffrey and Gordon H. Hanson. (2008). "Income Maximization and the Selection and Sorting of International Migrants", NBER Working Paper #13821.

Hoddinott, John. (1994). "A Model of Migration and Remittances Applied to Western Kenya." Oxford Economic Papers, 46(3), 459-476.

Hunt, Jennifer. (2004). "Are migrants more skilled than non-migrants' Repeat, return, and same-employer migrants." Canadian Journal of Economics, 37(4), 830-849.

Ibarraran, Pablo, and Darren Lubotsky. (2007). "Mexican Immigration and Self-Selection: New Evidence from the 2000 Mexican Census." In George Borjas (ed.), Mexican Immigration. Cambridge: National Bureau of Economic Research, Inc.

International Monetary Fund (2005). "Kenya: Report on the Observance of Standards and Codes (ROSC) - Data Module." IMF Country Report 05/388.

Lanzona, Leonardo A. (1998). "Migration, Self-Selection and Earnings in Philippine Rural Communities." Journal of Development Economics, 56, 27-50.

McKenzie, David, John Gibson, and Steven Stillman. (2006). "How Important is Selection-Experimental vs. Non-Experimental Measures of the Income Gains from Migration." IZA Discussion Paper #2087.

MEASURE DHS STATcompiler. (2007). http://www.measuredhs.com.

Miguel, Edward and Michael Kremer. (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." Econometrica, 72(1), 159-217.

Miguel, Edward, Sarah Baird and Michael Kremer (2007). "The Long Run Impacts of a School Health Intervention in Kenya", unpublished working paper, U.C. Berkeley.

Orr, Larry, Judith D. Feins, Robin Jacob, Erik Beecroft, Lisa Sanbonmatsu, Lawrence F.

Katz, Jeffrey B. Liebman, and Jeffrey R. Kling (2003). "Moving to Opportunity Interim Impacts Evaluation." Washington, DC: U.S. Department of Housing and Urban Development.

Republic of Kenya: Central Bureau of Statistics. (2001). "1999 Population and Housing Census, Volume I: Population Distribution by Administrative Areas and Urban Centers." Government Printer: Nairobi.

Republic of Kenya: Central Bureau of Statistics. (2002a). "1999 Population and Housing Census, Volume III: Analytical Report on Population Dynamics." Government Printer: Nairobi.

Republic of Kenya: Central Bureau of Statistics. (2002b). "1999 Population and Housing Census, Volume VI: Analytical Report on Migration and Urbanization." Government Printer: Nairobi.

Republic of Kenya: Central Bureau of Statistics. (2002c). "1999 Population and Housing Census, Volume IX: Analytical Report on Labor Force." Government Printer: Nairobi.

Republic of Kenya: Central Bureau of Statistics. (2003). "Report of the 1998/99 Labor Force Survey." Government Printer: Nairobi.

Republic of Kenya: Central Bureau of Statistics. (2005). Statistical Abstract: 2005. Government Printer: Nairobi.

Republic of Kenya: National Bureau of Statistics. (various years). Economic Survey. Nairobi.

Rosenzweig, Mark R. (1988). "Labor Markets in Low-Income Countries." In H. Chenery and T.N. Srinivasan (eds.), Handbook of Development Economics, Volume I (pp. 713-762). New York: Elsevier Science Publishers B.V.

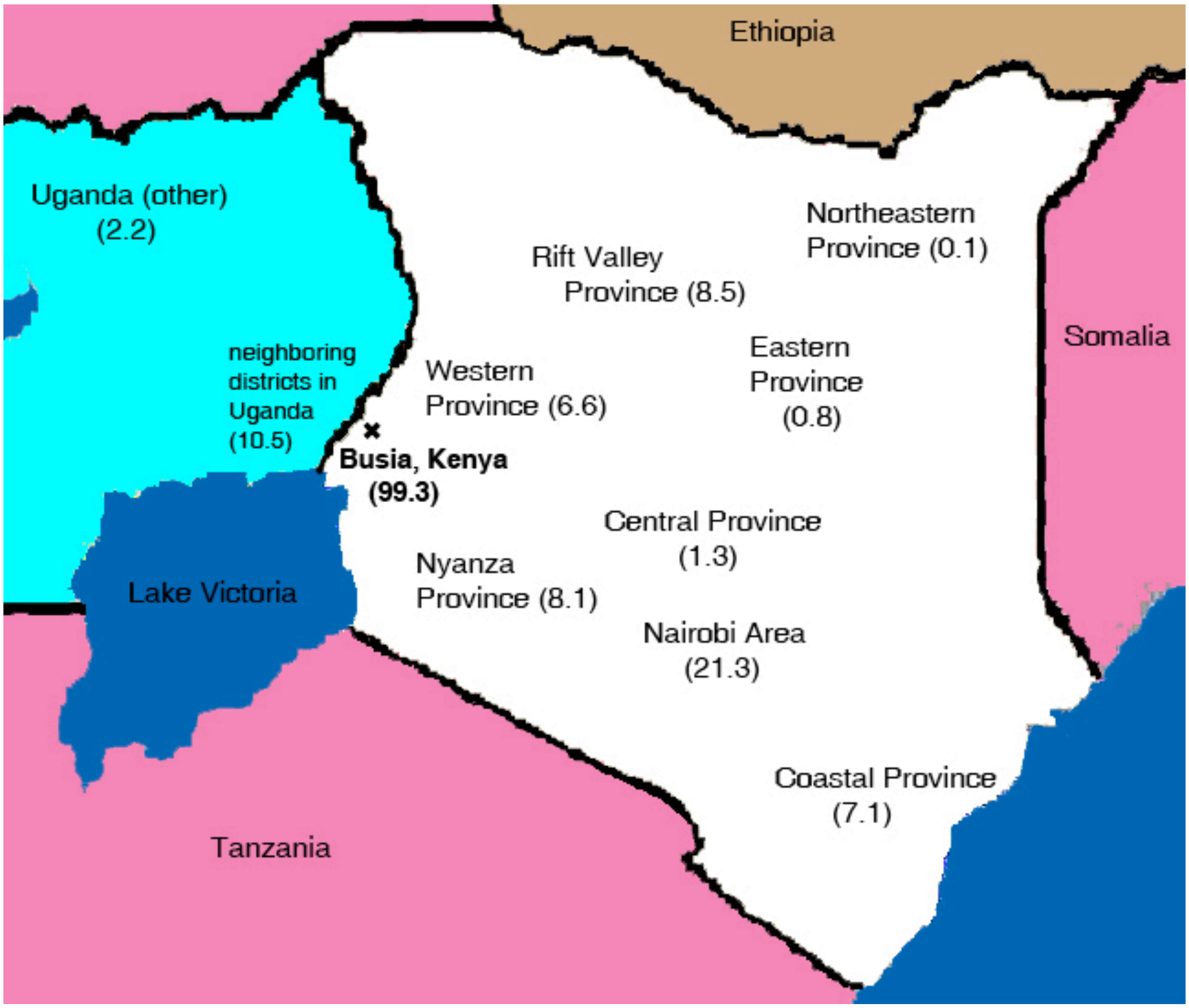
Roy, A.D. (1951). "Some Thoughts on the Distribution of Earnings." Oxford Economic Papers, 3(2), 135-146.

World Bank. (2007). World Development Indicators. Washington, DC.

World Bank. (2008). "Kenya Poverty and Inequality Assessment: Executive Summary and Synthesis Report." Draft Report No. 44190-KE.

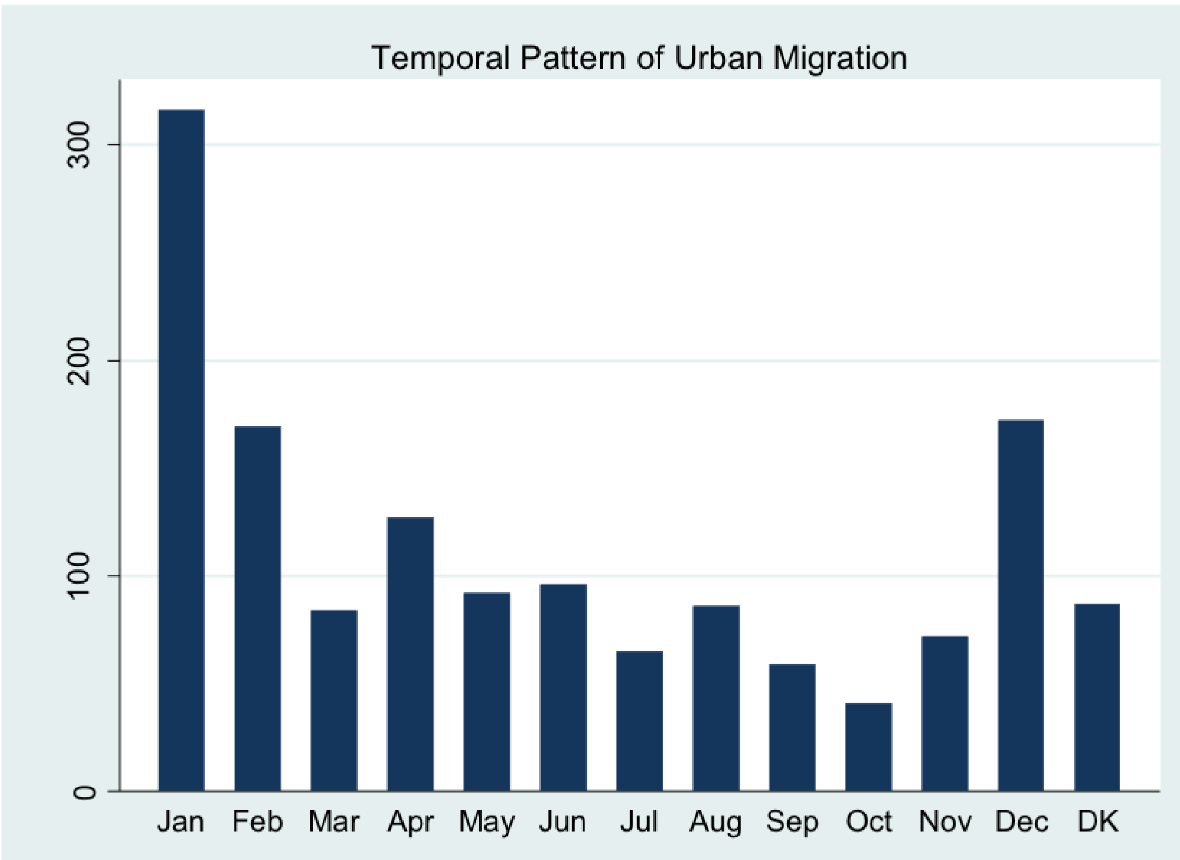
Zhao, Yaohui. (1999). "Labor Migration and Earnings Differences: The Case of Rural China." Economic Development and Cultural Change, 47(4), 767-782.

Figure 1: Locations of residence during 1998-2008



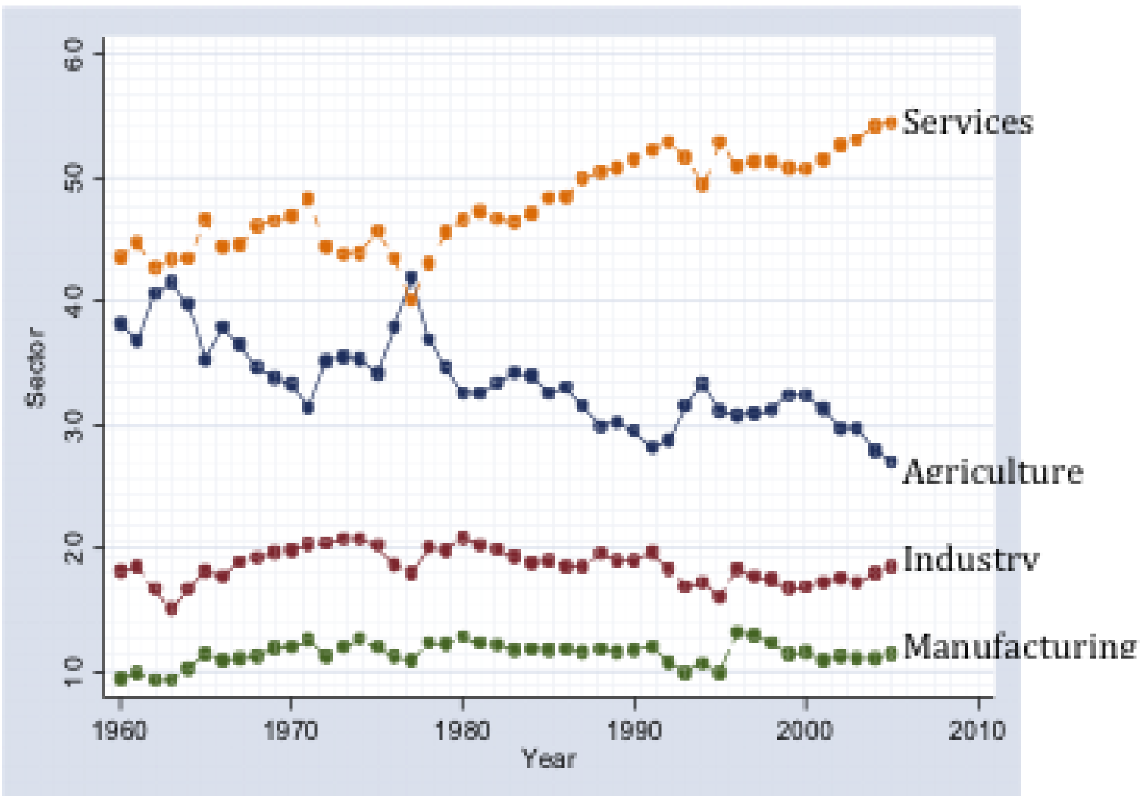
Notes: The sample here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data. One observation with an extreme 1998 ICS test score was dropped from the sample, as well as six observations missing date of survey or age information. Values signify percentage of sample that inhabited a given location at some point during 1998-2008. Values will sum to greater than 100, as individuals lived in multiple locations during the survey period. These figures are not weighted to maintain initial population proportions.

Figure 2: Temporal pattern of migration, among urban migrants



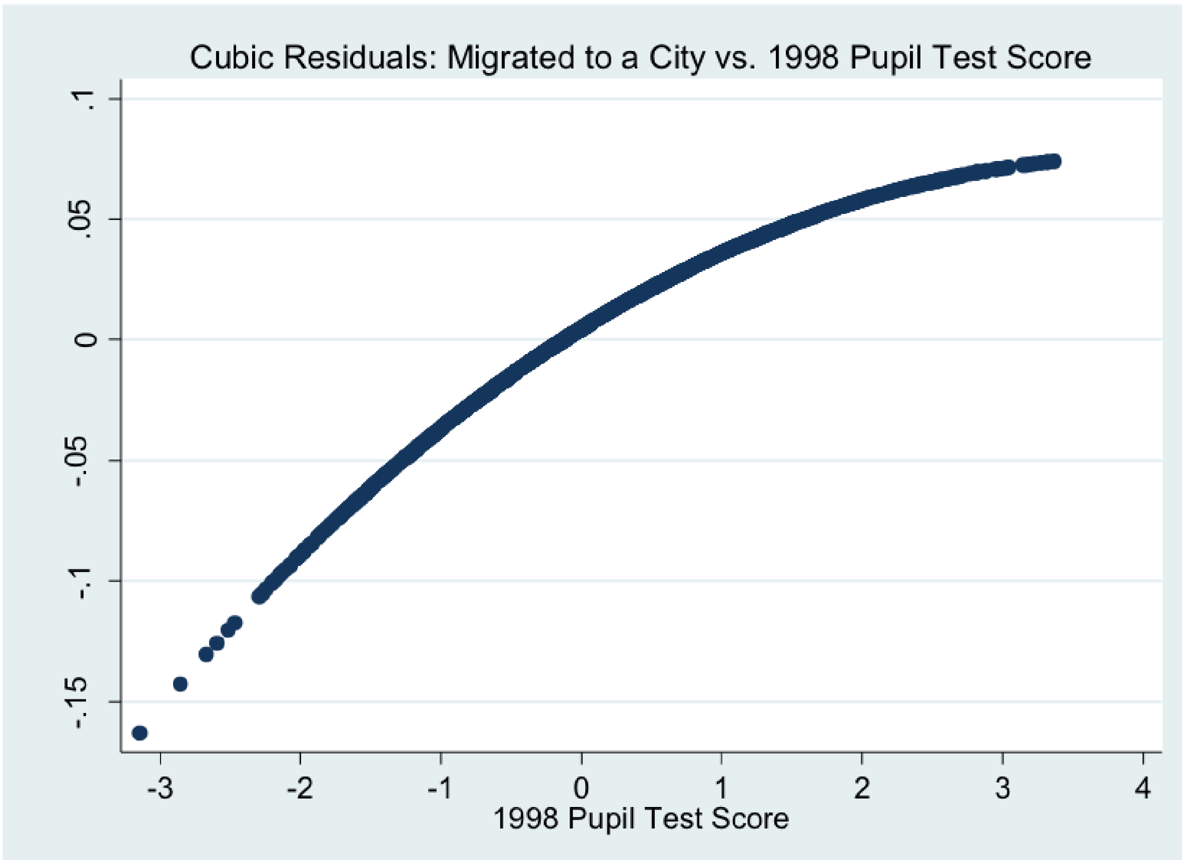
Note: The sample used here includes all individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data, who were surveyed and report migration to a city during 1998-2009. Date of migration information is missing for 131 individuals. In addition, two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations residential location or age information. Figures are not weighted to maintain initial population proportions.

Figure 3: Share of Value-Added in GDP by Sector



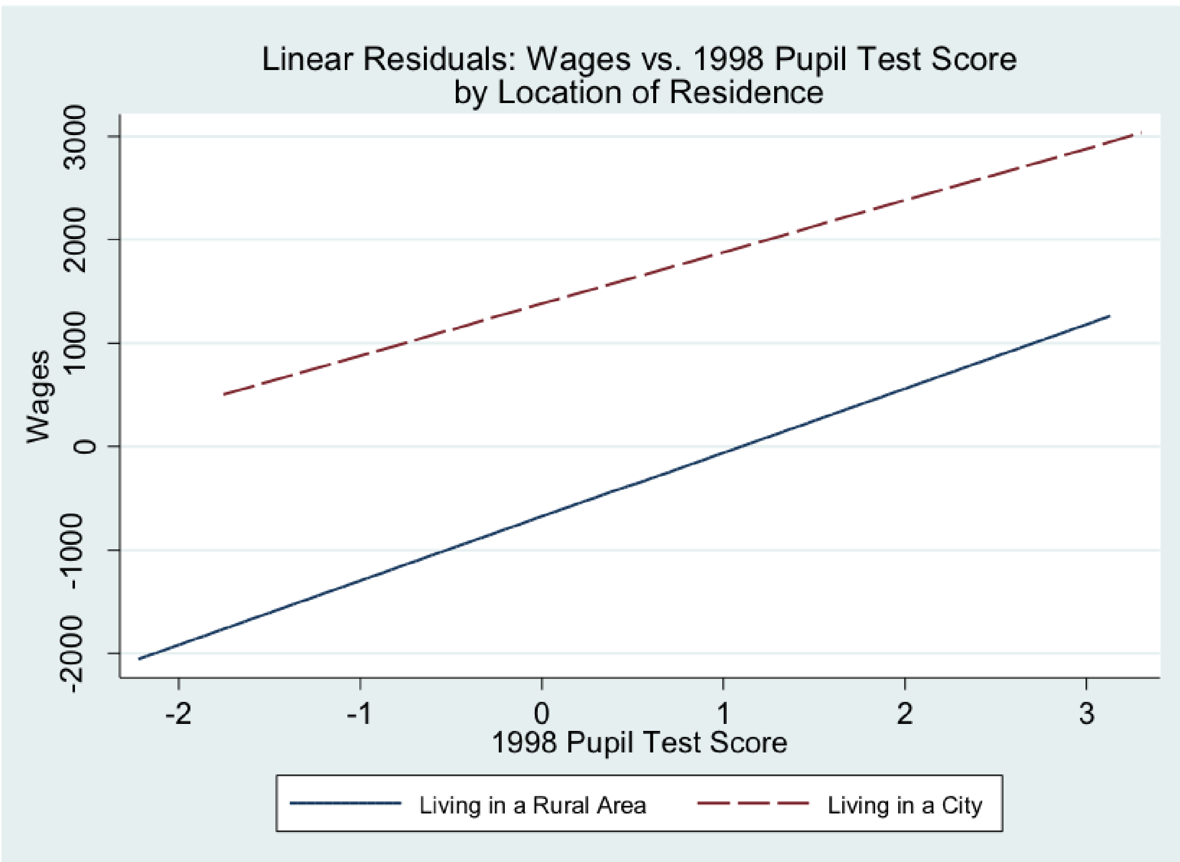
Source: World Bank (2007)

Figure 4: Cubic plot of urban migration on test score



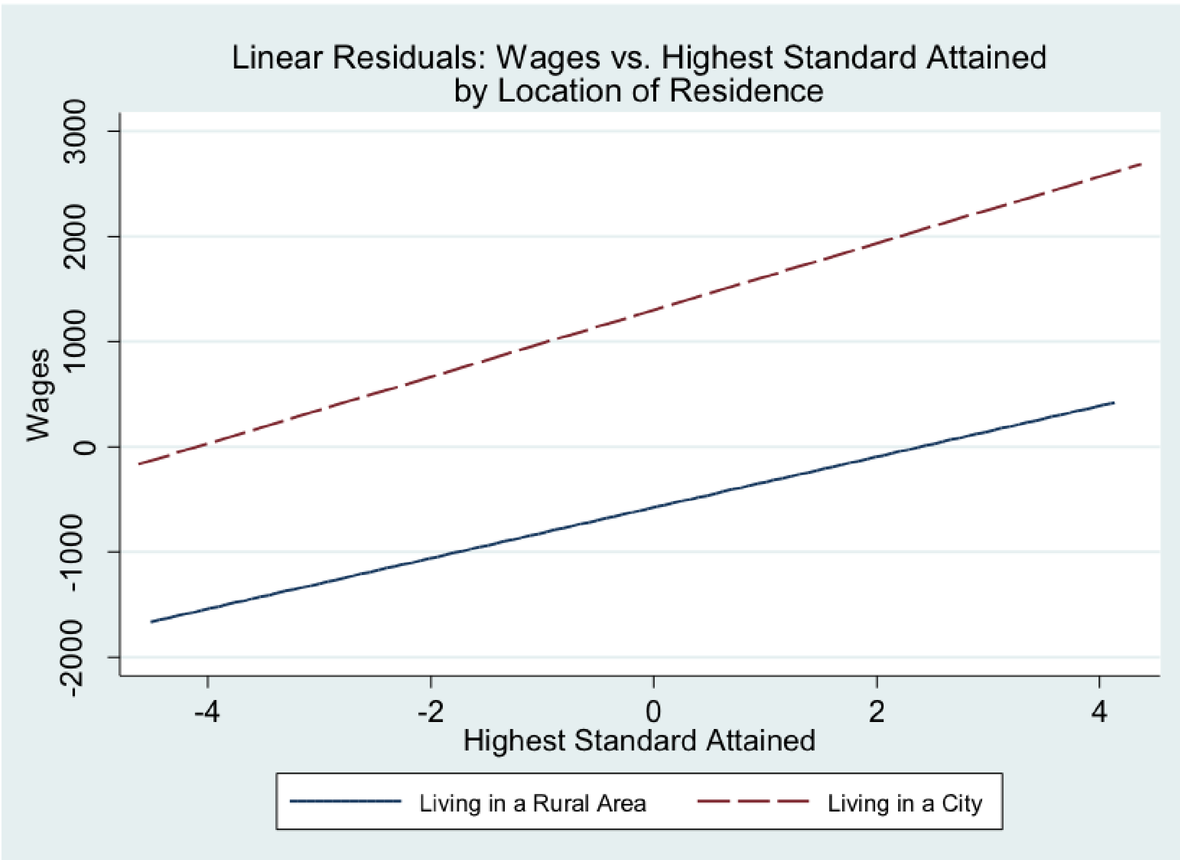
Note: The sample used here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations residential location or age information. Residuals result from regressions of migration to a city and test score using the specification reported in Table 12, column (2).

Figure 5: Linear residuals fit of wages on test score, by location of residence



Note: The sample used here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data, as well as information on wages. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location or age information. Wages are measured as cash salary in the last month. Both wages and test score are presented here as residuals from a regression of each on a set of individual and household-level controls.

Figure 6: Linear residuals fit of wages on schooling attainment, by location of residence



Note: The sample used here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, ICS test score and schooling attainment data, as well as information on wages. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location and age information. Wages are measured as cash salary in the last month. Both wages and schooling attainment are presented here as residuals from a regression of each on a set of individual and household-level controls.

Table 1: Summary statistics on sample attrition and residential location

Means	All	Treatment Group 1	2	3	Gender Female	Male
Panel A: Sample attrition, KLPS-2 I-Module						
Found (effective tracking rate)[a]	0.866	0.865	0.848	0.886	0.848	0.883
Surveyed (effective response rate)	0.833	0.832	0.823	0.846	0.820	0.845
Not surveyed, dead	0.015	0.018	0.012	0.014	0.012	0.017
Not surveyed, refused	0.015	0.014	0.010	0.022	0.016	0.015
Panel B: Residential location information[b]						
Residence in Busia District[e]	0.671	0.654	0.696	0.667	0.647	0.692
Residence in districts neighboring Busia District[c]	0.080	0.091	0.075	0.071	0.105	0.058
Residence outside of Busia and neighboring districts[d]	0.249	0.255	0.230	0.262	0.249	0.250
Residence in a city	0.199	0.212	0.184	0.198	0.200	0.198
In Nairobi	0.123	0.120	0.102	0.148	0.121	0.125
In Mombasa	0.041	0.049	0.048	0.023	0.037	0.044
In Nakuru	0.010	0.017	0.007	0.005	0.011	0.010
In Kisumu	0.017	0.015	0.019	0.017	0.021	0.013
Residence outside of Kenya (in Uganda)	0.052	0.054	0.053	0.049	0.065	0.040
Number of Observations	3363	1185	1098	1080	1666	1697

Note: The sample used here includes all individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data, who were surveyed, found deceased, refused participation, found but unable to survey, or not found but searched for during intensive tracking. All figures are weighted in order to maintain initial population proportions.

a. The proportion found is the combined rates of pupils surveyed, found deceased, refused and found but unable to survey.

b. Residential location information is available for surveyed individuals only.

c. Districts neighboring Busia include Siaya, Busia (Uganda), Bugiri (Uganda) and other districts in Kenya's Western Province.

d. The categories of "Residence outside of Busia and neighboring districts", "Residence in a city" and "Residence outside of Kenya" are not mutually exclusive.

e. In 2007, Busia District was separated into three districts - Busia, Samia and Bunyala. The present definition of Busia District contains all three.

Table 2: Summary statistics on sample attrition and residential location, by age group

Means	1998 Age				
	6-11	12-13	14-15	16-20	Missing Age
Panel A: Sample attrition, KLPS-2 I-Module					
Found[a]	0.916	0.870	0.870	0.856	0.735
Surveyed	0.882	0.842	0.842	0.799	0.689
Not surveyed, dead	0.009	0.015	0.016	0.022	0.013
Not surveyed, refused	0.022	0.012	0.011	0.023	0.021
Panel B: Residential location information[b]					
Residence in Busia District[e]	0.740	0.681	0.619	0.648	0.656
Residence in districts neighboring Busia District[c]	0.047	0.071	0.098	0.071	0.156
Residence outside of Busia and neighboring districts[d]	0.213	0.249	0.283	0.281	0.188
Residence in a city	0.163	0.196	0.245	0.184	0.151
In Nairobi	0.080	0.145	0.148	0.104	0.071
In Mombasa	0.052	0.033	0.039	0.056	0.036
In Nakuru	0.004	0.009	0.018	0.016	0.000
In Kisumu	0.021	0.005	0.029	0.005	0.022
Residence outside of Kenya (in Uganda)	0.027	0.045	0.054	0.067	0.136
Number of Observations	697	1081	1020	318	247

Note: TThe sample used here includes all individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data, who were surveyed, found deceased, refused participation, found but unable to survey, or not found but searched for during intensive tracking. All figures are weighted in order to maintain initial population proportions.

a. The proportion found is the combined rates of pupils surveyed, found deceased, refused and found but unable to survey.

b. Residential location information is available for surveyed individuals only.

c. Districts neighboring Busia include Siaya, Busia (Uganda), Bugiri (Uganda) and other districts in Kenya's Western Province.

d. The categories of "Residence outside of Busia and neighboring districts", "Residence in a city" and "Residence outside of Kenya" are not mutually exclusive.

e. In 2007, Busia District was separated into three districts - Busia, Samia and Bunyala. The present definition of Busia District contains all three.

Table 3: Impact of deworming and test score on being surveyed

	Dependent Variable: Indicator for Individual Surveyed					
	(1)	(2)	(3)	(4)	(5)	(6)
Pupil test score (1998)		0.007	0.006	-0.058	0.006	-0.012
		(0.034)	(0.035)	(0.086)	(0.010)	(0.020)
Years assigned deworming	-0.008	-0.008	-0.009	-0.004	0.025	0.008
	(0.040)	(0.040)	(0.040)	(0.047)	(0.052)	(0.054)
Pupil test score * Female				0.068		0.022
				(0.087)		(0.021)
Pupil test score * Age (1998)				0.061**		0.014**
				(0.024)		(0.006)
Pupil test score * Deworming				0.016		0.004
				(0.021)		(0.005)
Deworming * Female				0.023		0.004
				(0.074)		(0.018)
Deworming * Age (1998)				-0.017		-0.003
				(0.013)		(0.004)
Age, demeaned (1998)	-0.075**	-0.075**	-0.073**	0.004	-0.019**	-0.004
	(0.033)	(0.033)	(0.033)	(0.051)	(0.009)	(0.014)
Falls sick often, self-report (1998)	-0.017	-0.017	-0.021	-0.007	-0.007	-0.006
	(0.092)	(0.092)	(0.091)	(0.091)	(0.022)	(0.021)
Household owns cattle (1998)	0.065	0.066	0.062	0.068	0.024	0.024
	(0.079)	(0.078)	(0.079)	(0.080)	(0.021)	(0.021)
Household has a latrine (1998)	-0.081	-0.081	-0.077	-0.067	-0.029	-0.025
	(0.114)	(0.114)	(0.114)	(0.113)	(0.028)	(0.028)
Weight, kg (1998)	-0.006	-0.006	-0.006	-0.011	-0.002	-0.003
	(0.009)	(0.009)	(0.009)	(0.008)	(0.002)	(0.002)
Average school participation (1998)			0.249			
			(0.328)			
Controls for gender, 1998 grade and tracking wave	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	3,363	3,363	3,363	3,363	3,363	3,363
R-squared	0.041	0.041	0.041	0.052	0.069	0.076
Mean [std dev] of dependent variable	0.833 [0.373]	0.833 [0.373]	0.833 [0.373]	0.833 [0.373]	0.833 [0.373]	0.833 [0.373]

Note: Columns (1)-(4) contain probit specifications, with marginal effects evaluated at mean values. Columns (5) and (6) contain linear probability specifications, including school fixed effects. The sample used for all regressions includes individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data, who were surveyed, found deceased, refused participation, found but unable to survey, or not found but searched for during intensive tracking. Regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Test scores are standardized within grade. Years assigned deworming is calculated using treatment group of school and individual's grade in 1998, and is not adjusted for females over the age of 13. Missing age data was replaced with mean values. All specifications include a control for missing age data, and (4) and (6) include interactions between this indicator, deworming and test score. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Tests of joint significance for the test score terms in columns (4) and (6) fail to reject the hypothesis that the coefficients are jointly equal to zero. A test of joint significance for the deworming terms in column (4) rejects the hypothesis that the coefficients are jointly equal to zero at the 1% level, and in column (6) at the 5% level.

Table 4: Summary statistics for other variables, subsample with KLPS-2 data

Variable	Mean	Std	# Obs
Female	0.477	0.500	3075
Grade (1998)	4.89	1.40	3075
Age (1998)	12.58	2.31	3075
Years of assigned deworming treatment during 1998-2003	2.94	1.77	3075
Falls sick often, self-report (1998)[a]	1.93	0.51	3075
Weight (kg, 1998)	35.08	8.58	3075
Test score (1998)[b]	0.030	0.977	3075
Highest grade attended	9.22	2.31	3075
Average school participation (1998)	0.916	0.169	3075
Years of mother's education	5.87	3.90	2549
Years of father's education	8.93	4.26	2508
Household owns cattle (1998)	0.494	0.500	3075
Household has a latrine (1998)	0.797	0.402	3075
Group 1 school	0.384	0.486	3075
Group 2 school	0.311	0.463	3075
Budalangi division school	0.349	0.477	3075

Note: The sample used here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location or age information. All figures are weighted in order to maintain initial population proportions.

a. Child falls sick often takes on values of 1 (never/rarely), 2 (sometimes), and 3 (often).

b. Test score is standardized by 1998 grade.

Table 5: Summary statistics, urban migrants versus non-migrants

	Individuals Who Have Lived in a City	Individuals Who Have Not Lived in a City	Difference
Female	0.506	0.458	0.048**
			[0.025]
Age (1998)	12.83	12.42	0.41***
			[0.11]
Years of assigned deworming treatment during 1998-2003	2.66	3.12	-0.46***
			[0.09]
Test score (1998)[a]	0.125	-0.032	0.157***
			[0.047]
Average school participation (1998)	0.931	0.906	0.026***
			[0.008]
Highest grade attained at time of survey	9.50	9.04	0.46***
			[0.12]
Falls sick often, self-report (1998)[b]	1.95	1.92	0.03
			[0.03]
Weight (kg, 1998)	37.03	34.06	2.97***
			[0.58]
Years of mother's education	6.20	5.65	0.55***
			[0.22]
Years of father's education	9.38	8.65	0.72***
			[0.23]
Household owns cattle (1998)	0.482	0.501	-0.019
			[0.025]
Household has a latrine (1998)	0.817	0.784	0.033
			[0.022]
Number of observations	1197	1878	3075

Note: The sample used here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location or age information. All figures are weighted in order to maintain initial population proportions.
a. Test score is standardized by 1998 grade.
b. Child falls sick often takes on values of 1 (never/rarely), 2 (sometimes), and 3 (often).

Table 6: Summary statistics, international migrants versus non-migrants

	Individuals Who Migrated to Uganda	Individuals Who Did Not Migrate to Uganda	Difference
Female	0.460	0.480	-0.020
			[0.038]
Age (1998)	13.10	12.51	0.61***
			[0.16]
Years of assigned deworming treatment during 1998-2003	2.77	2.99	-0.19
			[0.13]
Test score (1998)[a]	0.035	0.008	0.006
			[0.075]
Average school participation (1998)	0.930	0.921	0.016
			[0.011]
Highest grade attained	9.03	8.93	-0.22
			[0.20]
Falls sick often, self-report (1998)[b]	2.01	1.91	0.09**
			[0.04]
Weight (kg, 1998)	36.27	34.95	1.20**
			[0.60]
Years of mother's education	4.88	6.03	-1.15***
			[0.29]
Years of father's education	7.76	9.14	-1.38***
			[0.37]
Household owns cattle (1998)	0.503	0.492	0.011
			[0.038]
Household has a latrine (1998)	0.748	0.805	-0.057
			[0.038]
Number of observations	390	2685	3075

Note: The sample used here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location or age information. All figures are weighted in order to maintain initial population proportions.
a. Test score is standardized by 1998 grade.
b. Child falls sick often takes on values of 1 (never/rarely), 2 (sometimes), and 3 (often).

Table 7: Summary statistics on migration history

Means	All	Gender		1998 Age[a]	
		Female	Male	At/Below Median	Above Median
Panel A: Urban Migration					
Individuals who lived in a city during 1998-2008	0.394	0.418	0.372	0.367	0.438
Among those with information on date of move[b]:					
Number of total moves	2.44	2.62	2.27	2.41	2.49
Number of urban moves	1.31	1.33	1.30	1.28	1.35
Length of stay in urban area (yr)[c]	2.41	2.44	2.38	2.05	2.90
Panel B: International Migration					
Individuals who live in Uganda during 1998-2008	0.146	0.141	0.151	0.117	0.192
Among those with information on date of move[d]:					
Number of total moves	2.49	2.45	2.53	2.39	2.59
Number of moves to Uganda	1.22	1.24	1.20	1.18	1.26
Length of stay in Uganda (yr)[c]	3.12	3.62	2.63	2.93	3.29
Number of Observations	3075	1557	1518	1841	1234

Note: The sample used here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location or age information. All figures are weighted in order to maintain initial population proportions.

a. Median age in 1998 is 13.

b. This data exists for 89% of those who report living in an urban area during 1998-2009.

c. Note that this is an underestimate, as many of these stays are still ongoing.

d. This data exists for 80% of those who report living outside of Kenya during 1998-2009.

Table 8: Reasons for migration

Means[a]		Gender		1998 Age[b]	
	All	Male	Female	At/Below Median	Above Median
Panel A: Among Urban Migrants					
Schooling/training	0.324	0.385	0.262	0.359	0.261
To look for work	0.327	0.447	0.204	0.249	0.470
To start a new job	0.118	0.110	0.126	0.110	0.132
Marriage	0.053	0.011	0.095	0.045	0.067
Parent/guardian moved	0.021	0.023	0.020	0.021	0.022
Return to permanent home	0.003	0.002	0.005	0.003	0.004
Just visiting	0.309	0.199	0.421	0.341	0.250
Other	0.086	0.069	0.104	0.058	0.137
Number of observations[c]	969	495	474	588	381
Panel B: Among International Migrants (to Uganda)					
Schooling/training	0.357	0.346	0.371	0.440	0.260
To look for work	0.251	0.373	0.127	0.164	0.353
To start a new job	0.113	0.198	0.027	0.064	0.170
Marriage	0.175	0.011	0.352	0.177	0.174
Parent/guardian moved	0.010	0.000	0.008	0.012	0.008
Return to permanent home	0.057	0.044	0.070	0.060	0.053
Just visiting	0.163	0.096	0.231	0.210	0.107
Other	0.064	0.024	0.105	0.065	0.063
Number of observations[d]	289	145	144	163	126

Note: The sample used here includes all individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data, who were surveyed and report migration to a city during 1998-2009 (panel A) or to a foreign country (panel B). Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location or age information. All figures are weighted in order to maintain initial population proportions.

a. It is possible for respondents to move multiple times, and to have multiple reasons for each move. An indicator was thus generated to take on a value of 1 if the person migrated for a given reason, and a zero if they did not migrate for that reason. Thus, proportions likely sum to greater than one.

b. Median age in 1998 is 13.

c. Information on reasons for migration is missing for 16% of individuals reporting living in a city since 1998. Statistics presented here are fractions of the non-missing information.

d. Information on reasons for migration is missing for XXX individuals reporting living outside of Kenya since 1998. Statistics presented here are fractions of the non-missing information.

Table 9: Activities of individuals at time of enumeration, by urban migration status

Means	Living in a Rural Area	Living in a City	Among those living in a city: Gender: Male	Female	1998 Age[a]: At/Below Median	Above Median
Ever been married	0.436	0.348	0.315	0.385	0.204	0.485
Ever been pregnant[b]	0.536	0.458	0.400	0.511	0.286	0.611
In school[c]	0.237	0.146	0.172	0.116	0.206	0.096
Ever received vocational training	0.235	0.345	0.359	0.330	0.326	0.364
Working, self-employed[d]	0.145	0.100	0.099	0.102	0.059	0.139
Working, not self-employed	0.139	0.367	0.470	0.255	0.292	0.437
Unemployed[e]	0.274	0.306	0.267	0.347	0.365	0.249
Number of Observations	2481	576	307	269	255	321

Note: The sample used here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location or age information. Figures are weighted to maintain initial population proportions.

a. Median age in 1998 is 13.

b. For males, indicates "a partner has ever been pregnant with your child".

c. Respondent attended school at some time during year of survey enumeration.

d. Individuals who farm for themselves are not included among those who are self-employed.

e. Note that the working and unemployment categories do not add up to one, as the remainder of individuals are out of the labor force (which in our definition includes those engaged in agricultural activities for the home).

Table 10: Distribution of working persons by industry, by urban migration status

Means	Living in a Rural Area	Living in a City	Among those living in a city: Gender: Male	Female	1998 Age[a]: At/Below Median	Above Median
Manufacturing	0.035	0.100	0.133	0.068	0.129	0.098
Trade contractors	0.067	0.050	0.071	0.013	0.049	0.051
Wholesale trade	0.040	0.041	0.047	0.031	0.045	0.038
Retail	0.306	0.198	0.161	0.264	0.224	0.183
Restaurants, cafes, etc.	0.035	0.075	0.089	0.051	0.028	0.102
Domestic Service	0.023	0.126	0.011	0.330	0.164	0.104
Government Services	0.050	0.097	0.108	0.079	0.170	0.055
Passenger transport	0.042	0.010	0.015	0.000	0.014	0.007
Medical, dental and health services	0.009	0.004	0.000	0.012	0.012	0.000
Other services	0.087	0.189	0.228	0.122	0.109	0.236
Other	0.308	0.101	0.140	0.032	0.057	0.126
Number of observations	715	277	177	100	89	188

Note: The sample used here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data, who were either self-employed or employed by someone else at the time of survey. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations residential location or age information. Figures are weighted to maintain initial population proportions.
a. Median age in 1998 is 13.

Table 11: Summary of employment characteristics, by urban migration status

Means			Among those living in a city:			
			Gender		1998 Age[a]	
	Living in a Rural Area	Living in a City	Male	Female	At/Below Median	Above Median
Employment Status						
Permanent	0.093	0.111	0.141	0.052	0.146	0.089
Temporary	0.256	0.312	0.311	0.316	0.207	0.379
Casual	0.566	0.521	0.498	0.567	0.568	0.490
Unpaid	0.091	0.052	0.057	0.055	0.069	0.041
Working Pattern						
Full time	0.645	0.883	0.843	0.964	0.934	0.851
Part time	0.266	0.090	0.125	0.021	0.048	0.117
Seasonal	0.089	0.027	0.033	0.015	0.018	0.033
Earnings (Ksh)[b]						
Cash salary[c]	3140	5072	5525	4171	4251	5595
In kind	222	266	162	476	206	304
Benefits/allowances	204	191	172	229	94	252
Number of observations	345	224	153	71	76	148

Note: The sample used here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data, who were either self-employed or employed by someone else at the time of survey. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location or age information. Figures are weighted to maintain initial population proportions. Between August 2007 and October 2008, the average exchange rate was 0.0154.

a. Median age in 1998 is 13.

b. Earnings data is only available for individuals employed by a person or business.

c. Defined as cash salary in the previous month.

Table 12: Impact of deworming treatment and test score on urban migration

	Dependent Variable: Indicator for Ever Moved to a City						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pupil test score (1998)		0.034***	0.036***	0.035**	0.045*	0.032**	0.044*
		(0.013)	(0.014)	(0.014)	(0.024)	(0.014)	(0.023)
Years assigned deworming	0.014	0.015	0.016	0.015	0.008	0.045	0.087
	(0.011)	(0.011)	(0.012)	(0.012)	(0.015)	(0.054)	(0.054)
Pupil test score * Female					-0.004		0.001
					(0.023)		(0.023)
Pupil test score * Age at tracking					-0.000		-0.000
					(0.006)		(0.005)
Pupil test score * Deworming					-0.003		-0.004
					(0.007)		(0.007)
Deworming * Female					0.020		0.014
					(0.016)		(0.015)
Deworming * Age at tracking					0.010***		0.010***
					(0.004)		(0.003)
Age at tracking, demeaned	0.001	0.002	0.003	0.003	-0.026*	-0.002	-0.032***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.014)	(0.006)	(0.012)
Years of mother's education	0.005	0.004	0.005	0.005	0.005	0.003	0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Years of father's education	0.008***	0.008***	0.008**	0.008**	0.008**	0.008***	0.008***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Falls sick often, self-report (1998)	0.026	0.028	0.029	0.029	0.027	0.038*	0.035
	(0.022)	(0.022)	(0.023)	(0.023)	(0.023)	(0.022)	(0.022)
Household owns cattle (1998)	-0.009	-0.005	-0.006	-0.006	-0.007	-0.001	-0.002
	(0.028)	(0.028)	(0.030)	(0.030)	(0.032)	(0.029)	(0.029)
Household has a latrine (1998)	0.015	0.014	0.018	0.019	0.021	-0.008	-0.006
	(0.034)	(0.034)	(0.036)	(0.036)	(0.036)	(0.037)	(0.037)
Weight, kg (1998)	-0.000	-0.000	-0.000	-0.000	0.000	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Average school participation (1998)				0.070			
				(0.086)			
Controls for gender, 1998 grade and tracking wave	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	3,073	3,073	3,073	3,073	3,073	3,073	3,073
Adjusted R2	0.070	0.074	0.064	0.064	0.068	0.099	0.103
Mean [std dev] of dependent variable	0.394 [0.489]	0.394 [0.489]	0.394 [0.489]	0.394 [0.489]	0.394 [0.489]	0.394 [0.489]	0.394 [0.489]

Note: Columns (1), (2), (6) and (7) contain linear probability model specifications, with (6) and (7) also including school fixed effects. Columns (3)-(5) contain probit specifications, with marginal effects evaluated at mean values. The sample employed in all regressions includes surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations residential location or age information. Regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Test scores are standardized within grade. Years assigned deworming is calculated using treatment group of school and individual's standard in 1998, and is not adjusted for females over the age of 13. Missing parent education data is replaced with the mean, and all specifications include a control for missing parent education data. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Tests of joint significance for years assigned deworming and its interactions in column (4) reject the hypothesis that the coefficients are jointly equal to zero at the 5% level. These same tests in column (7), as well as tests of joint significance for pupil test score and its interactions, cannot reject this hypothesis.

Table 13: Impact of test score and educational attainment on urban migration

	Dependent Variable: Ever Moved to a City			
	(1)	(2)	(3)	(4)
Pupil test score (1998)			0.028**	0.028**
			(0.014)	(0.014)
Highest grade attended	0.018***	0.016***	0.011*	0.010*
	(0.006)	(0.006)	(0.006)	(0.006)
Years of mother's education		0.004	0.004	0.004
		(0.004)	(0.004)	(0.004)
Years of father's education		0.008**	0.007**	0.007**
		(0.003)	(0.003)	(0.003)
Age at tracking, demeaned	0.007	0.009	0.008	0.008
	(0.007)	(0.008)	(0.007)	(0.007)
Falls sick often, self-report (1998)	0.029	0.026	0.029	0.028
	(0.022)	(0.023)	(0.022)	(0.022)
Household owns cattle (1998)	-0.013	-0.013	-0.009	-0.009
	(0.029)	(0.030)	(0.030)	(0.030)
Household has a latrine (1998)	0.025	0.017	0.017	0.017
	(0.035)	(0.036)	(0.036)	(0.036)
Weight, kg (1998)	0.000	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Average school participation (1998)				0.060
				(0.086)
Controls for gender, 1998 grade, and tracking wave	Yes	Yes	Yes	Yes
Controls for years assigned deworming	No	Yes	Yes	Yes
Number of observations	3,073	3,073	3,073	3,073
Adjusted R2	0.057	0.063	0.065	0.065
Mean [std dev] of dependent variable	0.394	0.394	0.394	0.394
	[0.489]	[0.489]	[0.489]	[0.489]

This table displays probit specifications, with marginal effects evaluated at mean values. The sample employed in all regressions includes surveyed individuals with 1998 Pupil Questionnaire, school participation, ICS test score, and school attainment information. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location or age information. Regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Test scores are standardized within grade. Missing parent education data is replaced with the mean. All specifications include a control for missing parent education data. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 14: Impact of deworming treatment and test score on international migration

	Dependent Variable: Indicator for Ever Lived Outside of Kenya						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pupil test score (1998)		0.008	0.007	0.007	0.005	0.003	0.003
		(0.011)	(0.010)	(0.010)	(0.016)	(0.011)	(0.017)
Years assigned deworming	0.004	0.004	0.004	0.004	0.006	-0.011	-0.003
	(0.010)	(0.010)	(0.010)	(0.010)	(0.012)	(0.044)	(0.048)
Pupil test score * Female					0.039**		0.038**
					(0.017)		(0.019)
Pupil test score * Age at tracking					0.002		0.002
					(0.004)		(0.004)
Pupil test score * Deworming					-0.006		-0.005
					(0.005)		(0.005)
Deworming * Female					-0.009		-0.015
					(0.014)		(0.014)
Deworming * Age at tracking					0.003		0.000
					(0.002)		(0.002)
Age at tracking, demeaned	0.015***	0.015***	0.015***	0.015***	0.007	0.016***	0.015**
	(0.006)	(0.006)	(0.005)	(0.005)	(0.007)	(0.005)	(0.007)
Years of mother's education	-0.005*	-0.005*	-0.006**	-0.006**	-0.006**	-0.006**	-0.006**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Years of father's education	-0.006*	-0.006*	-0.006**	-0.006**	-0.006**	-0.005*	-0.005*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Falls sick often, self-report (1998)	0.041**	0.042**	0.042**	0.042**	0.042**	0.034**	0.034**
	(0.018)	(0.018)	(0.013)	(0.018)	(0.018)	(0.015)	(0.015)
Household owns cattle (1998)	0.016	0.017	0.017	0.017	0.017	0.014	0.014
	(0.021)	(0.021)	(0.020)	(0.020)	(0.019)	(0.020)	(0.020)
Household has a latrine (1998)	-0.035	-0.035	-0.033	-0.033	-0.033	-0.008	-0.009
	(0.029)	(0.029)	(0.027)	(0.027)	(0.026)	(0.031)	(0.031)
Weight, kg (1998)	-0.002	-0.002	-0.002	-0.002	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Average school participation (1998)				0.016			
				(0.065)			
Number of observations	3,073	3,073	3,073	3,073	3,073	3,073	3,073
Adjusted R2	0.030	0.031	0.050	0.050	0.056	0.086	0.089
Mean [std dev] of dependent variable	0.146 [0.353]	0.146 [0.353]	0.146 [0.353]	0.146 [0.353]	0.146 [0.353]	0.146 [0.353]	0.146 [0.353]

Columns (1), (2), (6) and (7) contain linear probability model specifications, with (6) and (7) also including school fixed effects. Columns (3)-(5) contain probit specifications, with marginal effects evaluated at mean values. The sample employed in all regressions includes surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location or age information. Regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Test scores are standardized within grade. Years assigned deworming is calculated using treatment group of school and individual's standard in 1998, and is not adjusted for females over the age of 13. Missing parent education data is replaced with the mean, and all specifications include a control for missing parent education data. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Tests of joint significance for years assigned deworming, pupil test score and their interactions in columns (4) and (7) fail to reject the hypothesis that the coefficients are jointly equal to zero.

Table 15: Estimation of the selection-corrected urban-rural wage gap in Kenya

	Dependent	Variable:	Wages (residual)
	(1)	(2)	(3)
Entire Sample			
Indicator for lives in a city (residual)	1,548.232***		1,483.383***
	(371.299)		(329.450)
Highest standard attained (residual)		89.807	21.333
		(68.978)	(61.914)
Pupil test score (residual)		228.377**	244.742**
		(105.443)	(101.739)
Indicator for lives in a city * highest standard (residual)			39.215
			(187.584)
Indicator for lives in a city * test score (residual)			8.136
			(312.791)
Number of observations	788	788	788
Adjusted R2	0.061	0.015	0.068
Mean [std dev] of dependent variable	1310.35 [2693.45]	1310.35 [2693.45]	1310.35 [2693.45]
Males Only			
Indicator for lives in a city (residual)	2,014.065***		1,811.030***
	(569.125)		(473.796)
Highest standard attained (residual)		80.409	-34.266
		(107.135)	(100.958)
Pupil test score (residual)		417.250***	420.657***
		(145.284)	(147.382)
Indicator for lives in a city * highest standard (residual)			166.503
			(267.506)
Indicator for lives in a city * test score (residual)			-34.094
			(428.265)
Number of observations	468	468	468
Adjusted R2	0.071	0.023	0.083
Mean [std dev] of dependent variable	1758.71 [3143.19]	1758.71 [3143.19]	1758.71 [3143.19]

Females Only			
Indicator for lives in a city (residual)	853.646**		766.045**
	(353.843)		(346.176)
Highest standard attained (residual)		100.058*	122.391*
		(58.357)	(68.800)
Pupil test score (residual)		-83.537	-69.509
		(143.358)	(83.148)
Indicator for lives in a city * highest standard (residual)			-262.346
			(197.752)
Indicator for lives in a city * test score (residual)			259.802
			(371.272)
Number of observations	320	320	320
Adjusted R2	0.052	0.007	0.065
Mean [std dev] of dependent variable	592.22 [1493.17]	592.22 [1493.17]	592.22 [1493.17]

The sample used here includes all surveyed individuals with 1998 Pupil Questionnaire, school participation, and ICS test score data, as well as information on wages. Two observations with extreme 1998 ICS test scores were dropped from the sample, as well as fourteen observations missing residential location or age information. Wages are measured as cash salary among employed individuals in the last month (replaced with zero for unemployed individuals). All variables presented here are residuals from a regression of each on the set of individual and household-level controls in Table 12.

Table 16

	Dependent Variable: Indicator for Individual Employed		
	(1)	(2)	(3)
Total Sample			
Indicator for lives in a city (residual)	0.213***		0.226***
	(0.043)		(0.043)
Highest standard attained (residual)		-0.005	-0.006
		(0.007)	(0.008)
Pupil test score (residual)		0.008	0.006
		(0.015)	(0.013)
Indicator for lives in a city * highest standard (residual)			-0.027
			(0.020)
Indicator for lives in a city * test score (residual)			0.031
			(0.045)
Number of observations	1,798	1,798	1,798
Adjusted R2	0.046	-0.000	0.049
Mean [std dev] of dependent variable	0.158 [0.365]	0.158 [0.365]	0.158 [0.365]
Males Only			
Indicator for lives in a city (residual)	0.196***		0.203***
	(0.073)		(0.076)
Highest standard attained (residual)		-0.010	-0.015
		(0.012)	(0.014)
Pupil test score (residual)		0.009	0.007
		(0.023)	(0.023)
Indicator for lives in a city * highest standard (residual)			-0.001
			(0.030)
Indicator for lives in a city * test score (residual)			0.031
			(0.053)
Number of observations	918	918	918
Adjusted R2	0.028	-0.000	0.029
Mean [std dev] of dependent variable	0.232 [0.423]	0.232 [0.423]	0.232 [0.423]
Females Only			
Indicator for lives in a city (residual)	0.231***		0.233***
	(0.063)		(0.060)
Highest standard attained (residual)		0.002	0.006
		(0.006)	(0.006)
Pupil test score (residual)		0.007	0.006
		(0.017)	(0.010)
Indicator for lives in a city * highest standard (residual)			-0.058**
			(0.025)
Indicator for lives in a city * test score (residual)			0.039
			(0.055)
Number of observations	880	880	880
Adjusted R2	0.088	-0.001	0.107
Mean [std dev] of dependent variable	0.074 [0.264]	0.074 [0.264]	0.074 [0.264]

Appendix Table 1: Existing literature on selective migration, and comparison to current study

Study	Country(ies)	Empirical ability measure	Relationship with migration
Chiquiar and Hanson (2005)	Mexico to U.S.	Schooling attainment	Positive
Grogger and Hanson (2007)	Cross country analysis	Schooling attainment	Positive
Hoddinott (1994)	Kenya (urban)	Schooling attainment	Positive
Hunt (2004)	Germany (urban)	Schooling attainment	Positive
Ibarraran and Lubotsky (2007)	Mexico to U.S.	Schooling attainment	Negative
Lanzona (1998)	Philippines (urban)	Schooling attainment	Positive
McKenzie et al (2006)	Tonga to New Zealand	Schooling attainment	Positive
Zhao (1999)	China (urban)	Schooling attainment	None
Current study:			
Hamory and Miguel (2009)	Kenya (urban)	Schooling attainment	Positive; but none conditional on a cognitive test score
		Cognitive tests	Positive
		Health status	None / weakly positive
	Kenya to Uganda	Schooling attainment	None
		Cognitive tests	None
		Health status	None