The Macroeconomics of Rural-Urban Migration

David Lagakos

UCSD and NBER

Mushfiq Mobarak

Yale University

Michael E. Waugh

New York University and NBER

February 2015

ABSTRACT

This paper uses new tracking surveys for several developing countries to analyze rural-urban migration and their macroeconomic implications. We document that migrants from rural to urban areas typically experience large consumption growth one year after migrating, though overall migration rates are low. To understand these facts we build a model that generates a rural-urban gap in average consumption due to three factors: income risk from migration, worker sorting, and disutility from migration. We structurally estimate the model and assess the relative importance of each factor in explaining rural-urban consumption gaps. We then cross-check the model's predictions using evidence from a controlled migration experiment. Quantitative experiments using the model provide guidance about the quantitative importance of migration policy on aggregate consumption growth.

Preliminary and incomplete; comments welcome. We thank the International Growth Centre for financial support, and Patrick Kiernan for outstanding research assistance. All potential errors are our own.

1. Introduction

In almost every developing country, there are large gaps in average income and consumption between rural and urban areas (for recent studies see e.g. Young (2013); Hnatkovska and Lahiri (2013); Gollin, Lagakos, and Waugh (2014).) One perspective on these gaps is that they represent a pure arbitrage opportunity, whereby reallocations of workers out of unproductive rural agricultural activities could yield huge productivity gains for some of the poorest countries in the world. An alternative perspective is that these gaps reflect efficient outcomes, driven by differences in worker composition across regions, compensating differentials or rural life, or simply the disutility of migrating to an new locale.

A natural way to distinguish between these two perspectives is to study *rural-urban migrants* in developing countries. These migrants can provide researchers with a wealth of direct information about the costs and benefits of migrating, most importantly the distribution of consumption gains experienced by migrants compared to their counterparts who stayed in rural areas. Recent experimental evidence by Bryan, Chowdhury, and Mobarak (2014) show that even temporary, seasonal moves to cities are associated with large income and consumption gains. Until recently, however, no nationally representative panel surveys with information on migrants have been conducted. The simple reason is that migrants are expensive to survey, since they have to be tracked through space.[1]

In this paper we draw on new evidence from several new nationally-representative tracking surveys from the developing world. These surveys cover a representative sample of households in some base year, and then the same sample of households one year later. Importantly, they attempt to track any individual who was sampled in the base year and then migrated in year two. The surveys are available for Tanzania, Uganda and Malawi, though we have the most results so far for Tanzania. In Tanzania, between 2009 and 2010, the surveyors successfully tracked 97 percent of all migrants, allowing one to generalize about essentially individuals that changed locations between the two survey years.

We find that rural-urban migrants experience large increases in consumption on average, with relatively few migrants experiencing consumption declines. On average, migrants doubled their consumption between survey years, while the median migrant had a consumption gain of 49 percent. Just 5 percent of rural-urban migrants had consumption declines after migrating. Relative to households that stayed in rural areas, migrants had 40 percent higher consumption

[1]Several panel surveys from developing countries exist, though to our knowledge, none are nationally representative tracking surveys. One example is the Tanzanian Kagera Tracking survey that tracked workers from one region of Tanzania from 1994 to 2004; see Beegle, De Weerdt, and Dercon (2011). In the United States, the Current Population Survey (CPS) and Panel Study of Income Dynamics (PSID) are nationally representative surveys with a panel component, but both are surveys of housing units, and hence neither tracks workers that have moved between survey years. Hence, neither survey is suitable for a study of migration.

growth on average. As a frame of reference, the cross-sectional gap in average consumption between rural and urban areas is a factor two. Thus, those that migrate from rural to urban areas experience substantial consumption gains on average, those the gains are not as high as a naive look at the cross-sectional income gap would suggest.

To understand these facts, and to better understand rural-urban gaps in average consumption, we build a model that generates a rural rural-urban gap in average consumption due to three factors. The first is income risk from migrating, which is an idea that dates back to Harris and Todaro (1971). In our model, migration risk is captured as the probability that migration is followed by a stint of unemployment of potentially large length before the migrants finds a reliable income source. The second factor is working sorting, which has been explored recently by Lagakos and Waugh (2013), Young (2013), among others. We model sorting a la Roy (1951), where workers have heterogenous productivity draws in rural and urban areas, and choose where to locate (in part) based on their comparative advantage. The final factor is the disutility of migrating, which we take as heterogenous across individuals, and could capture any cost of leaving one's existing environment and moving to a new one.

We then structurally estimate the model using the Tanzanian tracking survey. In particular, we choose parameters of the model so as to match the rural-urban gap in average consumption, the average consumption gains for rural-urban migrants relative to rural stayers, the cross-sectional variance of consumption among rural and urban stayers, and the fraction of households that migrate, among other moments. We then use the model to assess the quantitative importance of the three factors: migration risk, sorting, and disutility of migration.

To assess the model's quantitative plausibility, we cross-check the model's predictions against the experimental data from the randomized controlled trial of Bryan et al. (2014). In particular, we simulate the effect of provide a payment to a poor subset of the rural stayers conditional on migrating to the urban area. This check assures that the model's predictions for rural residents that do not migrate are as accurate as possible.

We conclude by conducting several quantitative experiments that may improve aggregate consumption growth. In particular, we simulate the effect of "rural recruitment policies" that reduce the migration risk for rural residents. These policies are similar in nature to subsidies for urban producers to recruit workers in rural areas. The market may under-provide rural recruitment efforts because the firms doing the recruiting bear all the cost of recruitment, but enjoy on a share of the benefits, since workers are free to change jobs once settled in the urban area. We also simulate the effects of large-scale conditional transfers for migration. For both policies, the outcome of interest is the increase in average consumption and the increase in average welfare from the policy.

2. Facts on Rural-Urban Migration and Consumption

2.1. Tracking Surveys

Our main source of data are the panel surveys available through the World Bank's Living Standards Measurement Studies (LSMS). These surveys are available for several recent years for Tanzania, Uganda and Malawi. In what follows, we focus on the panel survey for Tanzania, which cover a nationally representative sample of 3,265 households in the years 2009, 2010 and 2011. A notable feature of these data is that track each surveyed individual over time, even if that person migrates from one location to the next. This feature allows us to compare consumption and other measures of those that migrated versus those that chose to remain in their initial location. The surveys have a very high success rate, successfully tracking some 97 percent of those that migrate from one year to the next.[2]

We use the Tanzanian survey data to construct statistics at both the household level and the individual level. Consumption measures are defined at the household level, and hence we report those at the household level. Migration variables, as well as demographic characteristics, are recorded at the individual level, and many of our statistics about migration are computed at the individual level. We also aggregate migration information information to the household level, as in many cases multiple members of a household migrate jointly. We define a household to be a migrant (say a rural-urban migrant) if more than 75 percent of the members in the second year are new to the household.

When focusing on individual-level variables, we restriction attention to individuals aged 15 or higher. Those younger than 15 are much less likely to be making their own decisions abut migration, labor supply or other variables in which we are interested. We also drop all individuals that have missing values to key variables, such as migration status or educational attainment, though there were few such individuals. We also drop the small percentage households that had a missing consumption aggregate. For much of what follows, we divide households into two geographic categories: urban and rural, which are defined based on population density.

As our main measure of a household's living standard, we use the consumption aggregate calculated as part of the survey. The consumption aggregate is constructed at the household level to be the average consumption of durables and non-durables, including the value of home production, per household member. For most households, the primary components are purchased food and home-produced food items. The value of home-produced food and other items are calculated from questions about the quantity of each food item produced (say in kilograms),

[2]The surveys were conducted by the national statistical agencies of the countries under study with assistance from the World Bank and other organizations. The data are all publicly available from the World Bank's website. To our knowledge these surveys are the only nationally representative panel surveys that track individuals through space.

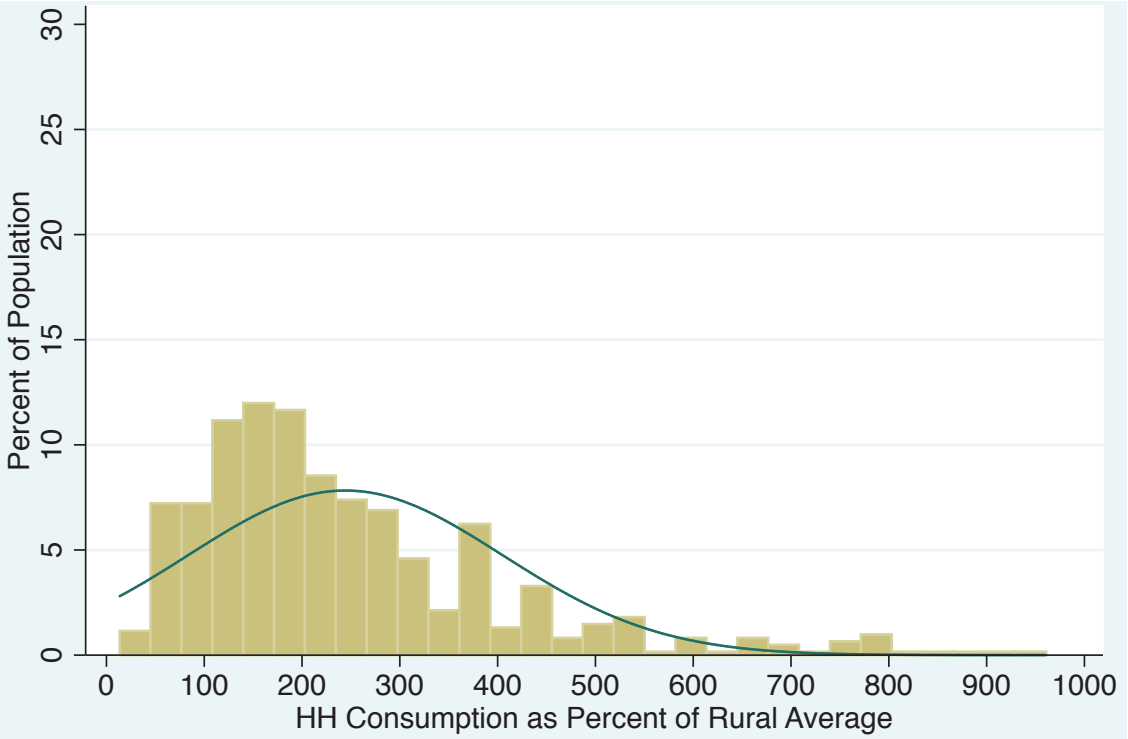

Figure 1: Cross Sectional Distributions of Consumption

plus local price information for these food items.

2.2. Cross-Sectional Distribution of Consumption

[Replace above figure with rural on top and urban on the bottom. Or, better yet, rural on the left and urban on the right.]

Figure 1 plots the distribution of household consumption in the rural and areas of Tanzania in 2009, the first year of the survey. For exposition, we divide each household's consumption

2.3. Consumption Growth by Migration Status

Figure 2 plots the distribution of consumption growth from 2009 to 2010, by migration status, using the panel dimension of the survey. The upper left panel plots the distribution of consumption growth for rural stayers, i.e. households that were located in a rural area in both years. These are the largest category of households, constituting xx% of the total. The mean rural household had consumption growth of factor 1.35 (???), meaning growth of 35 percent. The standard deviation was yy%, and zz% of the households experienced consumption losses.

The lower left panel plots the distribution of consumption growth for rural-urban migrants. On average, households in this category increased their consumption by factor of 1.63 and the

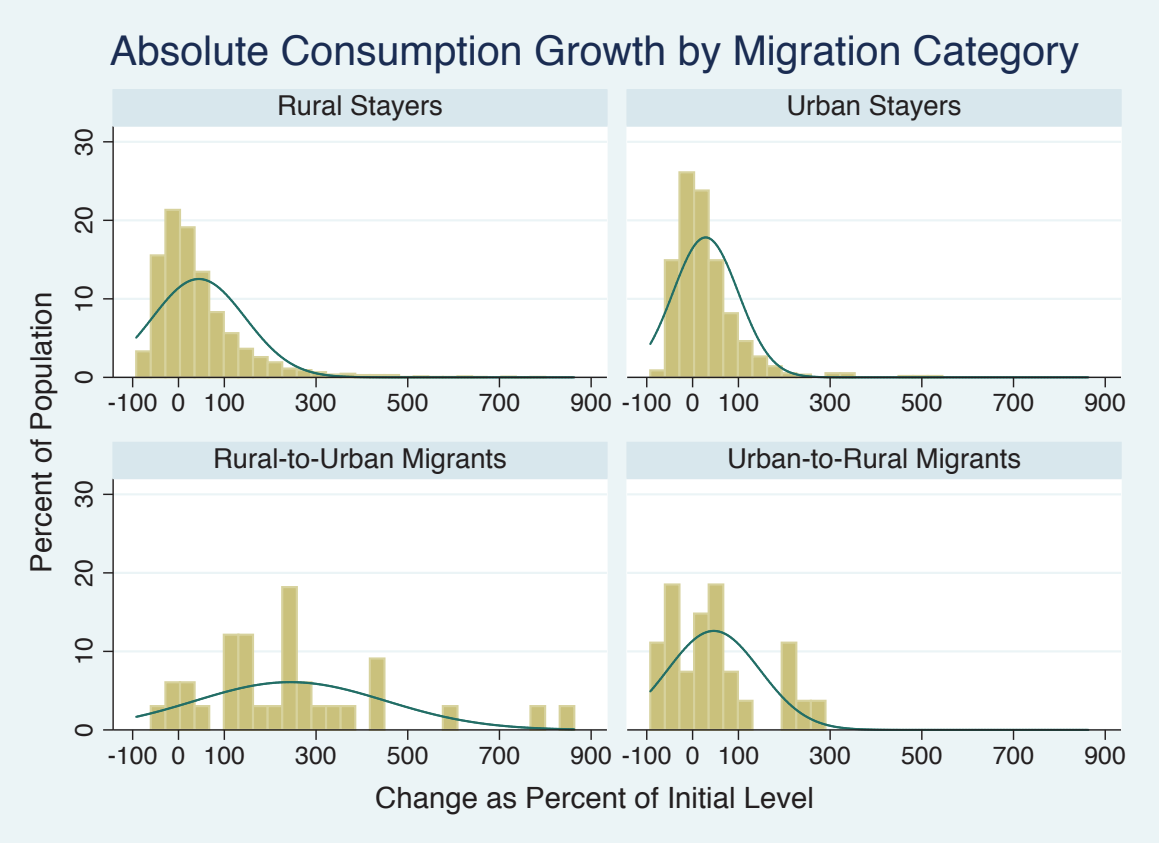

Figure 2: Consumption Growth by Migration Status

median growth was similar. The standard deviation was bb%, and just cc% of migrants had lower consumption in 2010 than in 2009, when they were located in the rural area. Thus, at least when looking one year later, households that migrated to urban areas had much larger consumption gains on average than households that stayed in rural areas. Furthermore, few households that migrated had consumption drops relative to one year earlier.

The final two panels of Figure 2 plot consumption growth for urban stayers and urban-rural migrants, the smallest category of individuals. The average urban stayer had a consumption gain of xx%, and the urban-rural migrants had gains of yy% on average. The variance of consumption growth is larger for urban-rural migrants, and fewer urban stayers experienced consumption losses than urban-rural migrants, at aa% and bb%, respectively.

Several caveats are in order about these data. First, given that the surveys are one year apart, they potentially miss migration stints that last less than one year. For example, a household that migrates for several months but then returns to the rural area, would be counted as a rural stayer. To the extent that these ex-post temporary migrants return because they experience a bad outcome in the urban area, this would likely be missed in these data.

To get some sense of the magnitude of these shorter term migration stints, we calculate the

fraction of all rural households in 2010 that report having lived in a different household for more than one of the last twelve months.

2.4. Determinants of Migration

3. Model

This section develops an model of migration choice. The model is designed to allow for several different economics forces that will dictate the migration decision: selection, risk, and moving costs. We then use the model to study two questions: Why do large differences in income and consumption persists across locations? What are the macroeconomics consequences of these differences?

3.1. Preferences, Moving, and Budget Sets

Households are infinitely lived and seeks to maximize expected discounted utility

$$\mathtt{E}\sum_{t=0}^{\infty}\beta^t u(c_t) \tag{1}$$

where β is the subjective discount factor and $\mathtt{E}$ is the expectation operator. Households have preferences over consumption of the following form

$$u(c_t) = \frac{(c_t - \bar{c})^{1-\sigma}}{1-\sigma} \tag{2}$$

where c is the consumption of a homogenous commodity that is produced in both rural and urban area. The preferences have the following parameters: $\bar{c}$ represents the minimum consumption requirement of the household and σ controls the households preference for risk. A key feature of these preferences are that they display decreasing relative risk aversion. That is as household become richer, their relative risk aversion will decline.

Each period and given a location of the household, households supply one unit of their labor inelastically with efficiency units $zs_{t,i}$. Efficiency units depend on a permanent component and a transitory component. The permanent component z and reflects the households innate skill, education, etc. We assume that the log of the permanent component is normally distributed across households with mean zero variance σ_z. Households also face transitory shocks which depend upon the region the household supplies labor. A household working in region i transitory shocks evolve stochastically according to a m-state Markov chain with transition matrix $\mathcal{P}_i$ and an associated stationary distribution $\tilde{\mathcal{P}}_i$.

Households can move across region with (i) uncertainty about the outcome and (ii) some cost.

Uncertainty comes in two forms. First, the ability to find a job is uncertain. We assume that with probability $1-\lambda$ a worker is "unemployed" and and receives zero labor income. With probability λ a worker receives a job. Second, upon moving and conditional on receiving a job, they receive a stochastic draw for efficiency units drawn from the invariant distribution associated with the transition matrix $\mathcal{P}_u$ in the urban areas.

They face a cost to moving, denoted $m(\epsilon)$ that could be household specific. We assume this cost is drawn log-normally with mean μ_m and variance σ_m. We think of this cost as being a catch all for many things: a pure physical cost, a utility cost of being away from their home location, etc.

The household's period t budget constraint is

$$a_{t+1} + c_t + m(\epsilon)\iota_t \leq Ra_t + w_i z s_{t,i} \tag{3}$$

where R is the gross interest rate on assets, c_t is consumption, $m(\epsilon)$ is the moving cost a household faces, and ι_t is an indicator function equalling one if a household moves in period t and zero otherwise. Households will be restricted in the amount of debt they can accumulate. For now, we constrain asset holing to lie on a grid $\mathcal{A} = [0 < a_1 < a_2 \ldots]$.

3.2. Optimization

A households state variable is its permanent type z and moving cost ϵ, its asset holdings a, and its transitory shock in the region it currently resides, $s_{t,i}$. The Bellman equation for the households residing in the rural area is

$$v(z,\epsilon,a,s_{i,r}) = \max_{c,a',\iota} \left[u(c) + (1-\iota)\beta \sum_{j=1}^{m} \mathcal{P}(i,j)_r v(z,\epsilon,a',s_{j,r}) + \iota\beta v_u(z,\epsilon,a') \right] \tag{4}$$

$$\text{subject to (3).}$$

$$\text{where } \; v_u(z,\epsilon,a') = \lambda \sum_{j=1}^{m} \tilde{\mathcal{P}}(j)_u v(z,\epsilon,a',s_{j,u}) + \beta(1-\lambda)v(z,\epsilon,a',0) \tag{5}$$

The states of the household are its current asset holdings a and its current efficiency units of labor $s_{i,r}$ in the rural area. Given its state variables, the households value function depends on current utility and next periods expected value function which can take on two values depending upon the migration choice. If the household decides to remain in the rural area (i.e. $\iota = 0$), the households expected continuation value is $\beta \sum_{j=1}^{m} \mathcal{P}(i,j)_r v(z,\epsilon,a',s_{j,r})$.

If the household migrates to the urban area (i.e. $\iota = 1$), the expected value of this choice is $v_u(a')$. This value reflects the two sources of income uncertainty in the urban area. With probability λ household finds a job with the efficiency units associated with that job being drawn with

probabilities $\tilde{\mathcal{P}}_u$. With probability $1-\lambda$, the household does not find a job and receives zero labor income.

The Bellman equation for the households residing in the urban area is

$$v(z,\epsilon,a,s_{i,u}) = \max_{c,a',\iota} \left[u(c) + (1-\iota)\beta \sum_{j=1}^{m} \mathcal{P}(i,j)_u v(z,\epsilon,a',s_{j,u}) + \iota\beta \sum_{j=1}^{m} \tilde{\mathcal{P}}(j)_r v(z,\epsilon,a',s_{j,r}) \right] \quad (6)$$

$$\text{subject to } (3).$$

This is essentially the same problem as the rural household. Note that we allow for the possibility of urban households to migrate to rural areas. If this is the case (i.e. $\iota = 1$), the households expected continuation value is $\beta \sum_{j=1}^{m} \tilde{\mathcal{P}}(j)_r v(z,\epsilon,a',s_{j,r})$ where the expectation is taken with respect to the invariant distribution associated with the Markov transition matrix for the rural area.

References

BEEGLE, K., J. DE WEERDT, AND S. DERCON (2011): "Migration and Economic Mobility in Tanzania," *Review of Economics and Statistics*, 93, 1010–1033.

BRYAN, G., S. CHOWDHURY, AND A. M. MOBARAK (2014): "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 82, 1671–1748.

GOLLIN, D., D. LAGAKOS, AND M. E. WAUGH (2014): "The Agricultural Productivity Gap," *Quarterly Journal of Economics*, 129, 939–993.

HNATKOVSKA, V. AND A. LAHIRI (2013): "Structural Transformation and the Rural-Urban Divide," Unpublished Working Paper, University of British Columbia.

LAGAKOS, D. AND M. E. WAUGH (2013): "Selection, Agriculture, and Cross-Country Productivity Differences," *The American Economic Review*, 103, 948–980.

ROY, A. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–46.

YOUNG, A. (2013): "Inequality, the Urban-Rural Gap and Migration," *The Quarterly Journal of Economics*, 129, 939–993.

	(1)	(2)	(3)
Household head or spouse	-0.005	-0.003	-0.003
	(0.003)*	(0.003)	(0.003)
Male	0.000	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
Ages 25 - 40	-0.009	-0.007	-0.007
	(0.003)***	(0.003)**	(0.003)**
Ages 41 - 54	-0.010	-0.008	-0.008
	(0.004)***	(0.004)**	(0.004)**
Ages 55+	-0.011	-0.008	-0.008
	(0.004)***	(0.004)**	(0.004)**
Currently in school		0.013	0.012
		(0.004)***	(0.004)***
Primary edu. completed		0.002	0.001
		(0.002)	(0.002)
Some secondary edu. completed		0.009	0.009
		(0.004)**	(0.004)**
Secondary edu. or higher completed		0.032	0.030
		(0.010)***	(0.010)***
Consumption quartile 2			0.003
			(0.003)
Consumption quartile 3			-0.001
			(0.003)
Consumption quartile 4			0.004
			(0.003)
Constant	0.017	0.011	0.010
	(0.002)***	(0.003)***	(0.003)***
R^2	0.01	0.01	0.01
N	7,108	7,108	7,108

Table 1: Linear Probability Model of Rural-Urban Migration

Note: Dependent variable is 1 if household is a rural-urban migrant and 0 if household is a rural stayer. Regression includes all households that were rural in 2009. ***, **, and * denote results that are significant at the 99 percent, 95 percent, and 90 percent.