Preliminary and incomplete-
do not cite without permission

Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata*

Joan Hamory Hicks,[a] Marieke Kleemans,[b] Nicholas Y. Li,[a] and Edward Miguel [a,c]

First version: December 2016
This version: December 2016

Abstract

Recent research has pointed to large and persistent gaps in labor productivity between the agricultural and non-agricultural sectors in low-income countries, as well as between workers in rural and urban areas. Yet most of these estimates are based on national accounts data or repeated cross-sections of micro-survey data, and as a result typically struggle to fully account for individual selection between sectors. This paper contributes to the literature on sectoral wage gaps using unusually long-run individual-level panel data from two low-income countries, Indonesia and Kenya. Accounting for individual fixed effects leads to much smaller estimated productivity gains from moving into the non-agricultural sector (or into urban areas), reducing estimated productivity gaps by between 75 and 100 percent. Per capita consumption gaps between non-agricultural and agricultural sectors, as well as between urban and rural areas, are also close to zero once selection is accounted for. Estimated productivity gaps do not emerge up to five years after a move between sectors, nor are they larger in big cities. We evaluate whether these findings imply a re-assessment of the current conventional wisdom regarding sectoral gaps, discuss how to reconcile them with existing cross-sectional estimates, and consider implications for the desirability of the reallocation of labor across economic sectors.

---

* We would like to thank David Albouy, Lori Beaman, Ben Faber, Fred Finan, Cecile Gaubert, Seema Jayachandran, Supreet Kaur, Andrés Rodríguez-Clare, and seminar participants at the U.C. Berkeley, University of Illinois and Northwestern University for many helpful discussions and suggestions. We are grateful to Brian Feld and Vedika Ahuja for excellent research assistance. All errors remain our own.
[a] University of California, Berkeley.
[b] University of Illinois, Urbana-Champaign.
[c] National Bureau of Economic Research.

I. Introduction

The shift out of agriculture and into other more "modern" sectors (e.g., manufacturing) has long been seen as a central component of the process of economic development. This structural transformation of the economy was a focus of influential early development scholars (including Rosenstein-Rodan 1943, Lewis 1955, Rostow 1960, and Kuznets 1971), with the issue stretching back to early Soviet debates over whether to "squeeze" farmer surplus to hasten industrialization (Preobrazhensky 1921).

A more recent macroeconomic empirical literature has revived interest in these issues, typically using data from national accounts (Gollin, Parente and Rogerson 2002, Caselli 2005, Restuccia, Yang and Zhu 2008). They have documented several important patterns that help shed light on the source of income differences across countries. First, they show that the share of labor in the agricultural sector correlates strongly with levels of per capita income: most workers in the poorest countries work in agriculture while only small shares do in wealthy countries. Importantly, they demonstrate that the difference in income per worker in the non-agricultural sector versus agriculture are typically much larger in poor countries than in wealthy countries. While income per worker is only moderately larger (on average) for non-agricultural workers in wealthy countries relative to poor countries, agricultural workers are many times more productive in rich countries. This creates a sort of double whammy for poor countries: agricultural work tends to be far less productive in low-income countries, and these countries' workforces are concentrated in that sector. Note that existing studies explore both the productivity gap between the non-agricultural and agricultural sectors, as well as the closely related question of gaps between the urban and rural sectors (a distinction we return to below), and reach similar conclusions.

Several recent studies have examined the extent to which these productivity gaps across sectors can be viewed as *causal impacts*, rather than mainly reflecting *selection*, namely, the possibility that differences are driven by the fact that workers of different average ability and skill levels are concentrated in particular sectors. By a causal impact of sector, we specifically mean that a given worker employed in the non-agricultural (or urban) sector is more productive than the same worker employed in the agricultural (rural) sector. The main contribution of the current study lies in disentangling these two explanations.

If there are causal impacts of this sort, then the large share of the workforce employed in the non-agricultural sector in low income countries could be viewed as a form of input misallocation, along the lines of Hsieh and Klenow (2009) and Restuccia and Rogerson (2008). The resolution of this econometric identification issue, namely, distinguishing causal impacts from selection, is not solely of scholarly interest. To the extent that productivity gaps are causal, then the movement of population out of agricultural and rural jobs and into other sectors could durably raise living standards in low-income countries, narrowing cross-country differences. The existence of large and persistent causal sectoral productivity gaps also raises a series of questions about the source of these differences, and relatedly, the nature of the frictions that limit individual movement into more productive employment, and the types of public policies that might promote such moves (e.g., Tanzania's attempts to move rural population into towns in the 1970s), or hinder them (China's *hukou* urban residential permit system) (see e.g. Stren, Halfani, and Malombe, 2004 and Au and Henderson, 2006, respectively).

Gollin, Lagakos and Waugh (2014, henceforth GLW) and Young (2013) are two important recent studies that address this identification issue, and reach differing conclusions. GLW examine

labor productivity gaps in non-agricultural employment versus agriculture using a combination of national accounts and repeated cross-sectional data from micro-surveys, and document a roughly three-fold average productivity gap across sectors. In their main contribution, GLW show that accounting for differences in hours worked and in average worker schooling attainment across sectors – and thus partially addressing the issue of worker selection – substantially reduces the average agricultural productivity gap by one third, from roughly 3 to 2. They also find that non-agricultural productivity gaps and per capita consumption gaps based on household data tend to be somewhat smaller than those they estimate using national accounts data, possibly in part due to differences in their measures of economic activity.

To the extent that individual schooling captures the most important dimensions of worker skill, and thus largely addresses selection issues, GLW's estimates would imply that the causal impact of moving workers from agriculture to the non-agricultural sector in low-income countries would be roughly a doubling of individual productivity, a substantial effect. GLW conclude (p. 940) that: "These large agriculture productivity gaps suggest … that labor is misallocated across sectors, particularly so in developing countries. By reallocating workers out of agriculture, where the value of their marginal product is low, and into other activities, aggregate output would increase even without increasing the amount of inputs employed in production. These gains could be particularly large in developing countries, where the agriculture productivity gaps and shares of employment in agriculture are largest." Of course, to the extent that educational measures alone fail to capture important dimensions of individual human capital and skill, then controlling for it would not fully account for selection.

Young (2013) examines the closely related question of urban-rural differences in consumption (as proxied with measures of household asset ownership, education and child health), rather than productivity, and similarly estimates cross-sectional large gaps.[1] Young's interpretation of these gaps differs from GLW, in emphasizing the role that selective migration across sectors could be playing in driving them. Using Demographic and Health Surveys (DHS) that feature retrospective information on individual birth district, Young shows that rural individuals with greater schooling than average in their sector are more likely to move to urban areas, while urban-born individuals with less schooling tend to move to rural areas. Young makes sense of this pattern through a model which assumes that there is more demand for skilled labor in urban areas, and shows that this could generate two-way flows of exactly the kind he documents. Young argues that he can fully explain urban-rural consumption gaps once he accounts for sorting by education in this model.[2]

The current study directly examines the issue of whether measured productivity gaps are causal or mainly driven by selection using long-term individual-level longitudinal (panel) data on worker productivity, as well as consumption in some cases. The use of this data allows us to account for individual fixed effects, capturing all time-invariant dimensions of worker heterogeneity, not just educational attainment (as GLW do). We focus on two country cases, namely, Indonesia and Kenya, that have long-term (over 10 years) panel micro data sets with relatively large sample sizes (thousands of respondents), featuring rich measures of individual earnings in both the formal and

[1] While Young (2013) focuses on urban gaps, he sometimes employs data on non-agricultural versus agricultural differences when urban-rural data is missing, and GLW similarly utilize urban versus rural data when they lack data on non-agricultural versus agricultural sectors.

[2] Porzio (2016) argues that a model of worker sorting can explain a large share (roughly 40 percent) of intersectoral productivity gaps, considering agriculture as well as a range of non-agricultural sectors. Lagakos and Waugh (2013) similarly model how worker sorting across sectors could generate sectoral productivity differences in equilibrium.

informal sector, and high rates of respondent tracking over time (over 80 percent).[3] The two datasets we use, the Indonesia Family Life Survey (IFLS) and Kenya Life Panel Survey (KLPS), are described in greater detail below.

For both countries, we start by characterizing the nature of selective migration between both non-agricultural versus agricultural economic sectors, and between urban versus rural residence. Like Young (2013), we show that individuals born in rural areas who attain more schooling are significantly more likely to migrate to urban areas, and also more likely to hold non-agricultural employment, while those born in urban areas with less schooling are more likely to move to rural areas and into agriculture. We exploit the unusual richness of our data, in particular, the existence of measures of cognitive ability in both datasets (namely, a Raven's Progressive Matrices score), to show that those with greater ability (as proxied by this score) in both Indonesia and Kenya are far more likely to move into urban and non-agricultural sectors, even conditional on measured educational attainment. This is a strong indication that conditioning on completed schooling alone may not be sufficient to fully capture differences in average worker skill levels across sectors.

We next estimate sectoral productivity differences, and, in our main finding, show that the inclusion of individual fixed effects reduces estimated sectoral productivity gaps by between 75 and 100 percent. This pattern is consistent with the bulk of the measured productivity gaps between sectors being driven by selection rather than causal impacts.

[3] There are several other high-quality panel data sets where similar approaches could be employed, for instance, the Mexican Family Life Survey (MxFLS). We leave this for future work. A member of the OECD, Mexico is considerably richer.

We first re-produce the differences documented by GLW for Indonesia and Kenya, showing both the unconditional gaps as well as accounting for differences in labor hours and education across sectors as they do (see Figure 1, Panels A and B). These are large for both countries, at over 100 log points in all cases, implying roughly a doubling of productivity in the non-agricultural sector. We then carry out estimation that treats our data as a series of repeated cross-sections, an econometric approach related to existing estimates that do not have panel data. We show that gaps remain large in this case, on the order of 50 log points for both Indonesia and Kenya. These are somewhat smaller than GLW's main estimates, though recall that GLW's estimates using household survey data tend to be smaller than their main estimates. Conditioning on individual demographic characteristics (age, gender) as well as hours worked and educational attainment partially narrow gaps to approximately 30 log points. Finally, including individual fixed effects reduces the non-agricultural productivity gap in Indonesia to zero (precisely estimated) and to 6.6 log points in Kenya (not statistically significant). Analogous estimates show that productivity gaps between urban and rural areas are also far smaller, at zero in Indonesia and 17 log points in Kenya. The estimated productivity gaps in GLW are an order of magnitude larger than even the largest of our estimates.

Beyond robustness to both the non-agricultural/agricultural and urban/rural labor productivity distinction (as just discussed), we also obtain similar results for the gap in per capita consumption levels across sectors; show that this is not simply a short-run effect by demonstrating that gaps do not emerge even up to five years after an individual moves to urban areas; and find that productivity gaps are no larger even when considering only moves to the largest cities in Indonesia and Kenya, namely, the capitals of Jakarta and Nairobi, respectively.

Our methodological approach is closely related to Hendricks and Schoellman (2016), who use panel data on the earnings of international migrants to the United States, including on their earnings in their home country. Mirroring our main results, the inclusion of individual fixed effects in their case greatly reduces the return to international migration (by roughly 60 percent). Similarly, McKenzie et al. (2010) show that cross-sectional estimates of the returns to international immigration (to New Zealand) greatly exceed those using individual panel data, or those derived from a randomized lottery. Bryan et al (2014) estimate positive gains in consumption (of roughly 30 percent) in the sending households of individuals randomly induced to migrate within Bangladesh, although no statistically significant gains in total earnings. Bazzi et al (2016) argue that cross-sectional estimates of productivity differences across rural areas within Indonesia are likely to overstate estimates derived from panel data using movers. Other related studies on the nature of selective migration include Chiquiar and Hanson (2005), Yang (2006), Beegle et al. (2011), and Kleemans (2016).

A limitation of the current study is that we focus on two countries, in contrast to GLW and Young (2013), who admirably utilize data from dozens of countries. This is due to the relative scarcity of long-run individual panel data in low-income countries that contain the measures necessary for our analysis. That said, we find broadly similar patterns in both countries that we study, each with large populations (Indonesia 250 million and Kenya 45 million) in two different world regions, which suggests some generalizability.

Another important issue relates to the local nature of our estimates, namely, the fact that the fixed effects estimates are derived from movers, those with productivity (or consumption) observations in both the non-agricultural and agricultural (or urban and rural) sectors. It is possible

that productivity gains could be different among non-movers, an issue that we discuss in detail in section 2 below. There we argue that, to the extent that typical Roy (1951) model conditions hold and those with the largest net benefits are more likely to move, then our estimates could be upper bounds on the true causal impact of moving between sectors on productivity. That said, other forms of selection are possible, as is the possibility that very long-run and even inter-generational "exposure" to a sector could persistently change individual productivity due to skill acquisition, and this opens up the possibility that selection and causal impacts are both important. We discuss these important issues of interpretation in the conclusion.

The rest of the paper is organized as follows. Section 2 presents a theoretical framework and discusses its implications for estimating sectoral productivity gaps, including a treatment of the core econometric issue of disentangling causal impacts from worker selection. Section 3 describes the two datasets we use in the analysis (IFLS and KLPS), characterizes the non-agricultural versus agriculture and the urban-rural distinctions, and presents evidence on patterns of systematic individual selection between sectors. Section 4 contains the main empirical results on productivity gaps (summarized in Figure 1), as well as results on the dispersion of labor productivity across individuals by sector, consumption gaps, dynamic effects up to five years after migration, as well as effects in big cities versus other urban areas. The final section contains alternative interpretations of the main results, discusses explanations that could reconcile our findings with existing estimates of large sectoral gaps, and concludes.

II. Theoretical Framework

II.A. The Agricultural Productivity Gap Through the Lens of an Aggregate Production Function

We present a standard development accounting framework in order to disentangle misallocation from selection in explaining aggregate productivity gaps within a country. Following Hendriks and Schoellman (2016), we allow production in sector k to be written as $Q_k = K_k^{\alpha}(A_k H_k L_k)^{1-\alpha}$. Dropping subscripts for notational convenience, a representative firm in sector k will solve

$$\max_{K,HL} K^{\alpha}(AHL)^{1-\alpha} - R(1+\tau^K)K - Z(1+\tau^H)HL$$

where R and Z represent returns per unit of physical capital K and a labor aggregate comprised of the product of human capital per unit of labor H and quantity of labor L, respectively, and τ_H and τ_K represent wedges that prevent factors from receiving their marginal product.

Solving the first order condition with respect to the quantity of human capital yields:

$$Z = \frac{1-\alpha}{1+\tau^H}\left(\frac{K}{Q}\right)^{\alpha/1-\alpha} A$$

While we solved sectoral production function for a representative worker, the compensation per unit of the labor aggregate remains the same. An individual's income in sector k is thus given by $Y_{ik} = Z_k H_{ik} L_{ik}$. Denoting logs with lower case letters, one can write the average log-income gap across the non-agricultural (n) and agricultural (a) sectors as:

$$\overline{y_n} - \overline{y_a} = \underbrace{(z_n - z_a)}_{residual\ income\ gap=\beta} + \underbrace{(\overline{l_n} - \overline{l_a})}_{labor\ supply\ gap} + \underbrace{(\overline{h_n} - \overline{h_a})}_{human\ capital\ gap} \quad (1)$$

Thus, the agricultural productivity gap is comprised of a labor supply gap, a human capital gap, and a residual, β, which is the key parameter of interest.

This gap allows for systematic sorting of workers into sectors on the basis of intensive margin labor supply and differences in human capital. Young (2013) argues that two-way migration between urban and rural areas is strong evidence that sorting on the basis of skill is what drives the urban-rural gap and not the labor supply gap or a residual productivity gap. Utilizing a mix of micro and macro data, GLW argue strongly that labor inputs across sectors are quite similar and therefore unlikely to be driving the aggregate gap. However, they find that well-measured observed average schooling differences are unable to explain away a residual gap, and they continue to find large aggregate gaps across countries.

The residual gap β captures not only wedges that directly prevent equalization of marginal products of labor between sectors, but also wedges that may impact wages indirectly by causing misallocation in capital. These wedges are the focal point of many of the theoretical models in structural transformation (see e.g. Restuccia, Yang, and Zhu 2008, and Graham and Temple 2006), a summary parameter for a country's degree of underdevelopment. In what follows, we do not take a stand on specific components of this gap.

We assume that an individual's human capital takes a Mincerian form $H_i = \exp[\boldsymbol{x_i'b} + \eta_i]$ where $\boldsymbol{x_i}$ is a vector of observed characteristics (e.g., years of schooling) with corresponding returns $\boldsymbol{b}$, and η_i represents unobserved individual skill. Substituting into our wage equation, this suggests that average log wages in sector k can be written as

$$y_i = z_a + \beta 1[\![k = n]\!]_i + l_i + \boldsymbol{x_i'b} + \eta_i \tag{2}$$

The agricultural productivity gap becomes

$$\overline{y_n} - \overline{y_a} = \beta + \left(\overline{l_n} - \overline{l_a}\right) + (\overline{\boldsymbol{x_n}} - \overline{\boldsymbol{x_a}})'\boldsymbol{b} + (\overline{\eta_n} - \overline{\eta_a}) \tag{3}$$

It is immediate that differences in unobserved components of human capital per worker will be absorbed into the residual wage gap here, and an OLS estimate of β will be biased.[4]

A principal objective of GLW, understanding η_i is crucial for estimating β, even before trying to understand which frictions are at play. There are two approaches for obtaining better estimates of β. First, one can obtain a richer set of observable characteristics $\boldsymbol{x_i}$, reducing the potential for unobserved (to the econometrician) ability in determining income. Second, one can utilize panel data and estimate *within person* wage differences over time to purge the estimation of the time-invariant components of unobserved characteristics. While our estimation explores both avenues to obtain improved estimates of the productivity gap, we focus on the second approach, the fixed effects estimation.

In our dynamic setting, we assume that the agricultural productivity gap does not change over time, which would preclude time varying changes in frictions or the production function.[5] Our

---

[4] This model can be generalized to allow for sector specific human capital with $h_{ik} = \exp\left[\boldsymbol{x_i'b_k} + \eta_i\right]$ yielding an urban-rural gap described by $\overline{y_n} - \overline{y_a} = \beta + \left(\overline{l_n} - \overline{l_a}\right) + (\overline{\boldsymbol{x_n}} - \overline{\boldsymbol{x_a}})'\boldsymbol{b_a} + (\overline{\boldsymbol{x_a}})'(\boldsymbol{b_n} - \boldsymbol{b_a}) + (\overline{\eta_n} - \overline{\eta_a})$ which motivates an Oaxaca decomposition where $\boldsymbol{b_n} - \boldsymbol{b_a}$ represent different returns paid to observable characteristics in non-agriculture. Our main specifications focus on human capital differences such as those described in equation (2) rather than this more flexible conception of human capital. We will discuss possibilities and consequences of these different factor prices, as well as scope for comparative advantage, below.

Mincerian human capital equation changes slightly to become: $H_{ikt} = \exp[\boldsymbol{x_i'b} + \eta_i + \omega_{ikt}]$. Here, η_i is again unobserved skill, and ω_{ikt} is a mean zero, individual, sector-specific, time-varying shock. An individual's time-invariant human capital (which we will estimate below as an individual fixed effect) is thus $\theta_i = \boldsymbol{x_i'b} + \eta_i$. Equation (2) becomes:

$$y_{it} = z_a + \beta 1[\![k = n]\!]_{it} + l_{it} + \theta_i + \omega_{it} \tag{4}$$

where $\omega_{it} = \omega_{iat} 1[\![k = a]\!]_i + \omega_{int} 1[\![k = n]\!]_i$ and the analogue of equation (3) is:

$$\overline{y_n} - \overline{y_a} = \beta + \left(\overline{l_{nt}} - \overline{l_{at}}\right) + (\overline{\theta_n} - \overline{\theta_a}) + (\overline{\omega_{nt}} - \overline{\omega_{at}}) \tag{5}$$

Here, the time varying, sector-specific components of human capital ω_{ikt}, are potential sources of omitted variable bias. Equation (4) is our key estimating equation; we explore potential limitations and pitfalls to this estimating equation in the following subsections.

Estimating the agricultural productivity gap via equation (4) captures the role that absolute advantage and selection may play in explaining the agricultural productivity gap while remaining agnostic to the drivers of sectoral choice. However, our data also allow us to explore Lagakos and Waugh's (2013) hypothesis that comparative and absolute advantage in a Roy (1951) model of self-selection can explain sectoral productivity gaps in countries. To explore these issues, we allow a richer formulation of our Mincerian human capital equation: $H_{ikt} = \exp[\theta_{ik} + \omega_{ikt}]$, where $\theta_{ik} = \boldsymbol{x_i'b_k} + \eta_{ik}$ allows for different returns to elements of observable human capital as well as differing unobserved ability by sector. Correspondingly, we are able to compute the distributions of individual

5 This contrasts with longer-term views of development (see e.g. Herrendorf, Rogerson, and Valentinyi, 2014), but seems sensible since the time scale of our analysis (one or two decades) is still dwarfed by the time scale of historical economic development.

time-invariant human capital in each sector (θ_{ik}) as well as examine the joint distribution of these productivities to assess whether those who have an absolute advantage in both sectors also have a comparative advantage in the non-agricultural sector.

II.B. Remaining Estimation Issues Related to Selection of Sector

Departing from the general equilibrium model previously specified, consider an agent facing a choice of working in agriculture or non-agriculture. The utility v they obtain by working in sector k is given by:

$$v_{ikt} = f(y_{it}, \boldsymbol{x}_i) + \xi_{ikt}$$

where ξ_{ikt} is an independent idiosyncratic preference shock for sector k in time t. For now, we assume that these preference shocks are uncorrelated with individual level sectoral wage innovations ω_{ikt}. We further assume that the non-stochastic component of utility of the agent is linearly separable as $f(y_{it}, \boldsymbol{x}_i) = y_{it} + \boldsymbol{x}_i'\Gamma_k$.

Substituting equation (4) with the single individual specific productivity term in the random utility function, an individual will choose the non-agriculture sector $k = n$ if and only if $v_{int} - v_{iat} > 0$; the probability of this occurring is given by

$$\begin{aligned}\Pr\{v_{int} - v_{iat} > 0\} = \Pr\{&\beta + (\omega_{int} - \omega_{iat}) \\ &+(\Gamma_n - \Gamma_a)'\boldsymbol{x}_i \\ &+(\xi_{int} - \xi_{iat}) > 0\}\end{aligned} \tag{6}$$

The possible selection bias here is classic simultaneity bias: wage innovations ω_{ikt} are simultaneously determining the sectoral choice of the worker and the worker's wage. Receiving a positive productivity shock in non-agriculture ω_{int} is both positively correlated with an indicator variable for non-agriculture and positively correlated with wages, but receiving a positive productivity shock in agriculture ω_{iat} is negatively correlated with an indicator for non-agriculture and positively correlated with wages.[6]

The requirements for a convincing instrumental variable to remove all selection biases in this context are stringent. Such an instrument would ideally affect preferences for migration but be excludable from any model of wages. This rules out using local rainfall shocks as an instrument precisely because the most straightforward channel for it to operate is by changing an individual's potential wages. Moreover, the threats to identification in this panel data setting are via time-varying shocks. Both the IFLS and KLPS data provide stated reasons for migration (subsequent to the move), but in order for these reasons to be used as instruments, the data would also need to provide reasons for staying, because not moving is also a choice for the agent. The dearth of credible natural experiments in the study of migration makes the experimental variation found in Bryan et al. (2014) and McKenzie et al. (2010) all the more valuable.

In a richer formulation of human capital with comparative advantage, the modified aggregate productivity gap in (5) is

---

[6] Explicitly, estimates of the non-agricultural productivity gap are biased if $\boldsymbol{E}\{\omega_{int}|v_{int} > v_{iat}\} - \boldsymbol{E}\{\omega_{iat}|v_{iat} > v_{int}\} \neq 0$. The term that dominates will be governed by the larger of $\boldsymbol{Var}\{\omega_{int}\}$ and $\boldsymbol{Var}\{\omega_{iat}\}$. There are two ways to see this. First, one can set $\omega_{iat} = 0$ to zero, and only the upwardly biasing ω_{int} will persist (and vice versa with $\omega_{int} = 0$). Alternatively, one can assume that ω_{int} and ω_{iat} are drawn from independent normal distributions. The conditional expectation $\boldsymbol{E}\{\omega_{int}|v_{int} > v_{iat}\}$ will be directly proportional to $\frac{\boldsymbol{Cov}\{\omega_{int},\omega_{int}-\omega_{iat}\}}{\sqrt{\boldsymbol{Var}\{\omega_{int}\}\boldsymbol{Var}\{\omega_{int}-\omega_{iat}\}}} = \frac{\sqrt{\boldsymbol{Var}\{\omega_{int}\}}}{\sqrt{\boldsymbol{Var}\{\omega_{int}\}+\boldsymbol{Var}\{\omega_{iat}\}}}$. The other conditional expectation is similar, and the larger of the conditional expectations will be the one with a larger numerator.

$$\overline{y_n} - \overline{y_a} = \beta + \left(\overline{l_{nt}} - \overline{l_{at}}\right)$$
$$+(\boldsymbol{E}\{\theta_{in} + \omega_{int} | v_{int} > v_{iat}\} - \boldsymbol{E}\{\theta_{ia} + \omega_{iat} | v_{iat} > v_{int}\}) \tag{7}$$

The selection equation in equation 6 suggests that we would only observe those employed in non-agriculture who would benefit from it, i.e., $\beta + \theta_{in} + \omega_{int} - \theta_{ia} - \omega_{iat} > c_i$, where c_i is the utility cost of moving (i.e., the other terms in equation 6). While this may be the average causal effect of non-agriculture for this population—analogous to a local average treatment-on-the-treated in the program evaluation literature—extrapolating this treatment effect to the non-movers may be problematic. This is especially relevant in our fixed effects estimation which effectively estimates the productivity gap β using the wages of the movers, namely, those with productivity observations in both sectors. It is possible that one might observe positive migration flows into non-agricultural employment even in the case where the true productivity gap β was negative; in such a case, movers would consist of those with particularly large and positive returns to non-agricultural relative to agricultural employment ($\beta + \theta_{in} + \omega_{int} - \theta_{ia} - \omega_{iat} > c_i$), or who face sufficiently large idiosyncratic preferences for the move (among those with c_i negative).

In the face of one-way selection, fixed effects estimates, which can be thought of as the treatment effect on the treated (the movers) will be generally larger than the average population treatment effect, by this logic. This suggests that estimated effects based on those who were initially in the agricultural (or rural) sector are likely to be upper bounds on the magnitude of the true average productivity gap in the population as a whole. This will be the case with the Kenya data where the entire sample is baseline rural. However, in our Indonesian setting with sorting in both directions, it is in theory possible to observe a non-agricultural premium every time an individual selects into non-

agriculture, and an agricultural premium every time an individual selects into agriculture. By a parallel logic to above, the selection equation in equation 6 suggests that, among those initially working in the non-agricultural (urban) sector, we would only observe those that benefit from working in agriculture, i.e., $-\beta + \theta_{ia} - \theta_{in} + \omega_{iat} - \omega_{int} > -c_i$. The resulting estimates would serve as lower bounds on the magnitude of the true average productivity gap.

The Indonesian data that we use (IFLS) provides an ideal testbed to understand the role of these particular biases in estimating the related urban-rural gap. In the spirit of Alwyn Young's observation that migration flows in both directions in most countries, our data allow us to condition on birth location of the individual and measure the dynamic impacts on wages after migration. The argument above predicts that the estimated urban-rural productivity gap would be larger when estimated for movers from rural to urban areas than it is when estimated for movers from urban to rural areas. This is a prediction that we take to the data below, and largely confirm. The next section describes the data in Indonesia and Kenya in detail.

III. Data

This section introduces the main data sources. This paper uses detailed panel (longitudinal) data from Indonesia and Kenya to revisit worker productivity gaps between the non-agricultural and agricultural sectors, as well as between workers in urban and rural areas. The data we use from both countries is unusually rich and the long-term panel data structure features high rates of respondent tracking over time. At 250 million, the Southeast Asian country of Indonesia is the fourth most populous country in the world, and Kenya is among the more populous countries in Sub-Saharan Africa, with approximately 45 million inhabitants. The high tracking rates of the datasets we employ allow us to construct multiyear panels of individuals' location decisions with high coverage.

Moreover, both data collection efforts include employment information on formal as well as informal sector employment. The latter is often difficult to capture in standard administrative data sources, yet often employs a large share of the labor force in low income countries. If informal employment is more common in rural areas and in agriculture, and is partially missed in national accounts data, this might generate a bias in measured sectoral and geographical productivity gaps.

IIIA. Indonesia

Detailed individual and household-level data were collected in four rounds of the Indonesia Family Life Survey, henceforth "IFLS," between 1993 and 2008 (Strauss et al., 2004). The survey is representative of 83 percent of the country's population and includes 28,841 individuals. The sample from the first survey conducted in 1993 included individuals from 13 of the 27 provinces who have since moved throughout Indonesia. Subsequent rounds of data collection were conducted in 1997-1998, in 2000, and in 2007-2008.

Attrition is often high in panel data; however, with an intensive focus on respondent tracking over space and time, the IFLS is uniquely well-suited to study migration. In particular, the panel data is characterized by low attrition rates of less than five percent in the 15 years between the first and third rounds (Thomas et al., 2001).

Detailed employment data were collected during each survey round. In addition to current employment information, the survey included questions on previous employment, allowing us to create an annual employment panel at the individual level, in line with the migration panel. Employment status and sector of employment are available for each year, but in the fourth IFLS round, earnings were collected only for the current job. Therefore, the panel has annual data on

employment status and sector of employment from 1988 to 2008, and earnings data annually from 1988 to 2000 and in the year 2008. IFLS includes information on the respondent's principal as well as secondary employment. Respondents are asked to include any type of employment, including wage employment, self-employment, temporary work, and unpaid family work. In addition to wages and profits, people are asked to estimate the value of their compensation in terms of share of harvest, meals provided, transportation allowance, housing and medical benefits, and credit; the main earnings measure is the sum of wages, profits, and all of these benefits.

For each job, individuals are asked to describe the sector that they work in. The single largest sector is "agriculture, forestry, fishing, and hunting", with 34 percent reporting working in this sector as their main employment, and 47 percent as their secondary occupation. Other common sectors are (in order of primary employment importance) "wholesale, retail, restaurants, and hotels" (21 percent), "social services" (19 percent), manufacturing (15 percent) and construction (5 percent). Men are more likely than women to work in agriculture (38 compared to 27 percent) and less likely to work in wholesale, retail, restaurants, and hotels, in manufacturing, and in social services. Smaller male-dominated sectors include construction (7 percent for men compared to 0.8 percent for women) and "transportation, storage, and communications" (7 vs. 0.3 percent).

Throughout the analysis that follows, we employ an indicator variable for non-agricultural employment. This variable equals 1 if a respondent does not have any agricultural employment, in other words, when he or she works only in the non-agricultural sector. If respondents hold any work in agriculture, for their main and/or secondary employment, the non-agricultural employment variable equals 0. In this way, individuals who work in both sectors are included in the agricultural sector. We perform several robustness checks, including categorizing the individual as working in

the non-agricultural sector when they working in both sectors, and obtain similar results (as described in the appendix).

Along with labor market data, all rounds of the IFLS collected a full history of migration within Indonesia, including all residential moves that last at least six months in duration. There is no minimum distance requirement for moves to be included; even moves within a village are reported. We combine data across IFLS rounds to construct a 21-year panel, from 1988 to 2008, resulting in 112,914 individual-year pairs from 18,068 individuals.[7] Refer to Kleemans (2016) and Kleemans and Magruder (2016) for more information on the IFLS migration panel.

When studying consumption gaps, we expand the sample to include individuals, including those both with earnings (used in the productivity analysis) as well as those without earnings or employment. IFLS consumption data was collected by directly asking households the value in Indonesian Rupiah of all food and non-food purchases and consumption in the last month, along the lines of a standard World Bank LSMS-style survey.[8] In contrast to the retrospective earnings data in IFLS, our consumption data is contemporaneous to the time of the survey. The sample used to analyze consumption includes 37,491 individual-year observations from 19,554 individuals in IFLS rounds 1–3.

We present the sample area in Indonesia, with each dot representing an IFLS respondent's residential location in Figure 2, Panel A. Location data is available at three geographical levels: the

[7] The panel is unbalanced due to sample attrition, death, and limiting observations to those where the respondent is at least age 16.

[8] These data are currently used directly without spatial or temporal price index adjustments; we will include these adjustments in future versions of this study. However, note that given that urban prices tend to be higher than rural prices, including a spatial price adjustment is likely to only strengthen our main findings.

province, district ("kabupaten") and subdistrict ("kecamatan"). While the location of respondents within Java is particularly dense, we observe considerable geographic coverage throughout the country. For our analysis, we utilize a survey-based measure of an individual's urban status: if the person reported living in a village, we define the area to be rural and if they answered "town" or "city," they are defined to live in an urban area.

We next present the correspondence between this survey-based urban residential indicator variable and employment in the non-agricultural sector (as defined above), in Table 1, Panel A. In 60 percent of individual-year observations, people are employed in the non-agricultural sector, and in 20 percent of the observations, they live in urban areas. One can see that a substantial portion of rural employment is in both agriculture and non-agricultural work, while urban employment is almost exclusively non-agricultural.

Given the migration focus of the analysis, it is useful to report descriptive statistics both for the full analysis sample, as well as separately for individuals in four mutually exclusive categories (Table 2, Panel A): those who always reside in rural areas throughout the IFLS period ("Always Rural"), those who move from rural to urban areas at some point ("Rural-to-Urban migrants"), those who are "Always Urban," and finally, the "Urban-to-Rural migrants." As discussed above, the fixed effects analysis is be driven by individuals who move between sectors.

In the full IFLS sample, 80 percent of individuals had completed at least primary education, and a quarter had completed secondary education. However, levels of tertiary education remain quite

low, at less than 10 percent. Among those who are baseline rural[9] in columns 2 and 3 (of Table 2, Panel A), we see that migrants to urban areas are highly positively selected in terms of both educational attainment, and in terms of cognitive ability, with Raven's Progressive Matrices exam scores nearly 0.4 standard deviation units higher among those who migrate to urban areas, a substantial effect.[10] These relationships are presented in a regression framework in Table 3, Panel A (in columns 1 through 5), and the analogous relationships between education, cognitive ability and moves out of the agricultural sector and into non-agricultural employment are also evident (Table 4, Panel A). Importantly, the relationship between higher cognitive ability and likelihood of migrating to urban areas holds even conditional on schooling attainment and demographic characteristics (in column 7 of both tables), at over 99% confidence. This is indicates that selection on difficult to observed characteristics is a relevant issue in understanding sectoral productivity differences in this context, as suggested in the estimation framework developed in Section 2.

It is worth noting that, if we naively classify individuals on the basis of original rural and urban status and ignore who migrates, we observe that individuals who live in urban areas at baseline appear far more skilled than those who live in rural areas at baseline. As a stark contrast, "Always Urban" individuals score over 0.3 standard deviation units higher on Raven's matrices and have more than double the rate of secondary education and triple the rate of tertiary educational completion relative to "Always Rural" individuals.

---

[9] In this version of the paper, we have defined baseline urban and rural status based on the individual's residential location in their earliest observation in the IFLS data. In subsequent versions, we plan to classify individuals based on birth location. We do not anticipate that this will substantially change the analysis given the extensive correspondence between these two measures.

[10] Raven's Matrices were administered to only a subset of individuals in IFLS 3 and 4.

The urban-to-rural migrants in Indonesia are also negatively selected relative to those who are remain urban residents, which corroborates Young's (2013) claim that rural migrants are often negatively selected. These patterns emerge in Table 2, Panel A, where the urban-to-rural migrants are worse along all skill dimensions relative to those who remain urban, and appendix Tables A1 and A2 report regression results analogous to Tables 3 and 4, conditioning on those individuals who were urban residents at baseline.

IIIB. Kenya

The Kenya Life Panel Survey (KLPS) follows 8,999 individuals who attended primary school in western Kenya in the late 1990s and early 2000s through adolescence and early adulthood. These individuals are a representative subset of participants in one of two primary school-based randomized interventions: a scholarship program for upper primary school girls in 2001 and 2002 (Kremer, Miguel, and Thornton 2009) and a deworming treatment program for upper and lower primary school students during 1998–2002 (Miguel and Kremer 2004). To date, three rounds of KLPS data collection have been completed, during 2003–2005, 2007–2009, and 2011–2014, respectively.

Two key issues in any longitudinal data collection endeavor are representativeness and attrition. The KLPS sample contains a randomly selected subset of children enrolled in primary school in Busia, a rural district of western Kenya, at the time of the deworming or scholarship program launch. According to 1998 DHS data, 85 percent of children in Western Province aged 6–15 were enrolled in school at that time, making the sample generally representative of school-age children in the region. Lee et al (2015) shows that this area is quite representative of rural Kenya as a whole in terms of socioeconomic and educational characteristics.

KLPS data collection was designed with careful attention to minimizing bias related to survey attrition. Sample individuals who had left the original study area were tracked throughout Kenya, as well as neighboring Uganda. In addition, respondents were sought in two separate "phases" of data collection where the "regular tracking phase" proceeded until over 60 percent of target respondents had been located. At this point a representative subset of approximately 25 percent of the remaining sample was chosen for tracking during the "intensive tracking phase" (and the remaining unfound individuals were no longer sought). Survey weights were then created to effectively weight these intensive individuals nearly four times as much in the analysis, to maintain representativeness with the original sample. Overall effective tracking rates for each KLPS round are roughly 85 percent.[11]

Similar to the IFLS, the KLPS includes detailed information on educational attainment, labor market participation, and migration choices over time. Employment data was collected in a wage employment module, a self-employment module, and an agricultural home production module. Most individuals were quite young (typically teenagers) during data collection for KLPS Round 1, and thus only limited information on employment and self-employment was collected at that time. In contrast, full employment and self-employment histories, including a much more detailed set of questions, were collected during Rounds 2 and 3. It is from these two rounds that we draw data for this study. The latter rounds also included detailed information on agricultural home production, though this information was restricted to the 12 months preceding the survey rather than a full history.

[11] Please refer to Baird et al. (2008) for an explanation and calculation of the effective tracking rate.

Many individuals perform some agricultural activities for home production. In addition, agriculture is a prominent sector for wage employment, at 17 percent of wage earners. Whenever annual agricultural sales exceeded 40,000 Kenyan Shillings (approximately US$400), agricultural production was also counted as self-employment (rather than home production) and included in the self-employment module. In those cases, recall data on agricultural production in previous years is included in the monthly panel.

KLPS includes information any residential moves of a least four months in duration, a slightly more permissive definition than in IFLS. In the IFLS, the exact calendar month of relocation is often missing, but this is not the case in the KLPS, allowing us to more consistently construct a monthly panel. Combined with our retrospective employment data, we construct a monthly panel with 127,254 individual-month observations from 4,439 individuals aged 16 and above with information on location and earnings measures.

Because the KLPS does not contain a survey-based measure of urban/rural status, we define a location based measure. Respondents live in an urban area if they live in a county: (a) with a population size of at least 1,000,000; (b) with a density greater than 1,000 people per square kilometer and/or (c) with a central city with at least 250,000 people. Appendix Table A10 contains the list of all counties we classify as urban; it is immediately apparent that the vast majority of urban residential moves are to Kenya's largest cities (namely, Nairobi, Mombasa, Nakuru and Mombasa).[12]

[12] In future versions of this paper, we will explore robustness of results to alternate definitions of urban residence, for instance, dropping the condition on population density and total county population. We do not anticipate that this will change the main results presented here.

According to self-reports, most individuals in our KLPS sample move for jobs or job search (57 percent). Men are more likely to migrate for employment reasons (60 compared to 54 percent) and women are more likely to migrate for family reasons including marriage (13 percent for women compared to 1 percent for men). Approximately 6 percent of individuals have moved for education.

Summary statistics on sectoral and geographic choice for KLPS respondents, the correspondence between urban and rural residence, and non-agricultural employment, are presented in Table 1, Panels B and C. Panel B includes activity in subsistence agriculture from the agricultural module for economic activity contemporaneous to the time of survey, but as a result they focus on current wage and self-employment module data and exclude the retrospective data; Panel C includes all the retrospective information, but then must exclude subsistence agricultural employment captured by the agricultural module (since this is only collected for the survey year). When including subsistence agriculture, the agricultural employment share of employment in rural areas is 68.2%, compared to 15.0% in urban areas, but these fall to 24.2% and 2.9% in our main analysis sample that excludes information on subsistence agricultural activity. Though far less than before, the agricultural employment share in rural areas is still substantial, and is sufficient for the estimation of agricultural productivity gaps, in this case, based on the earnings of those who are employed as agricultural labor, as well as those who have at least moderate levels of agricultural sales (as described above). We focus on the sample in panel C in the main analysis because of the importance of the long-term panel for our purposes.

The Kenya sample is somewhat less educated than the Indonesia sample (Table 2, Panel B). Recall that the Kenya sample is all rural at baseline (they were first recruited while attending rural primary schools). Very similar patterns emerge regarding positive selection into urban migration,

with levels of educational attainment and normalized Raven's matrix scores far hire among those migrate to cities. In particular, there is a raw gap of nearly 0.3 standard deviation units between urban migrants and those who are always rural. Tables 3 and 4 (Panel B) report these same patterns in the form of regression estimates, for urban migration and employment in non-agricultural work, respectively. Even controlling for educational attainment and gender, the Raven's score is highly positively correlated with urban migration (at over 99% confidence), with a substantial magnitude.

IV. Results

IVA. Main Productivity Gap Estimates

GLW estimate raw and adjusted agricultural productivity gaps of 138 and 108 log points in Indonesia, respectively (Figure 1, Panel A). The estimate of this raw gap from the IFLS is 54 log points (Table 5, Panel A). The most straightforward explanation for this discrepancy is an issue of measurement. GLW observe that, in an analysis of 10 countries, the average agricultural productivity gap was 17 log points smaller when estimated in Living Standards Measurement Study (LSMS) data that is similar to IFLS.[13] That said, the gap we estimate remains considerable.

Inclusion of control variables similar to those used by GLW to adjust macro data gaps reduces our estimate of the agricultural productivity gap (columns 2 and 3), to 44 and 28 log points. Estimating on the subsample for which we have scores from Raven's matrix tests, the gap is reduced slightly, although note the far smaller sample size in this case.

[13] Estimates come from log transformed values from the "Average" row of GLW, Table 4, i.e., $\ln 2.6 - \ln 2.2 = 0.167$.

Limiting the analysis to those who have productivity measurements at some point in time in both agricultural and non-agricultural employment, the productivity gap drops substantially to only 6 log points (column 5), suggesting that the selection on unobservable characteristics alluded to in Section 2 may play a meaningful role. Inclusion of fixed effects reduces the gap further (column 6), and using our preferred labor productivity measure, the log wage (per hour), as the dependent variable eliminates the gap altogether, with a coefficient estimate of just -0.001 (SE 0.029), in column 7.[14]

We follow a similar approach for the analysis in Kenya, where our raw productivity gap falls from 49 log points to 33 with the inclusion of GLW's controls (Table 5, Panel B, columns 1-3), to 16 log points when including an individual fixed effect. Using the preferred hourly wage measure reduces the gap to 6 log points (column 6), and it is reduced slightly further when adjusted with an urban price deflator (column 7). None of these fixed effects estimates are statistically significant at traditional confidence levels.

If columns 1 and 7 are compared in Table 5 (both panels), the agricultural productivity gap is reduced by 100 percent in Indonesia (all the way to zero), and by 88 percent in Kenya (from 49 to 6 log points). The standard errors are somewhat larger for Kenya, so the upper end of the 95% confidence interval includes a sizable gap of around 30 percent there, consistent with some non-trivial gains to non-agricultural employment. That said, even this value remains far lower than the 107 and 71 log point effects that GLW estimate for Indonesia and Kenya, respectively, once they condition on observable labor characteristics (namely, hours worked and educational attainment). As

[14] Log wage is computed as earnings divided by hours worked.

noted in the introduction, these results for Indonesia and Kenya are presented graphically in Figure 1, Panels A and B, respectively, and compared to GLW's estimated productivity gaps[15].

Table 6 presents the closely related exercise of estimating the labor productivity gap between residents of urban and rural areas. While the existing empirical literature has sometimes conflated these two gaps, Table 1 shows that employment in rural areas is not exclusively characterized by agriculture. Focusing on measuring an urban productivity difference isolates a labor market friction that continues to puzzle economists (see e.g. Bryan et al., 2014). To the extent that residential migration is a costlier activity than shifting jobs (but not homes), and the urban and non-agricultural wage premia are related but distinct parameters, one might suspect that an urban wage premium might even be more pronounced than a non-agricultural wage premium.

The microdata estimates from Indonesia and Kenya appear to be consistent with this view: the raw gap reported in column 1 of Table 6 (Panels A and B) are 63.9 and 69.5 log points for Indonesia and Kenya, respectively. Similar to the agricultural productivity gap, the urban-rural productivity gaps falls when additional control variables are added in columns 2, 3 and 4, but remains substantial and statistically significant. Focusing the analysis only on those who have earnings measures in both urban and rural areas (column 5) leads to a further reduction. Finally, the urban-rural earnings gap falls to 2 log points with the inclusion of individual fixed effects in Indonesia, and to -1 log point for the preferred log wage measure (column 7). The analogous urban productivity effect estimate for Kenya is 17 log points (column 7). Thus the productivity gap in Indonesia falls by 100 percent in Indonesia (to zero), and the reduction for Kenya is 75 percent (from

[15] Table A9 shows similar patterns when using an alternative definition of non-agricultural employment, classifying simultaneous work in both sectors as agriculture instead of non-agriculture. Please refer to Section 3 for further details on the definition of non-agricultural employment.

69.5 to 17.2 log points, across columns 1 and 7) with the inclusion of individual fixed effects. Once again, these results are summarized in Figure 1 (Panels C and D).

The selection model (presented above in section 2) predicts that estimated productivity gaps would be higher among rural-to-urban migrants than urban-to-rural migrants. Table 7 explores this hypothesis in Indonesia by separately conditioning on birth location; panel A limits the sample to those born rural, and panel B those born in urban areas. One can observe the same pattern of declining productivity gaps in each subsample for non-agriculture (first four columns) and urban (last four columns) as additional controls are included. In our preferred specifications in columns 4 and 8, productivity gaps are indeed somewhat larger for those born in rural areas (although the difference with estimates for those born in urban areas is not significant), as predicted by the sorting model. These results recall the main prediction from the model presented by Young (2013), and provide suggestive evidence for selection into migration based on absolute advantage.

Table A3 studies whether there are differences in unemployment rates and search behavior between urban and rural areas. The sample sizes differ from previous analyses because questions about job search are contemporaneous to the time of the survey and are not retrospective. In columns 1, 2, and 3 of Panel A, we defined individuals as unemployed if they are searching for work and have no earnings from wage or self-employment. They may be engaged in subsistence agriculture. The last three columns of Panel A only count individuals as unemployed if they are not engaged in agricultural home production either. This more restrictive unemployment definition leads to lower unemployment rates at 7 compared to 30 percent and suggest that unemployment rates are higher in urban areas. The dependent variable in Panel B is the number of hours a person reports to be searching for work and finds in line with Panel A that individuals engage in more job search in

urban areas. All results on earnings gaps described above are conditional on having positive earnings measures so this excludes unemployed individuals.

IVB. Productivity versus Living Standards

The previous section has established a 75 to 100 percent reduction of productivity gaps once individual fixed effects and covariates are included. The wage measures so far are closest to the individual marginal productivity of labor parameters that are the focus of most existing macroeconomic empirical literature. However, productivity and "utility" may diverge for many reasons, including price differences across regions, as well as amenities. There could be considerable individual heterogeneity in the taste for rural versus urban amenities, e.g., comforts of home, ethnic homogeneity, safety, better informal insurance, etc. in rural areas versus cosmopolitan cities with better public goods and more excitement (but downsides too – more crime!).

To get closer to differences in living standards, we draw on consumption data that was collected in the IFLS. As described in more detail in Section 3, four rounds of the IFLS included questions on the value of household consumption which is divided by the number of household members to get individual consumption measures. In our main specification in Table 8 we include all individuals for whom consumption data is available. This is a slightly larger sample size – 19,501 compared to 18,068 individuals – because we do not restrict the sample to those with positive earnings measures. The initial consumption gap between non-agriculture and agriculture is 51.5 log points. The gaps reduces considerable when including time fixed effects and control variables in column 2, and the gap reduces to only 1 log point when also including individual fixed effects in column 3, and this difference is no longer statistically significant. A similar pattern is presented for the urban-rural consumption gap in columns 4, 5, and 6. The consumption gap reduces from 53 log

points to 3 log points (not significant). Note that prices may be higher in urban areas and we do not yet adjust for such differences, although a price adjustment in urban areas would presumably only lead this estimate to be more negative, thus arguably strengthening the finding. As a result, the sectoral gaps in per capita consumption may be even smaller.

Appendix Table A4 shows the gap in food and non-food consumption in Panels A and B, respectively. The raw consumption gap is largest for non-food consumption and both see a 90 to 100 percent reduction when including covariates and individual fixed effects. Appendix Tables A5 and A6 repeat the consumption analyses on our main analysis sample for total consumption (Table A5) and broken down by food and non-food consumption (Table A6). Results are consistent in both samples.

IVC. Sector-specific Productivity---Absolute and Comparative Advantage

The main results (in Tables 5 and 6, and in Figure 1) suggest that a large portion of the agricultural productivity gap is due to both observed and unobserved productivity differences. In the conceptual framework, the richest model of human capital allowed for individual sector-specific productivity θ_{ik}. Analysis of these productivities has been given renewed focus in Lagakos and Waugh (2013), which argue that self-selection on the basis of comparative advantage plays an important role in understanding the agricultural productivity gap. In particular, in their model, comparative advantage is positively correlated to absolute advantage—the most productive workers have the most to gain to selecting into non-agriculture.

Utilizing panel data, we estimate a modified version of equation (4) replacing the individual fixed effect with an individual-sector fixed effect. We recover these residuals, and normalize the

mean of the fixed effed effects of permanent rural residents to be mean zero.[16] Figure 3 presents the joint and marginal distribution of these estimated productivities. The first panel conditions on Indonesians for whom we first observe in rural areas. We can see that rural-to-urban migrants are positively selected relative to non-migrants with an average rural wage approximately 13 log points higher than non-migrants. These individuals experience on average a 1 log point decline in their wage upon migration to an urban area.

Panel B presents the same exercise with those who were initially observed in urban areas. Here, we observe very little selection and very little difference in rural wages among movers. Finally, panel C presents results in Kenya that are analogous to panel A. Compared to Indonesia, there is much more positive selection among migrants in Kenya (31 log points), as well as a modest urban premium of 16 log points.

We interpret the relationship between urban and rural productivities with caution as the productivities are estimated residuals and may be subject to measurement error and attenuation bias. That said, all three of these charts show that absolute advantage plays a strong role in wage determination with positive and remarkably similar slopes across settings and sets of individuals. However, we observe a slope of less than 1 in all cases (slopes are between 0.6 and 0.7), which suggests that in a relative sense, those with absolute advantage are gaining somewhat less than those without. In fact, in all the graphs, roughly half the individuals fall below the 45 degree line; taken literally, this means that they experience better earnings in rural areas than urban areas. This is consistent with our central finding of zero or small positive sectoral productivity gaps.

[16] This procedure is identical in spirit to correlated random coefficient models utilized to analyze heterogeneous returns to hybrid seed adoption (Suri, 2011), and union's effects on wages (Card, 1996 and Lemieux, 1995).

It is still possible that Lagakos and Waugh's hypothesis holds if one were to include the unobserved productivity outcomes of the never migrants. Nonetheless, though the selection effect is evident in these graphs in mean differences, the rural productivities of the never migrants share overlap and common support with the rural productivities of the migrants. Thus, it would seem that if this population faces disproportionate disadvantages in urban areas, they are perhaps through channels that would not affect rural productivity.

IVD. Dynamics

In unpacking our main result, we test to see if dynamics and experience effects produce productivity gains that do not materialize right away. In particular, while our main specification includes time fixed effects which would account for overall growth of wages as the sample ages, individuals may begin to earn more after longer time spent in urban areas.

Figures 4A and 4B present event study analyses that explore whether individuals earn more after migrating. We estimate regression equations of the form

$$y_{it} = \theta_i + \delta_t + \boldsymbol{X}_{it}'\boldsymbol{b} + \sum_{\tau=-5}^{-2} \beta_\tau 1[\![k = U]\!]_{i,t-\tau} + \sum_{\tau=0}^{5} \beta_\tau 1[\![k = U]\!]_{i,t-\tau} + \gamma_{pre} \sum_{\tau \leq -6} 1[\![k = U]\!]_{i,t-\tau} + \gamma_{post} \sum_{\tau \geq 6} 1[\![k = U]\!]_{i,t-\tau} + \varepsilon_{it} \tag{8}$$

These regressions are estimated on an unbalanced panel of person-time periods and include individual fixed effects θ_i, time fixed effects δ_t, squared age as a time-varying covariate $\boldsymbol{X_{it}}$, and

indicator variables for time periods exceeding five years pre- and post-move, γ_{pre} and γ_{post}, respectively. Indonesian event studies condition on an individual's first observation being in a rural area.

The β_τ parameters of interest are coefficients on indicators for time periods relative to the period of the individuals' move $\tau = 0$. So all estimates are relative to the year and month prior to the individuals' move in Indonesia and Kenya, respectively, we exclude an indicator for the period prior to the individuals' move. These coefficients are identified by individuals who have adjacent productivity measures in both the period they move to urban and the period immediately prior—421 in Indonesia, and 308 in Kenya. We do not enforce a requirement that individuals are observed in every period five years prior and post. If extensive margin decision to exit the labor force or attrition is correlated with one's experience in urban areas, these results may be biased, and we therefore interpret these results with caution. Nonetheless, smaller sample sizes are reflected in the somewhat larger standard errors.

These parameters represent the difference in mean wages between movers and non-movers net of the difference that existed in period prior to the move. An advantage of these event study analyses is that it allows us to see the dynamics of wages *prior* to the move, which may give some clues about whether individuals are moving due to negative shocks. Panel A shows that urban wages ultimately do not change relative to the year prior to moving in Indonesia, and even five years after the move, migrants see no mean difference in wages. Panel B shows similar results relative to the month prior to the move; perhaps individuals see some small wage gains initially, but these differences ultimately fade away. Nor is there any evidence of significant pre-trends before the move.

These results are shown regardless of where individuals are. The bottom half of panels A and B show a survival rate of only about 50 percent after five years. Naturally, one might suspect that those with poor outcomes after migration might return home; Appendix Figures A2A and A2B plot separately post-move wages separately for survivors and non-survivors and we find no evidence that gains are higher even among survivors whom we might suspect would be selected on gains.

The aggregate productivity gap is a combination of selection on absolute advantage and selection on gains. The fixed effects estimates in section IVA implicitly estimate the residual gap off of both those who switch from rural to urban and urban to rural. As we suggested in section II, there is no control group; both groups of switchers are treated, the former by urban, and the latter by rural. Thus, it is possible that an average of the positive treatment-on-the-treated effect for rural-to-urban migrants would cancel a positive treatment-on-the-treated effect for urban-to-rural migrants producing an attenuated urban premium. We observe no differences in our main specification, no gains in our urban event study, and thus our finding of no differences in our rural event study in Appendix Figure A3 is not surprising. If anything, it appears the urban wage gap is higher for urban-to-rural migrants than rural-to-urban migrants, but these differences are not practically distinguishable.

IVE. Big Cities

Using panel data from Spain, De la Roca and Puga (2016) show that job experience is particularly valuable in big cities and moreover, that these cities boost productivity over time. In this section, we assess both findings for Indonesia and Kenya. Table 9 repeats the main analysis of Table 6 but adds indicator variables for the five largest cities of each country. In Indonesia, all 5 cities are larger than 2 million inhabitants, with the capital Jakarta topping the list with a population of 10 million.

Kenya's capital Nairobi hosts 3.4 million people, the second city Mombasa 1.2 million and the other three cities in the top 5 are smaller. Focusing on column 4 in Panel A and B, we do not find evidence of large earnings gaps in big cities in particular. With the exception of Bandung in Indonesia and Mombasa in Kenya, the pattern of large cities is similar to urban areas in general.[17]

The analyses in Table 9 do not find a big city level effect that De la Roca and Puga find in Spain; we also assess whether large city effects can manifest over a longer time horizon. Appendix Figures A4 repeats the event study analysis from Section 4D separately for Jakarta in Indonesia and Nairobi in Kenya, respectively. Similar to the event studies in Figure 4, these graphs capture the difference in mean wages regardless of whether the respondent continues to live in the capital city or not. These figures show no evidence of dynamic effects. If anything, Nairobi appears to have a negative experience effect relative to the path of earnings for individuals in rural areas. Nonetheless, these estimates are somewhat imprecise and we can neither rule out moderate positive nor negative dynamic effects. The bottom line is no evidence for big city effects, either immediately or over a five year time horizon.

V. Conclusion

Several influential recent studies estimate large sectoral productivity gaps in low-income countries, and highlight an apparent puzzle, namely, as Gollin, Lagakos and Waugh (2014, p. 941) write, "why so many workers remain in the agricultural sector, given the large residual productivity gaps with the

[17] We can only speculate as to why there is a residual effects for Bandung in Indonesia and Mombasa in Kenya. Both cities are known for their large tourism sector. Perhaps this sector pays particularly high wages to low-skilled individuals.

rest of the economy." This study makes two main contributions, using data from two low-income countries with large populations (Indonesia and Kenya) located in two different regions. First, we show that estimating sectoral productivity gaps – both across non-agricultural and agricultural sectors, and across urban and rural areas – using panel data and including individual fixed effects leads to a large reduction of 75 to 100 percent in estimated gaps. The second main empirical contribution lies in demonstrating that there is extensive individual selection across sectors in both settings, both along relatively easily observable dimensions such as educational attainment as well as measures of skill (here, an intelligence measure) that most standard economic databases lack.

Taken together, the findings point to the importance of individual selection in driving observed sectoral gaps in productivity and living standards, and call into question strong causal interpretations. As a result, the puzzle of why the share of workers in agriculture (and rural areas) remains so high may not be as much of a puzzle as previously thought. Similarly, if gaps are mainly driven by selection, then policies to incentivize workers to move to urban areas (and out of agriculture), based on the logic of input misallocation, would not appreciably raise aggregate living standards and would not appear to be an appropriate policy direction. So why do people migrate? Our data do not allow us to capture potential effects of migration that occur over time scales beyond a couple of decades. The never movers in urban areas appear to be positively selected even before moving, which suggest that schooling quality or ecological factors may play important roles. Recent research has shown that wages grow at different rates over the life-cycle (Lagakos et al. 2016); such returns would not be observed in our main specifications.

A historical episode illustrates some of the potential risks of pro-urbanization policies. In the 1970s, the authoritarian socialist government in Tanzania attempted to move much of its rural

population into larger villages and towns in an attempt to speed up economic modernization. The underlying idea was that the provision of public services, and the shift into non-agricultural employment (including in manufacturing) would be hastened if households would only leave their traditional homesteads, which were often located on their own farmland and thus highly spatially dispersed. After initial rhetorical encouragement by the government led to little residential movement, the government began to resort to forced migration in certain regions in 1973, in the so-called "Operation Vijiji". The resulting economic and social dislocation is today widely viewed as a policy disaster within Tanzania (Stren, Halfani, and Malombe, 1994). While one could argue that observers are unable to assess the true economic effects of attempted villagization and urbanization in Tanzania, since the forced moves were quickly abandoned (within a year) in the face of large-scale popular resistance, at a minimum the Tanzanian case indicates that it can sometimes be very costly (from a welfare perspective) to induce a large share of the population to move out of rural agriculture.

As noted above, our productivity gap estimates are derived from individual movers, namely, those with productivity measured in both sectors. Thus a logical way to reconcile our finding of zero to small sectoral productivity gaps with the existing macroeconomic empirical evidence of large average gaps is the possibility that productivity effects among non-movers are much larger than those of movers. Given the nature of our data, it is impossible to rule out this possibility, and it clearly merits further investigation, although the lack of measured individual productivity or consumption in both sectors for non-movers naturally complicates the rigorous econometric identification of these relationships.

However, several factors lean against this interpretation in our view, at least in the short-run. First, it is natural to think of the migration decision in terms of a Roy (1951) model, in which those with the largest net benefits are most likely to move. This would lead our estimates to overstate gaps between sectors overall. While it is possible that those individuals who remain in the rural agricultural sector might receive large positive earnings gains from moving, their choice not to do so might simply reflect high financial or non-financial costs to migration. For instance, the bundle of amenities found in a large city are very different than those in rural areas (along many dimensions, including access to public services, crime, and the nature of social interactions with neighbors), and individuals may have strong and heterogeneous preferences for these amenities, leading to large reductions in utility for some migrants. Poor individuals may also face credit constraints or financial frictions that prevent them from moving to exploit wage gaps, and easing these constraints could boost migration rates (as argued in the Indian case by Munshi and Rosenzweig 2016).

A promising approach to estimating the returns to migration in low-income countries among those who are typically "non-movers" and may face such constraints is the recent Bryan et al (2014) study in Bangladesh. They find that a moderate subsidy did induce a small share of recipients (roughly one fifth) to move to towns and cities for temporary work (during the agricultural low season); the relatively low rate of migration again indicates that the utility costs of migration are non-trivial. Among movers, there is an estimated increase in per capita consumption among the sending household (excluding the migrant) of roughly 30 percent over two years, although effects appear more modest when the migrant is included, and 25 percent average gains in earnings among those assigned to the subsidy (and these effects are not statistically significant). Overall, the study provides some indication that there are positive returns to temporary seasonal migration among rural

workers who are typically non-movers, although they are fairly modest in size and closer in magnitude to those we estimate in this paper than to those found in many other recent contributions.

The case of urban-born non-movers is less well understood, and raises some intriguing possibilities. Recall from Table 2 above that the individuals raised in urban areas have considerably higher cognitive scores (as measured in a Ravens Matrices test) than those raised in rural areas. It is difficult to definitively determine the causes of this gap, but there are several plausible channels. One is simply that wave after wave of rural to urban (urban to rural) migration by positively (negatively) selected individuals over many decades, combined with partial heritability of cognitive ability, have reshaped the underlying ability distributions in these two sectors. This would simply be an inter-generational extension of the patterns of individual selection across urban and rural areas that we and Young (2013) document, and would not necessarily change the interpretation of our main results.

Another explanation, which is not mutually exclusive, is that there is a lower cost to skill acquisition in urban areas, either due to improved provision of schooling for children there or something else about the nature of social interactions (e.g., the density of such interactions or other forms of intellectual stimulation in childhood). In other words, given the importance of early childhood circumstances for lifetime cognitive development (Gertler et al, 2014), growing up in a city might generate high average skill levels. This would properly be understood as a causal effect of urban residence on individual labor productivity, albeit in the very long-run and on the movers' children rather than for themselves. These effects would not be captured even in the five-year follow-up period that we consider in this study (in Figure 4), but could be contributing to large and persistent urban-rural productivity gaps overall.

In our view, this remains a research area ripe for further empirical analysis. Some natural next steps include extending our long-run panel analysis to new countries, settings and time periods (as appropriate panel data becomes available, ideally including large-scale administrative data); conducting further experiments along the lines of Bryan et al (2014) and McKenzie et al (2010) to induce at least partially random selection in migration, thus generating "local" estimates in new sub-populations and better understanding the nature of costs and constraints facing potential migrants; and exploration of very long-run and even inter-generational effects of parent sectoral and residential choice on child ability, along the lines of the exercise we report above.

Bibliography

Au, C-C and J. Henderson, 2006. "How migration restrictions limit agglomeration and productivity in China." *Journal of Development Economics* 80 (2), 350–388.

Baird, S., J. Hamory, and E. Miguel, 2008. "Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)," University of California CIDER Working Paper #C08-151.

Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong, 2016. "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," *American Economic Review* 106(9), 2658–2698.

Beegle, K., J. De Weerdt, and S. Dercon, 2011. "Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey," *The Review of Economics and Statistics* 93(3), 1010–1033.

Bryan, G., S. Chowdhury, and A. M. Mobarak, 2014. "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica* Vol. 82(5), 1671–1748.

Caselli, F., 2005. "Accounting for Cross-country Income Differences," in Aghion, P. and S. Durlauf (ed.), *Handbook of Economic Growth*, Volume 1A, 681–741.

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

De la Roca, J. and D. Puga, 2016. "Learning by Working in Big Cities," *Review of Economic Studies* (advance access).

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. Chang, S. Grantham-McGregor, 2014. "Labor market returns to an early childhood stimulation intervention in Jamaica," *Science*, Vol. 344, Issue 6187, pp. 998-1001

Gollin, D., D. Lagakos and M. E. Waugh, 2014. "The Agricultural Productivity Gap," *The Quarterly Journal of Economics*, 129(2): 939–993.

Gollin, D., S. Parente and R. Rogerson, 2002. "The Role of Agriculture in Development," *American Economic Review Papers and Proceedings*, 92(2): 160-164.

Gollin, D., D. Lagakos and M. E. Waugh, 2014. "Agricultural Productivity Differences across Countries," *American Economic Review: Papers & Proceedings*, 104(5): 165–170.

Graham, B. S. and J. R.W. Temple, 2006. "Rich Nations, Poor Nations: How Much Can Multiple Equilibria Explain?" *Journal of Economic Growth* 11, 5–41.

Hendricks, L. and T. Schoellman, 2016. “Human Capital and Development Accounting: New Evidence from Wage Gains at Migration,” working paper.

Herrendorf, B., R. Rogerson and A. Valentinyi, 2014. “Growth and Structural Transformation,” in ed. Aghion, P. and S. Durlauf, *Handbook of Economic Growth*, Volume 2B, 855–941.

Hsieh, C. and P. J. Klenow, 2009. “Misallocation and Manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, 124(4): 1403–1448.

Kleemans M. and J. Magruder, 2016. “Labor Market Changes in Response to Immigration: Evidence from Internal Migration Driven by Weather Shocks,” accepted by *The Economic Journal*.

Kleemans, M., 2016. “Migration Choice under Risk and Liquidity Constraints,” working paper.

Kremer, M., and E. Miguel. 2004. “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities,” *Econometrica*, 72(1):159-217.

Kremer, M., E. Miguel, and R. Thornton. 2009. “Incentives to Learn,” *The Review of Economics and Statistics,* 91(3): 437-456.

Kuznets, S., 1971. "Modern Economic Growth: Findings and Reflections," Nobel Prize in Economics documents 1971-2, Nobel Prize Committee.

Lagakos, D., B. Moll, T. Porzio, N. Qian, and T. Schoellman, 2016. “Life-Cycle Wage Growth Across Countries.” Working Paper.

Lagakos, D. and M. E. Waugh, 2013. “Selection, Agriculture, and Cross-Country Productivity Differences,” *American Economic Review*, 103(2): 948–980.

Lee, K., E. Brewer, C. Christiano, F. Meyo, E. Miguel, M. Podolsky, Javier Rosa, and Catherine Wolfram, 2015. "Barriers to Electrification for “Under Grid” Households in Rural Kenya", *Development Engineering* 1: 26–35.

Lewis, W., 1955. *The Theory of Economic Growth*. London: George Allen & Unwin.

McKenzie, D., S. Stillman, and J. Gibson, 2010. “How Important is Selection? Experimental vs. Non-experimental Measures of the Income Gains from Migration,” *Journal of the European Economic Association*, 8(4); 913-945.

Munshi, K. and M. Rosenzweig, 2016. “Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap,” *American Economic Review*,106(1): 46–98.

Porzio, T., 2016. “Cross-Country Differences in the Optimal Allocation of Talent and Technology”. Working Paper.

Preobrazhensky, E., 1921. *The Crisis of Soviet Industrialization* (1980 ed.). London: MacMillan.

Restuccia, D., and R. Rogerson. 2008. "Policy Distortion and Aggregate Productivity with Heterogenous Establishment," *Review of Economic Dynamics*, 11, 707-720.

Restuccia, D., D. Yang and X. Zhu, 2008. "Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis," *Journal of Monetary Economics*, 55, 234–250.

Rosenstein-Rodan, P.N. 1943. "Problems of Industrialisation of Eastern and South-Eastern Europe," *The Economic Journal*, 53(210/211): 202-211.

Rostow, W.W. 1960. *The Stages of Economic Growth: A Non-communist Manifesto.* Third edition. New York, NY: Cambridge University Press.

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

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.

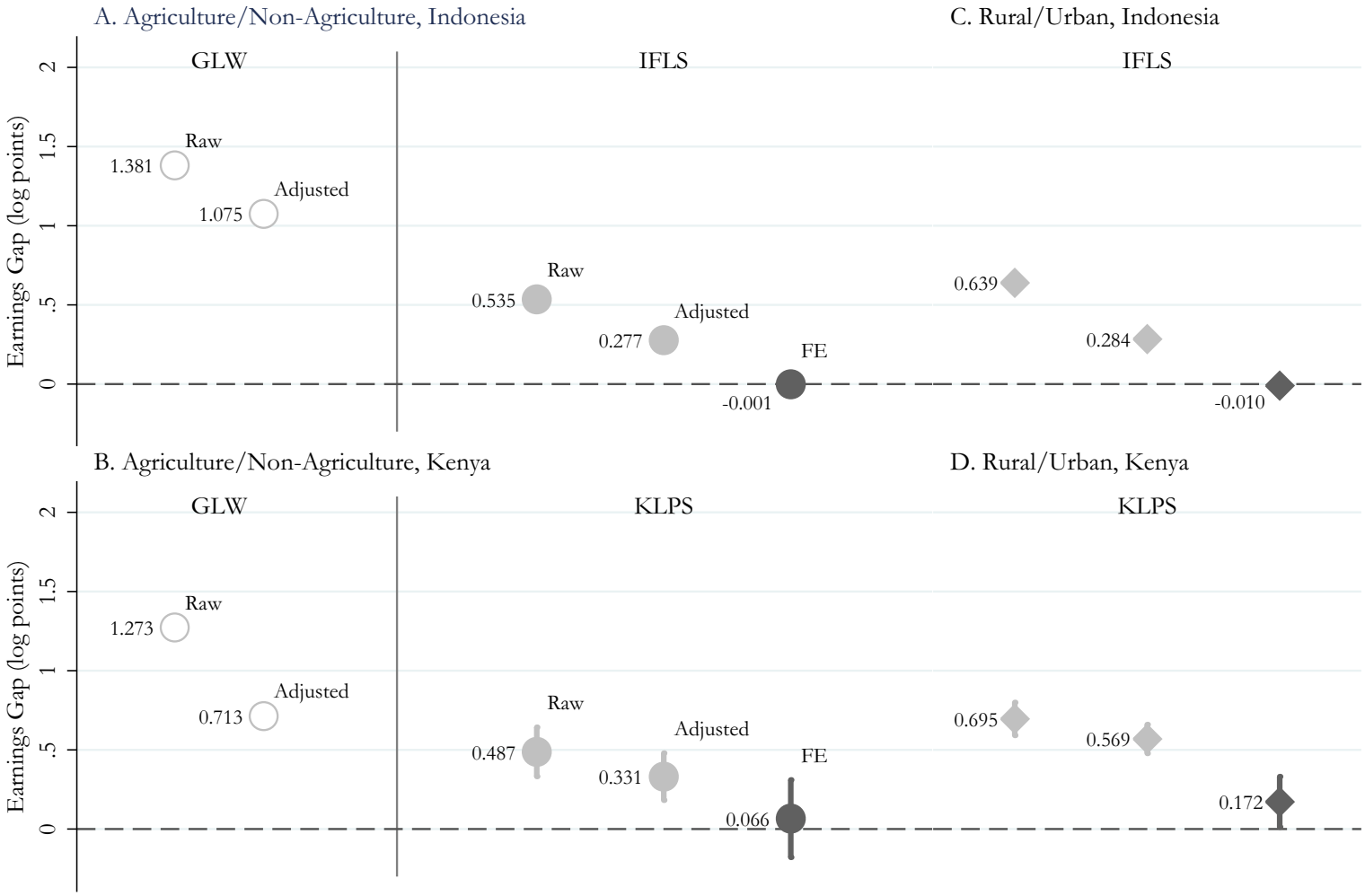
Stren, R., M. Halfani, J. Malombe, 2004. "Copying with Urbanization and Urban Policy." *Beyond Capitalism vs. Socialism in Kenya and Tanzania*. Ed. Joel D. Barkan. Colorado: Lynne Rienner Publishers.

Thomas, D., E. Frankenberg, and J. P. Smith. 2001. "Lost but Not Forgotten: Attrition and Follow-up in the Indonesia Family Life Survey," *The Journal of Human Resources*, 36(3): 556-592.

Young, A., 2013. "Inequality, the Urban-rural Gap, and Migration," *The Quarterly Journal of Economics*, 128 (4): 1727-1785.

Tables and Figures

Figure 1: Productivity Gap in Total Earnings

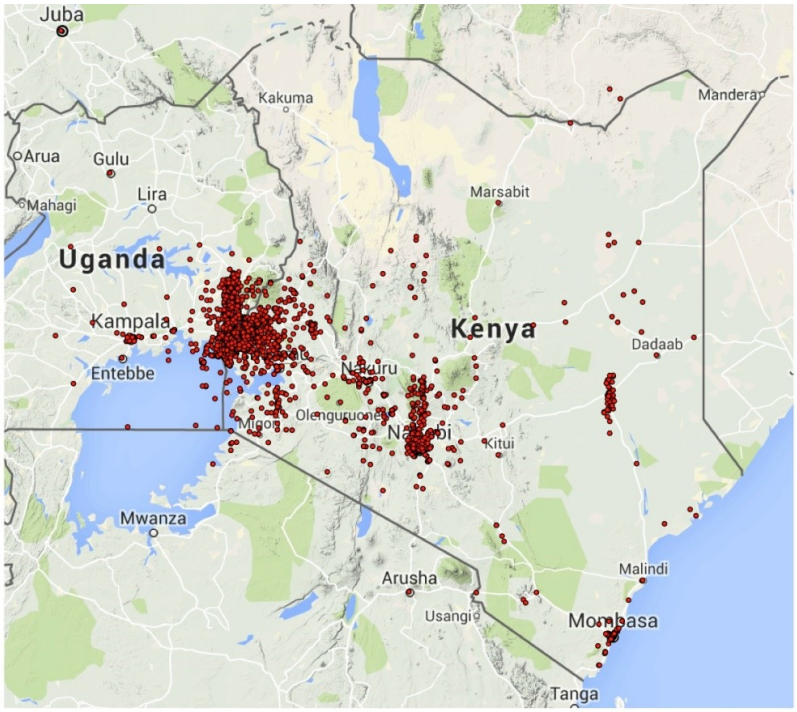


Notes: GLW refers to estimates from Gollin, Lagakos, and Waugh (2014), Online Appendix Table 4. For comparability, the figure reports log transformed numbers from their columns 4 and 5 for Indonesia and Kenya, respectively. Symbols here represent point estimates, and vertical lines represent 95% confidence intervals. Panel A estimates from IFLS come from Table 5, panel A: "Raw" is the mean difference estimate from column (1), "Adjusted" is the regression adjusted mean difference estimate from column (3), and "FE" is the fixed effects regression estimate from column (6). Corresponding estimates from KLPS come from Table 5, panel B. Estimates in panels C and D come from the same columns in Table 6, panels A and B, respectively. Note that the confidence intervals for the estimates from IFLS are smaller than the size of the symbols and are therefore not visible.

Figure 2: Sample Areas

(A) Indonesian Family Life Survey

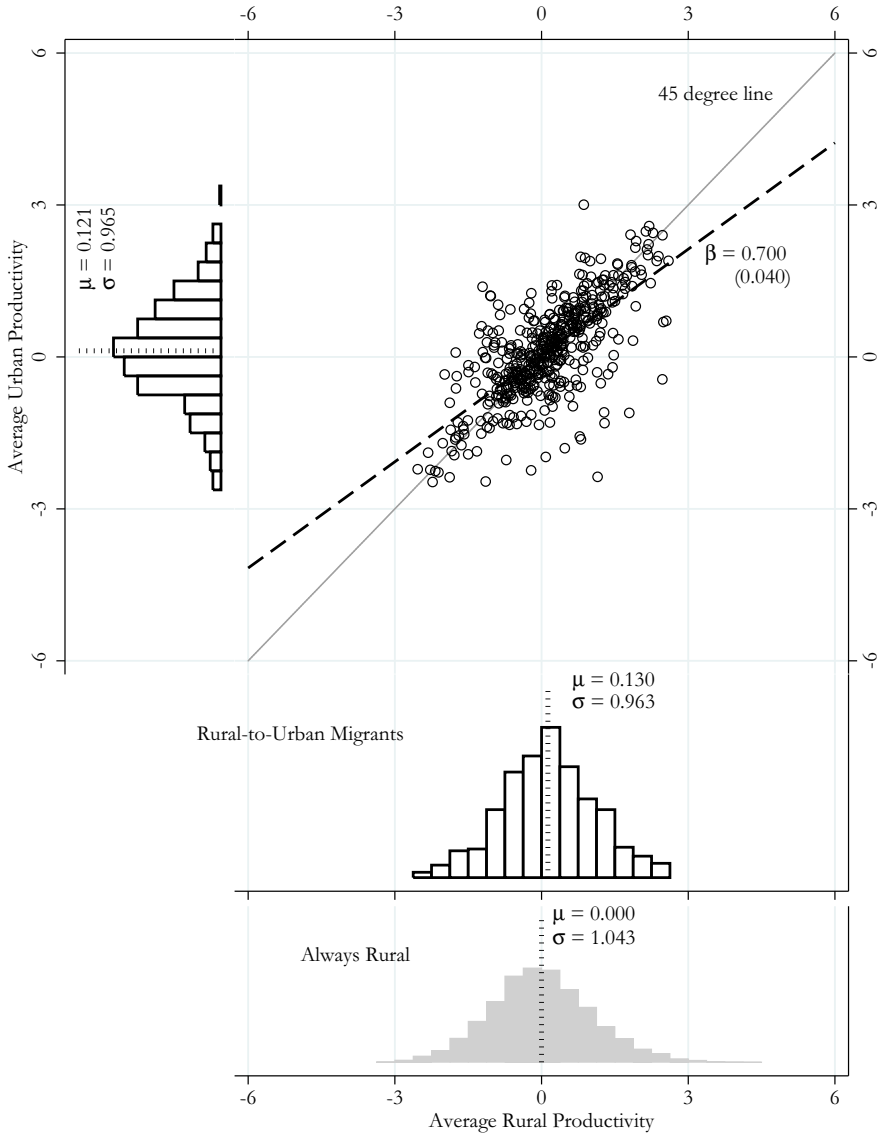
(B) Kenya Life Panel Survey



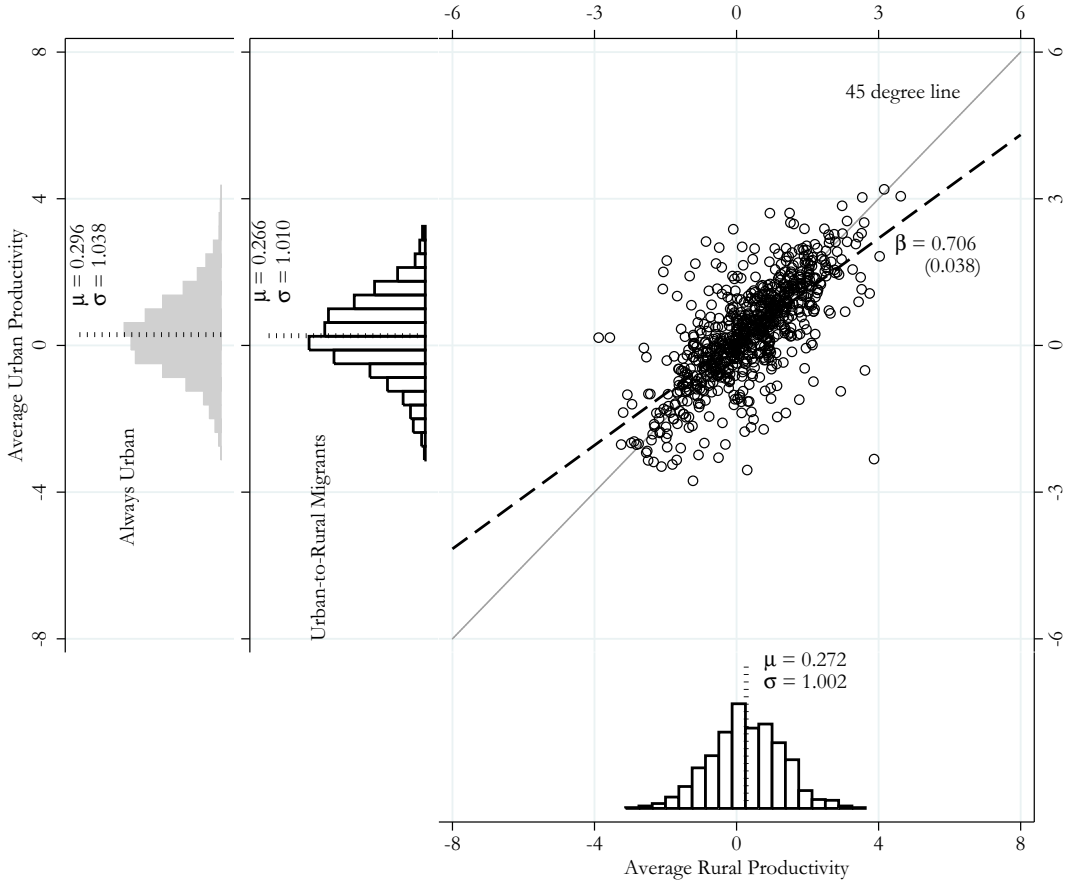
Notes: Panel A shows the residential locations of individuals during the 1988–2008 sample period of rounds 1–4 of the Indonesian Family Life Survey (IFLS). For the Kenyan sample, Panel B shows individuals' residential locations during the 1998–2014 sample period that was collected during rounds 2–3 of the Kenya Life Panel Survey (KLPS). The location information of both datasets are described in more detail in Section 3.

Figure 3: Joint Distribution of Rural and Urban Productivities

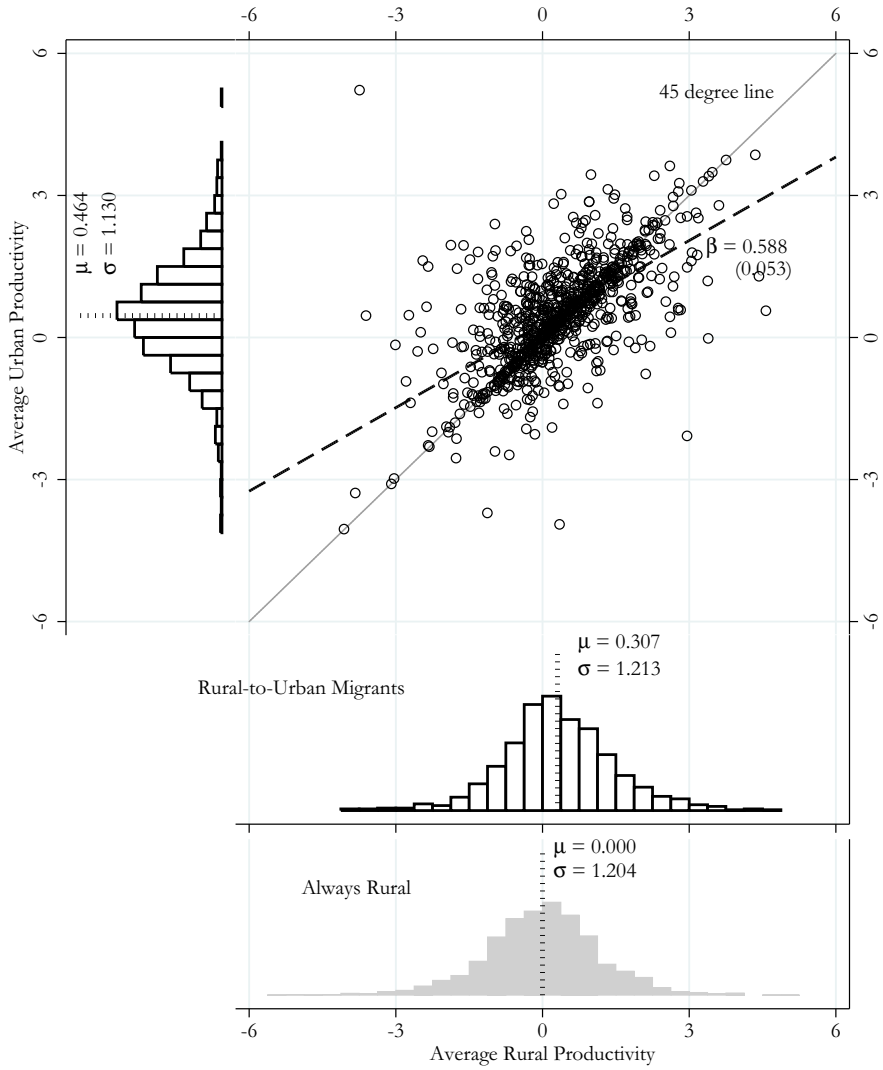
(A) Indonesia (Baseline Rural)



(B) Indonesia (Baseline Urban)



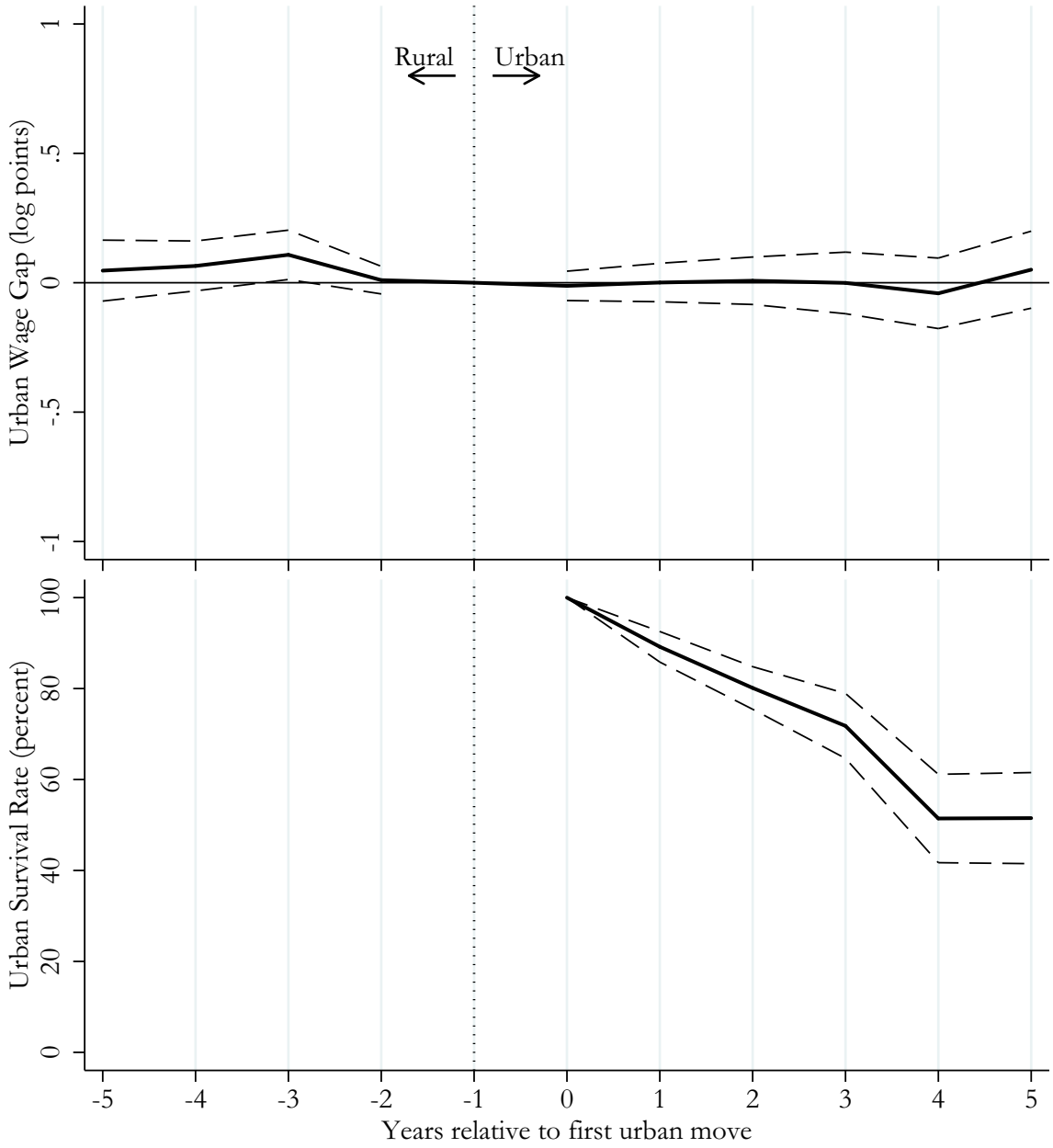
(C) Kenya (Baseline Rural)



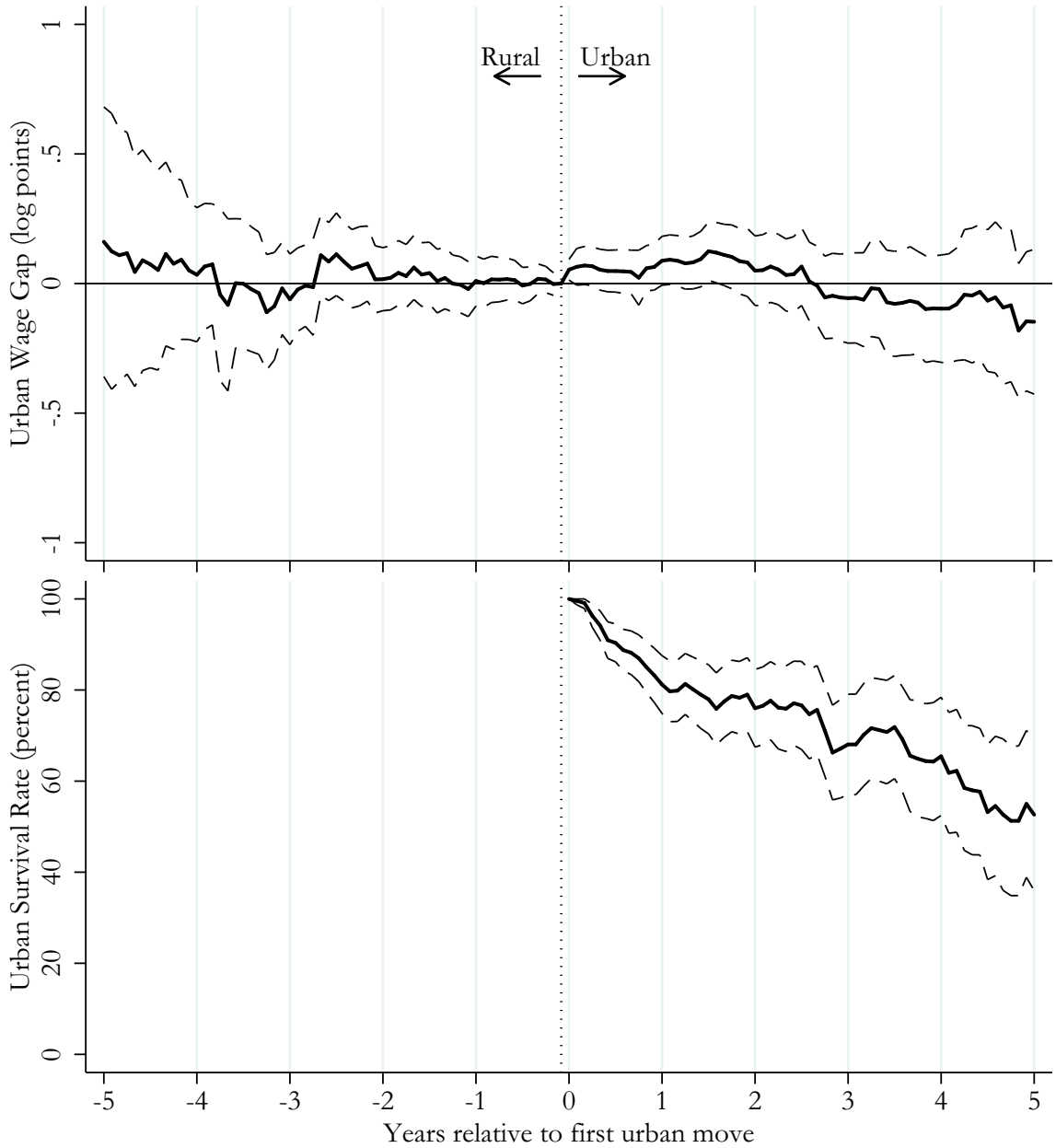
Notes: Productivities are recovered individual-urban status effects from a fixed effects regression of log wages on squared age and indicators for time period on the same sample found in Table 5, panel A, column 6. Productivities are normalized such that the average productivity of rural non-migrants has zero mean. Histograms on the bottom of Panel A represent marginal distributions of rural productivities for "Always Rural" non-migrants (grey) and migrants (hollow). Marginal distribution of estimated urban productivities for migrants reported on the left (hollow). Means and standard deviations reported in log points. Scatterplot presents joint distribution for migrants with best fit line. Bootstrapped standard error of the slope reported in parenthesis from 1,000 iterations of block sampling of individuals with replacement. The correlation coefficient for these productivities is 0.698 (0.035). Panel B presents a histogram of "Always Urban" urban productivities of non-migrants (grey) at the top left, an adjacent histogram of migrant urban productivities (hollow), and migrant rural productivities (grey) below. Joint distribution of urban and rural productivities and corresponding best fit line presented similar to panel A. The correlation coefficient for these productivities in Panel B is 0.700 (0.034). Panel C mimics the format of Panel A except uses data from KLPS. The correlation coefficient for these productivities in Panel C is 0.631 (0.050).

Figure 4: Event Study of Urban Migration

(A) Indonesia



(B) Kenya



Notes: Panel A uses data from individuals in IFLS who are baseline rural, and Panel B uses data from KLPS. Please refer to Section 3 for further details on the data. The top half of each panel reports event study coefficients β_τ from a regression of log wages described in equation 8 (in section IV.D). The solid line represents the point estimate, and the dashed lines represent the 95% confidence interval. Estimates represent the difference in mean wages between movers and non-movers net of the difference that existed in the period prior to the move. Regressions include individual fixed effects, time fixed effects, squared-age, and indicator variables that pool observations exceeding a five year window of the move. The lower half of each panel reports the fraction of people having no rural observations from period zero to the period of interest. (Thus the estimated fraction of survivors can increase in principle due to sample composition changes, as can be seen in the lower half of panel B.) In IFLS, there are 421 individuals who have observed wages in the year of the move and the year prior; 99 of these individuals report wages 5 years later. In KLPS, these numbers are 308 and 38, respectively.

Table 1: Non-Agriculture/Agriculture and Urban/Rural

(A) Indonesia

	Rural	Urban	Total
Agriculture	49.0%	3.9%	39.8%
Non-Agriculture	51.0%	96.1%	60.2%
Number of Observations	89,782	23,132	112,914

(B) Kenya (Including Agricultural Module)

	Rural	Urban	Total
Agriculture	68.2%	15.0%	54.4%
Non-Agriculture	31.8%	85.0%	45.6%
Number of Observations	4,359	1,530	5,889

(C) Kenya (Main Analysis Sample)

	Rural	Urban	Total
Agriculture	24.2%	2.9%	16.2%
Non-Agriculture	75.8%	97.1%	83.8%
Number of Observations	79,567	47,687	127,254

Notes: Panel A presents summary statistics from the Indonesian Family Life Survey (IFLS), and Panels B and C present data from the Kenya Life Panel Survey (KLPS); both are described in more detail in Section 3. Panel A shows the main Indonesian analysis sample of 112,914 individual-year observations, for individuals aged 16 and above for whom earnings measures are available. Panel B uses data contemporaneous to when KLPS rounds 2 and 3 were conducted (excluding retrospective information), and includes detailed employment sector data from the wage and self-employment module. Furthermore, during those rounds, some information was collected on small-scale agricultural home production for those with annual sales less than 40,000 Ksh (approximately 400 USD). No recall data is available on small-scale production in the agricultural module, so Panel B presents data from at most two points of time for each individual, for those aged 16 and above. Panel C excludes the agricultural production model and uses monthly employment data that was collected in the wage and self-employment modules of KLPS rounds 2 and 3. These modules collected detailed information on current as well as previous employment, and include wage employment in the agricultural sector as well as self-employment in agriculture if annual sales were greater than 40,000 Ksh. This results in a monthly panel dataset of 127,254 individual-month observations from 1998 to 2014, which is the main Kenyan analysis sample. Each cell reports the percentage of observations by agricultural and non-agricultural sector, and by rural and urban area. In both IFLS and KLPS, individuals are characterized as working in non-agriculture if they do not have any employment in agriculture during that time period, and zero otherwise. The urban indicator from IFLS is obtained from survey responses to the question: "Is the area you live in a village, a town or a city?" If the person reports living in a town or city, the urban indicator variable equals 1. For KLPS, the urban indicator equals 1 if the person lives in a city with a population size of more than 250,000 and/or if a person lives in a county with a population size greater than 1,000,000 and/or if the person lives in a county with a density greater than 1,000 per square kilometer.

Table 2: Summary Statistics

(A) Indonesia

	All N=18068	Always Rural N=13048	Rural-to-Urban Migrants N=471	Always Urban N=3769	Urban-to-Rural Migrants N=780	Obs
Primary Ed.	0.800	0.748	0.932	0.943	0.904	18068
	[0.400]	[0.434]	[0.252]	[0.231]	[0.295]	
Secondary Ed.	0.284	0.206	0.446	0.511	0.399	18068
	[0.451]	[0.405]	[0.498]	[0.500]	[0.490]	
College	0.072	0.042	0.110	0.159	0.124	18068
	[0.258]	[0.201]	[0.314]	[0.365]	[0.330]	
Female	0.420	0.416	0.380	0.444	0.392	18068
	[0.494]	[0.493]	[0.486]	[0.497]	[0.489]	
Raven's Z-score	0.004	-0.079	0.296	0.264	0.167	11127
	[0.961]	[0.961]	[0.935]	[0.911]	[0.942]	

(B) Kenya

	All N=4439	Always Rural N=1834	Rural-to-Urban Migrants N=2605	Always Urban	Urban-to-Rural Migrants	Obs
Primary Ed.	0.745	0.663	0.803			4439
	[0.436]	[0.473]	[0.398]			
Secondary Ed.	0.363	0.262	0.434			4439
	[0.481]	[0.440]	[0.496]			
College	0.037	0.018	0.051			4439
	[0.190]	[0.133]	[0.220]			
Female	0.520	0.495	0.538			4439
	[0.500]	[0.500]	[0.499]			
Raven's Z-score	0.063	-0.101	0.178			4439
	[0.981]	[0.967]	[0.974]			

Notes: Panel A reports summary statistics from IFLS and panel B reports summary statistics from KLPS. Sample standard deviations reported in brackets below sample means. The sample is limited to respondents who report age, gender, and years of education and have at least one person-time observation that has income, hours, location of residence, and sector of occupation. In panel A (Indonesia), baseline urban status is determined by the first person-year observation where an individual is observed. Migrants in columns three and five are those who have wage, hour, and sector observations in urban and rural areas, respectively. In panel B (Kenya), all individuals are baseline rural—; migrants are those who have subsequent observations with information on income, hours, and sector in the urban sector. Rows correspond to the fraction within each column who have completed primary education, secondary education, and college; the fraction female; and the average score from a Raven's matrices exam, normalized to be mean zero and standard deviation.

Table 3: Correlates of Urban Migration

(A) Indonesia (Baseline Rural)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.0334***					0.0218***	0.0191***
	(0.003)					(0.003)	(0.003)
Secondary Ed.		0.0478***				0.0373***	0.0277***
		(0.005)				(0.006)	(0.006)
College			0.0539***			0.0177	0.0258
			(0.012)			(0.013)	(0.015)
Female				-0.00498		0.000251	0.00828*
				(0.003)		(0.003)	(0.004)
Raven's Z-score					0.0130***		0.00723***
					(0.002)		(0.002)
Constant	0.00964***	0.0246***	0.0324***	0.0369***	0.0340***	0.00950***	0.00639
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	13519	13519	13519	13519	8610	13519	8610

(B) Kenya

	(1)	(2)	(3)	(4)	(5)	(6)
Primary Ed.	0.178***					0.0952***
	(0.017)					(0.020)
Secondary Ed.		0.180***				0.105***
		(0.015)				(0.018)
College			0.223***			0.0990**
			(0.032)			(0.033)
Female				0.0419**		0.0631***
				(0.015)		(0.015)
Raven's Z-score					0.0704***	0.0361***
					(0.007)	(0.008)
Constant	0.454***	0.521***	0.579***	0.565***	0.582***	0.439***
	(0.015)	(0.009)	(0.008)	(0.011)	(0.007)	(0.017)
Observations	4439	4439	4439	4439	4439	4439

Notes: Please see Table 2 for sample restrictions and row variable definitions. Each cell reports a regression coefficient with an indicator for being an urban migrant as the dependent variable. Panel A (Indonesia) is estimated on individuals whose first observation with income, hours, and sector information is rural, whereas panel B includes the full sample subject to previously defined sample restrictions. Columns 6 and 7 report coefficients from a multiple regression with corresponding rows as included covariates. Column 6 for Indonesia omits the Raven's matrix exam to preserve sample size. Robust standard errors reported below in parenthesis, *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Correlates of Employment in Non-Agriculture

(A) Indonesia (Baseline Rural)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.279***					0.240***	0.216***
	(0.010)					(0.010)	(0.014)
Secondary Ed.		0.298***				0.226***	0.193***
		(0.008)				(0.009)	(0.011)
College			0.277***			0.0369**	0.0292
			(0.013)			(0.014)	(0.018)
Female				0.104***		0.151***	0.159***
				(0.008)		(0.008)	(0.010)
Raven's Z-score					0.0655***		0.0311***
					(0.005)		(0.005)
Constant	0.421***	0.567***	0.619***	0.588***	0.663***	0.337***	0.366***
	(0.009)	(0.005)	(0.004)	(0.006)	(0.005)	(0.009)	(0.014)
Observations	13519	13519	13519	13519	8610	13519	8610

(B) Kenya

	(1)	(2)	(3)	(4)	(5)	(6)
Primary Ed.	0.123***					0.0969***
	(0.012)					(0.013)
Secondary Ed.		0.0823***				0.0295***
		(0.007)				(0.008)
College			0.0785***			0.0218**
			(0.007)			(0.008)
Female				0.0626***		0.0744***
				(0.008)		(0.008)
Raven's Z-score					0.0339***	0.0178***
					(0.004)	(0.005)
Constant	0.827***	0.889***	0.916***	0.886***	0.916***	0.795***
	(0.011)	(0.006)	(0.004)	(0.007)	(0.004)	(0.013)
Observations	4439	4439	4439	4439	4439	4439

Notes: Please see Table 2 for sample restrictions and row variable definitions. Each cell reports a regression coefficient with an indicator for being ever being employed in non-agriculture as the dependent variable. Panel A (Indonesia) is estimated on individuals whose first observation with income, hours, and sector information is rural, whereas panel B includes the full sample subject to previously defined sample restrictions. Columns 6 and 7 report coefficients from a multiple regression with corresponding rows as included covariates. Column 6 for Indonesia omits the Raven's matrix exam to preserve sample size. Robust standard errors reported below in parenthesis, *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Non-Agricultural/Agricultural Gap in Earnings

(A) Indonesia

	Dependent variable: Log Earnings (in IDR)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7) Log Wage
Non-agricultural employment	0.535***	0.444***	0.277***	0.252***	0.064**	0.047**	−0.001
	(0.016)	(0.015)	(0.014)	(0.018)	(0.026)	(0.023)	(0.029)
Log hours		0.521***	0.353***	0.388***	0.289***	0.326***	
		(0.026)	(0.023)	(0.027)	(0.060)	(0.034)	
Log hours squared		−0.016***	−0.002	−0.009*	0.010	−0.014**	
		(0.005)	(0.004)	(0.005)	(0.011)	(0.006)	
Female			−0.458***	−0.454***	−0.588***		
			(0.014)	(0.017)	(0.042)		
Age			0.063***	0.064***	0.063***		
			(0.002)	(0.004)	(0.007)		
Age squared			−0.001***	−0.001***	−0.001***	−0.001***	−0.000***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of education			0.031***	0.018***	0.050***		
			(0.005)	(0.007)	(0.012)		
Years of education squared			0.003***	0.003***	0.002***		
			(0.000)	(0.000)	(0.001)		
Normalized Ravens				0.060***			
				(0.009)			
Normalized Ravens squared				0.021***			
				(0.007)			
Time FE	N	Y	Y	Y	Y	Y	Y
Individual FE	N	N	N	N	N	Y	Y
Switchers					Y		
Number of observations	112914	112914	112914	67116	16204	112914	112914
Number of individuals	18068	18068	18068	11127	1998	18068	18068

(B) Kenya

	Dependent variable: Log Earnings (in KSh)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7) Log wage	(8) Log real wage
Non-agricultural employment	0.487***	0.324***	0.331***	0.318***	0.127	0.157	0.066	0.057
	(0.079)	(0.079)	(0.076)	(0.076)	(0.119)	(0.104)	(0.125)	(0.125)
Log hours		0.390**	0.300*	0.291*	−0.235	0.324		
		(0.189)	(0.173)	(0.170)	(0.387)	(0.254)		
Log hours squared		0.004	0.011	0.012	0.093	−0.006		
		(0.032)	(0.029)	(0.029)	(0.064)	(0.040)		
Female			−0.549***	−0.526***	−0.357*			
			(0.052)	(0.052)	(0.214)			
Age			0.154**	0.151*	0.213			
			(0.077)	(0.078)	(0.217)			
Age squared			−0.002	−0.002	−0.004	0.001	−0.000	−0.000
			(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)
Years of education			−0.053	−0.079*	−0.093			
			(0.048)	(0.048)	(0.134)			
Years of education squared			0.008***	0.009***	0.007			
			(0.003)	(0.002)	(0.008)			
Normalized Ravens				0.088***	−0.010			
				(0.030)	(0.082)			
Normalized Ravens squared				−0.046*	−0.181**			
				(0.026)	(0.071)			
Time FE	N	Y	Y	Y	Y	Y	Y	Y
Individual FE	N	N	N	N	N	Y	Y	Y
Switchers					Y			
Number of observations	127254	127254	127254	127254	13113	127254	127254	127254
Number of individuals	4439	4439	4439	4439	284	4439	4439	4439

Notes: Panel A uses data from rounds 1–4 of the Indonesian Family Life Survey (IFLS), described in Section 3. Panel B uses data from rounds 2–3 of the Kenya Life Panel Survey (KLPS), also described in Section 3. The dependent variable in columns 1 to 6 is log earnings, which are the combined earnings from wage and self-employment, reported in Indonesian Rupiah. Earnings from subsistence agriculture are included in IFLS data but not in KLPS data. If an individual has multiple jobs in the same time period, earnings from all employment are included. The dependent variable in column 7 is log wage, which is obtained by dividing log earnings by total hours worked. The dependent variable in column 8 is log wage adjusted for differences in prices between urban and rural areas. The covariate "Non-agricultural employment" is an indicator variable which equals 1 if the person only holds non-agricultural employment and does not work in agriculture during that time period. The variable equals 0 otherwise, so when the individual holds any employment in the agricultural sector. The covariate log hours sums up hours worked in all employment. The sample size in column 4 is smaller in Panel A because the Raven's test was administered only for a subset of the sample. The sample size in column 5 is smaller because it only includes "switchers" who have at least one observation in both the non-agricultural and agricultural sector. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Urban/Rural Gap in Earnings

(A) Indonesia

	Dependent variable: Log Earnings (in IDR)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7) Log Wage
Urban	0.639***	0.544***	0.284***	0.253***	0.075***	0.018	−0.010
	(0.017)	(0.016)	(0.014)	(0.019)	(0.025)	(0.017)	(0.021)
Log hours		0.509***	0.348***	0.387***	0.495***	0.327***	
		(0.025)	(0.023)	(0.027)	(0.135)	(0.034)	
Log hours squared		−0.012**	0.001	−0.007	−0.024	−0.014**	
		(0.005)	(0.004)	(0.005)	(0.022)	(0.006)	
Female			−0.418***	−0.415***	−0.300***		
			(0.014)	(0.017)	(0.042)		
Age			0.061***	0.064***	0.084***		
			(0.002)	(0.004)	(0.007)		
Age squared			−0.001***	−0.001***	−0.001***	−0.001***	−0.000***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of education			0.038***	0.025***	0.009		
			(0.005)	(0.007)	(0.018)		
Years of education squared			0.003***	0.003***	0.004***		
			(0.000)	(0.000)	(0.001)		
Normalized Ravens				0.063***			
				(0.009)			
Normalized Ravens squared				0.019***			
				(0.007)			
Time FE	N	Y	Y	Y	Y	Y	Y
Individual FE	N	N	N	N	N	Y	Y
Switchers					Y		
Number of observations	112914	112914	112914	67116	9158	112914	112914
Number of individuals	18068	18068	18068	11127	1251	18068	18068

(B) Kenya

	Dependent variable: Log Earnings (in KSh)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7) Log wage	(8) Log real wage
Urban	0.695***	0.601***	0.569***	0.554***	0.351***	0.251***	0.172**	0.094
	(0.053)	(0.051)	(0.046)	(0.046)	(0.075)	(0.072)	(0.081)	(0.081)
Log hours		0.386**	0.298*	0.292*	0.427	0.331		
		(0.184)	(0.170)	(0.168)	(0.335)	(0.253)		
Log hours squared		0.001	0.008	0.008	−0.007	−0.008		
		(0.031)	(0.028)	(0.028)	(0.052)	(0.040)		
Female			−0.536***	−0.517***	−0.442***			
			(0.051)	(0.052)	(0.106)			
Age			0.107	0.106	0.006			
			(0.079)	(0.079)	(0.150)			
Age squared			−0.001	−0.001	0.001	0.001	−0.000	−0.000
			(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Years of education			−0.041	−0.061	−0.063			
			(0.046)	(0.046)	(0.076)			
Years of education squared			0.007***	0.008***	0.009**			
			(0.002)	(0.002)	(0.004)			
Normalized Ravens				0.071**	0.004			
				(0.031)	(0.054)			
Normalized Ravens squared				−0.027	−0.041			
				(0.026)	(0.037)			
Time FE	N	Y	Y	Y	Y	Y	Y	Y
Individual FE	N	N	N	N	N	Y	Y	Y
Switchers					Y			
Number of observations	127254	127254	127254	127254	30470	127254	127254	127254
Number of individuals	4439	4439	4439	4439	863	4439	4439	4439

Notes: Panel A uses data from IFLS and Panel B uses data from KLPS. Please refer to Section 3 for further details on the data and to the notes of Table 5 for additional information on the variables. For IFLS, the urban indicator is obtained from survey responses to the question: "Is the area you live in a village, a town or a city?" If the person reports living in a town or city, the urban indicator variable equals 1. For KLPS, the urban indicator equals 1 if the person lives in a city with a population size of more than 250,000 and/or if a person lives in a county with a population size greater than 1,000,000 and/or if the person lives in a county with a density greater than 1,000 per square kilometer. Column 5 only includes switchers, who are defined as individuals with at least one observation in both an urban and rural area. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Gap in Earnings for those Born in Rural and Urban Areas, Indonesia

(A) Indonesian individuals born in rural areas

	Dependent variable: Log Earnings (in IDR)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	0.457***	0.262***	0.051*	0.011				
	(0.018)	(0.016)	(0.026)	(0.032)				
Urban					0.663***	0.331***	0.030	−0.003
					(0.026)	(0.022)	(0.023)	(0.028)
Time FE	N	Y	Y	Y	N	Y	Y	Y
Individual FE	N	N	Y	Y	N	N	Y	Y
Control variables	N	Y	Y	Y	N	Y	Y	Y
Number of observations	81805	81805	81805	81805	81805	81805	81805	81805
Number of individuals	12474	12474	12474	12474	12474	12474	12474	12474

(B) Indonesian individuals born in urban areas

	Dependent variable: Log Earnings (in IDR)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	0.491***	0.244***	0.026	−0.059				
	(0.043)	(0.036)	(0.047)	(0.057)				
Urban					0.429***	0.207***	−0.001	−0.022
					(0.026)	(0.022)	(0.027)	(0.033)
Time FE	N	Y	Y	Y	N	Y	Y	Y
Individual FE	N	N	Y	Y	N	N	Y	Y
Control variables	N	Y	Y	Y	N	Y	Y	Y
Number of observations	31109	31109	31109	31109	31109	31109	31109	31109
Number of individuals	5594	5594	5594	5594	5594	5594	5594	5594

Notes: Panels A and B repeat the analyses of Table 5A and Table 6A for those born in rural and urban areas, respectively. Please refer to Section 3 for further details on the data, and to the notes of Table 5 and 6 for additional information on the variables. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Gap in Consumption, Indonesia

	Dependent variable: Log Consumption (in IDR)					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.515***	0.263***	0.010			
	(0.011)	(0.010)	(0.018)			
Urban				0.539***	0.286***	0.031
				(0.013)	(0.012)	(0.024)
Time FE	N	Y	Y	N	Y	Y
Individual FE	N	N	Y	N	N	Y
Control variables	N	Y	Y	N	Y	Y
Number of observations	37491	37358	37358	37491	37358	37358
Number of individuals	19554	19501	19501	19554	19501	19501

Notes: All regressions use data from IFLS and, unlike previous tables, the sample includes individuals with and without earnings measures. Consumption data are obtained by adding up the value of food and non-food consumption at the household level and dividing this by the number of household members. The data was collected for each of the four waves so each household has four observations at most. Analyses by food and non-food consumption separately can be found in the Appendix, Table A6, and Appendix tables A7 and A8 provide consumption analyses when only using the sample with positive earnings measures. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared because the others are absorbed by the individual fixed effects. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Urban/Rural Gap in Wages for Top 5 Cities

(A) Indonesia

	Dependent variable: Log Wages (in IDR)			
	(1)	(2)	(3)	(4)
Urban	0.429***	0.309***	0.167***	0.010
	(0.017)	(0.023)	(0.019)	(0.019)
Jakarta (population 10 million)		0.257***	0.240***	0.015
		(0.032)	(0.025)	(0.034)
Surabaya (population 2.8 million)		−0.126*	−0.011	0.061
		(0.071)	(0.060)	(0.081)
Bandung (population 2.6 million)		0.383***	0.221***	0.183**
		(0.060)	(0.043)	(0.079)
Medan (population 2.5 million)		0.151**	0.201***	−0.028
		(0.071)	(0.059)	(0.080)
Bekasi (population 2.5 million)		0.454***	0.337***	−0.014
		(0.063)	(0.050)	(0.061)
Time FE	N	N	Y	Y
Individual FE	N	N	N	Y
Control variables	N	N	Y	Y
Number of observations	112914	110971	110971	110971
Number of individuals	18068	17933	17933	17933

(B) Kenya

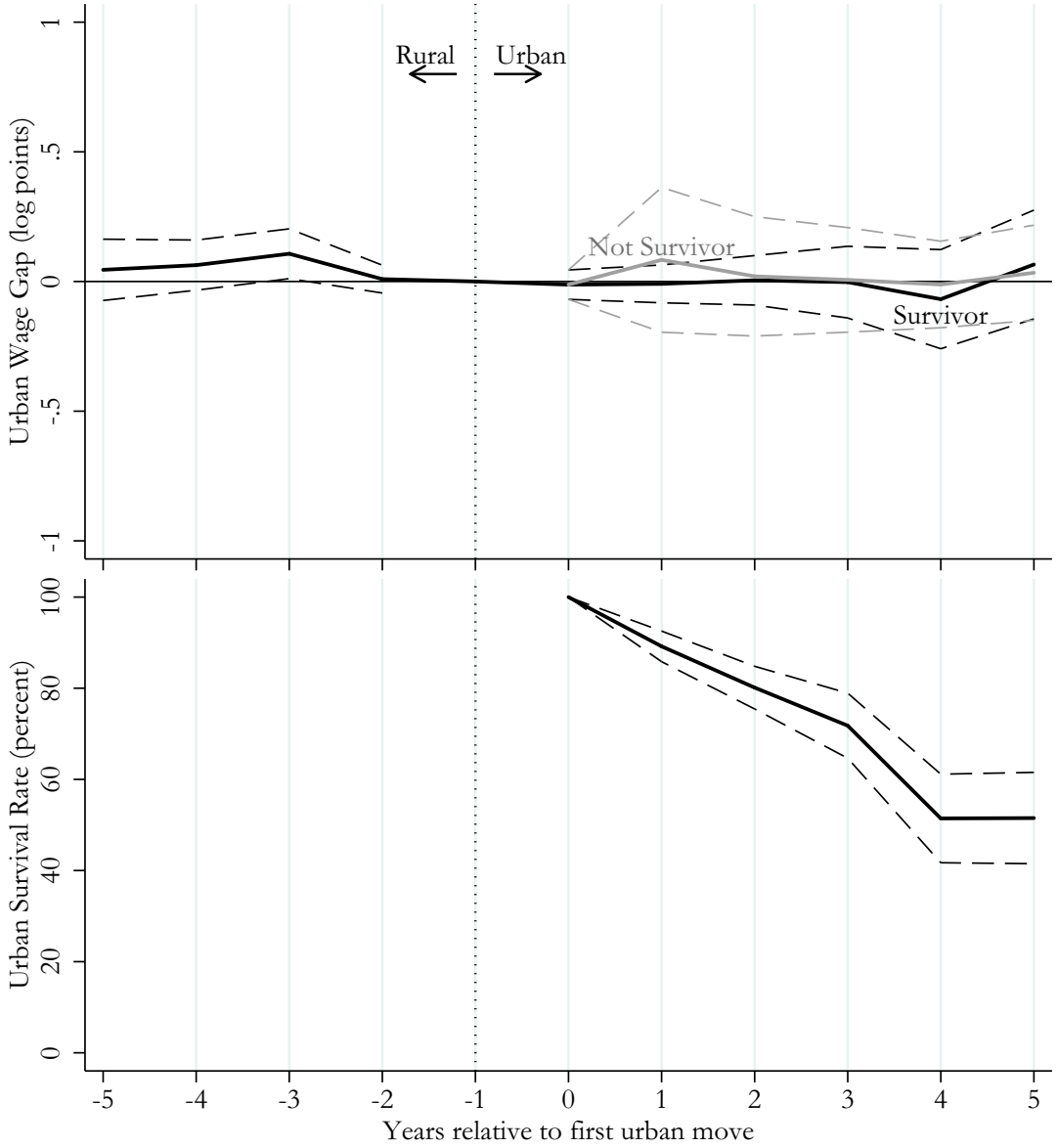
	Dependent variable: Log Wages (in KSh)			
	(1)	(2)	(3)	(4)
Urban	0.537***	0.489***	0.409***	0.195**
	(0.055)	(0.102)	(0.076)	(0.079)
Nairobi (population 3.4 million)		0.101	0.282***	0.031
		(0.104)	(0.075)	(0.072)
Mombasa (population 1.2 million)		0.045	0.134	0.176**
		(0.123)	(0.099)	(0.088)
Kisumu (population 0.4 million)		−0.149	0.061	0.171
		(0.203)	(0.141)	(0.411)
Nakuru (population 0.3 million)		0.062	0.133	−0.038
		(0.141)	(0.102)	(0.150)
Eldoret (population 0.3 million)		0.306	0.250	0.048
		(0.322)	(0.223)	(0.199)
Time FE	N	N	Y	Y
Individual FE	N	N	N	Y
Control variables	N	N	Y	Y
Number of observations	127254	127254	127254	127254
Number of individuals	4439	4439	4439	4439

Notes: Panel A uses data from IFLS and Panel B uses data from KLPS. Please refer to Section 3 for further details on the data and to the notes of Table 5 for additional information on the variables. The covariate "Urban" is an indicator variable that equals 1 if the person lives in an urban area. Similarly, five city indicators are included for the five most populous cities in Indonesia and Kenya, respectively. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 4, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

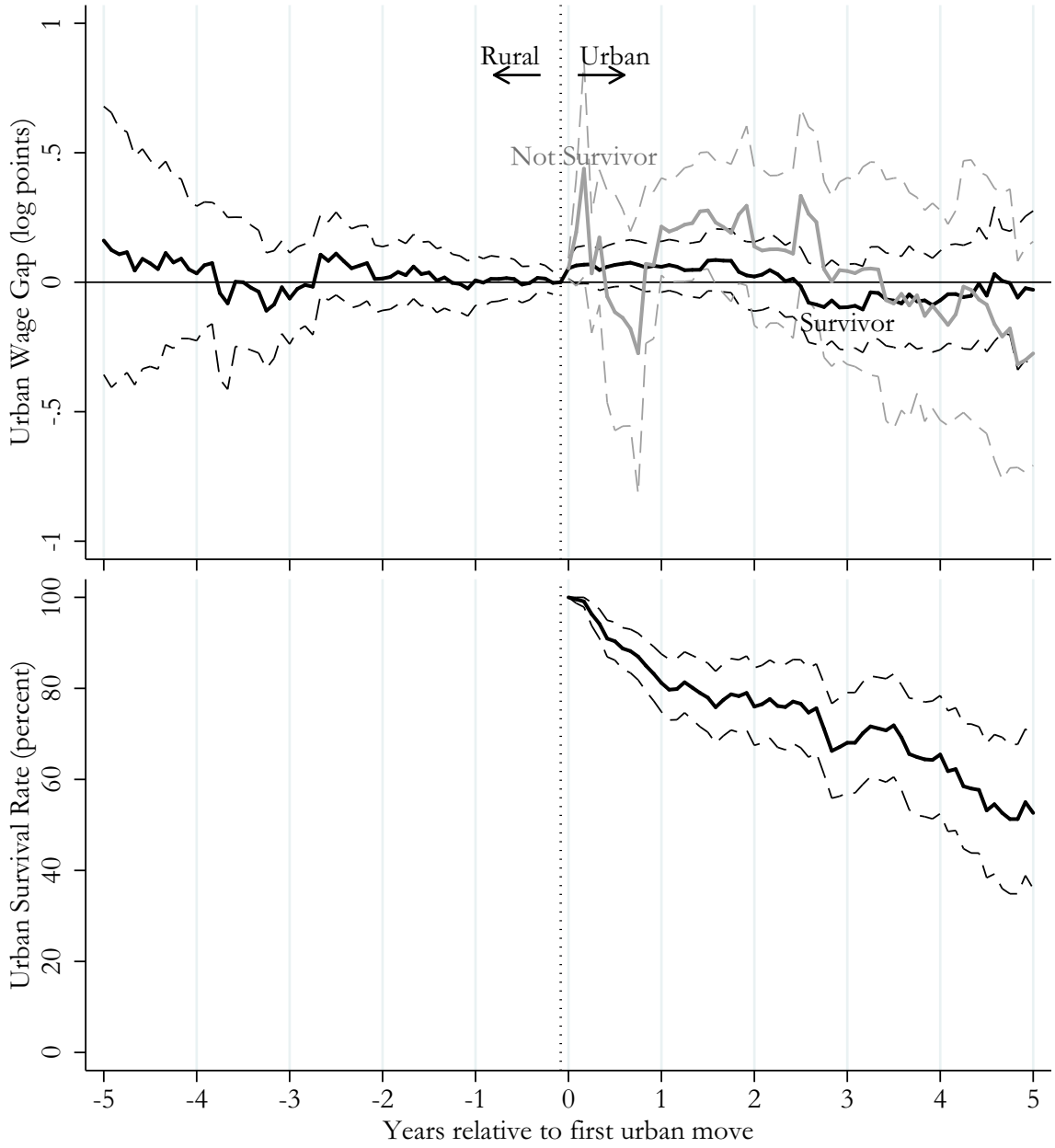
Appendix

Figure A1: Event Study of Urban Migration for Urban Survivors

(A) Indonesia

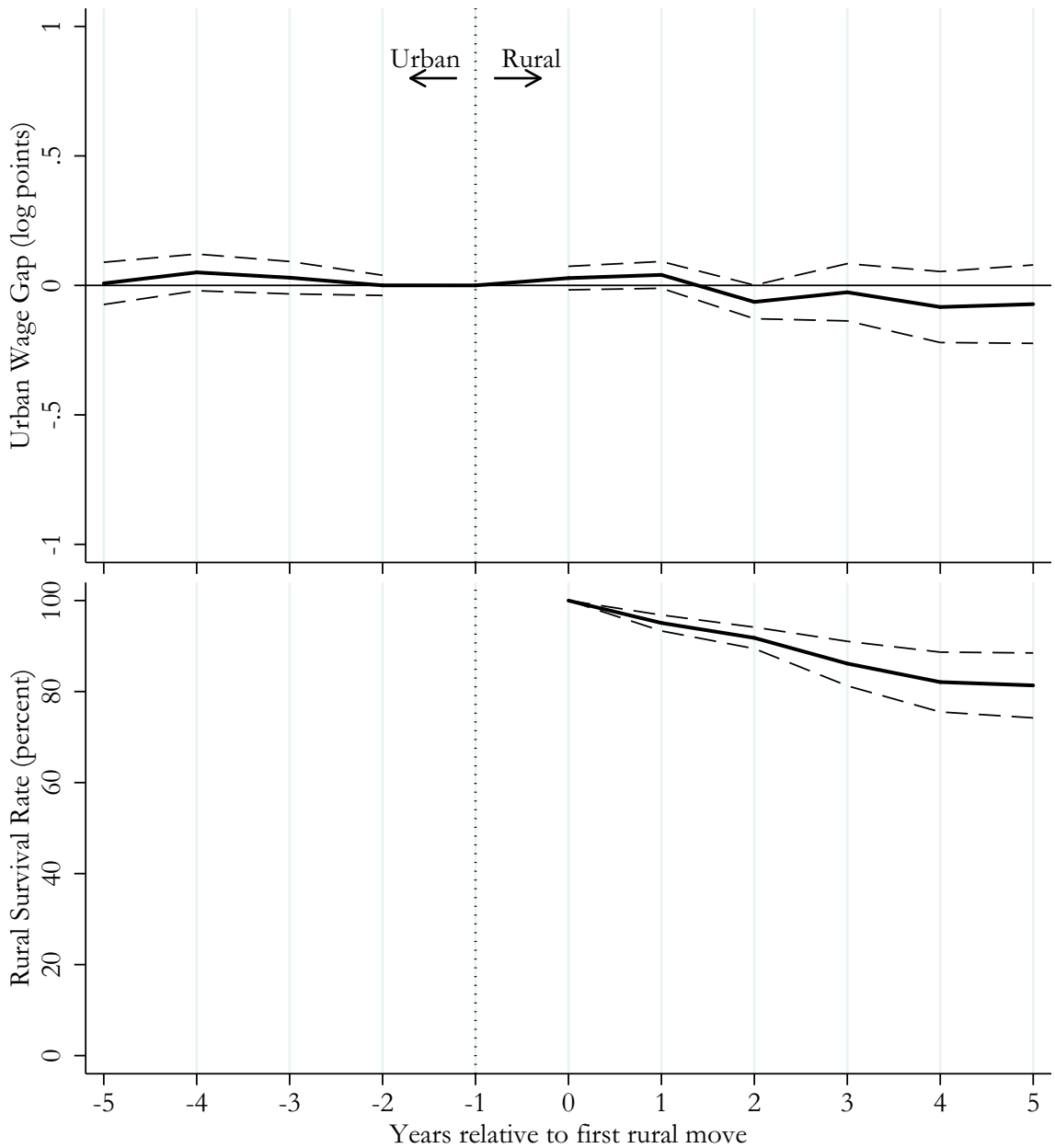


(B) Kenya



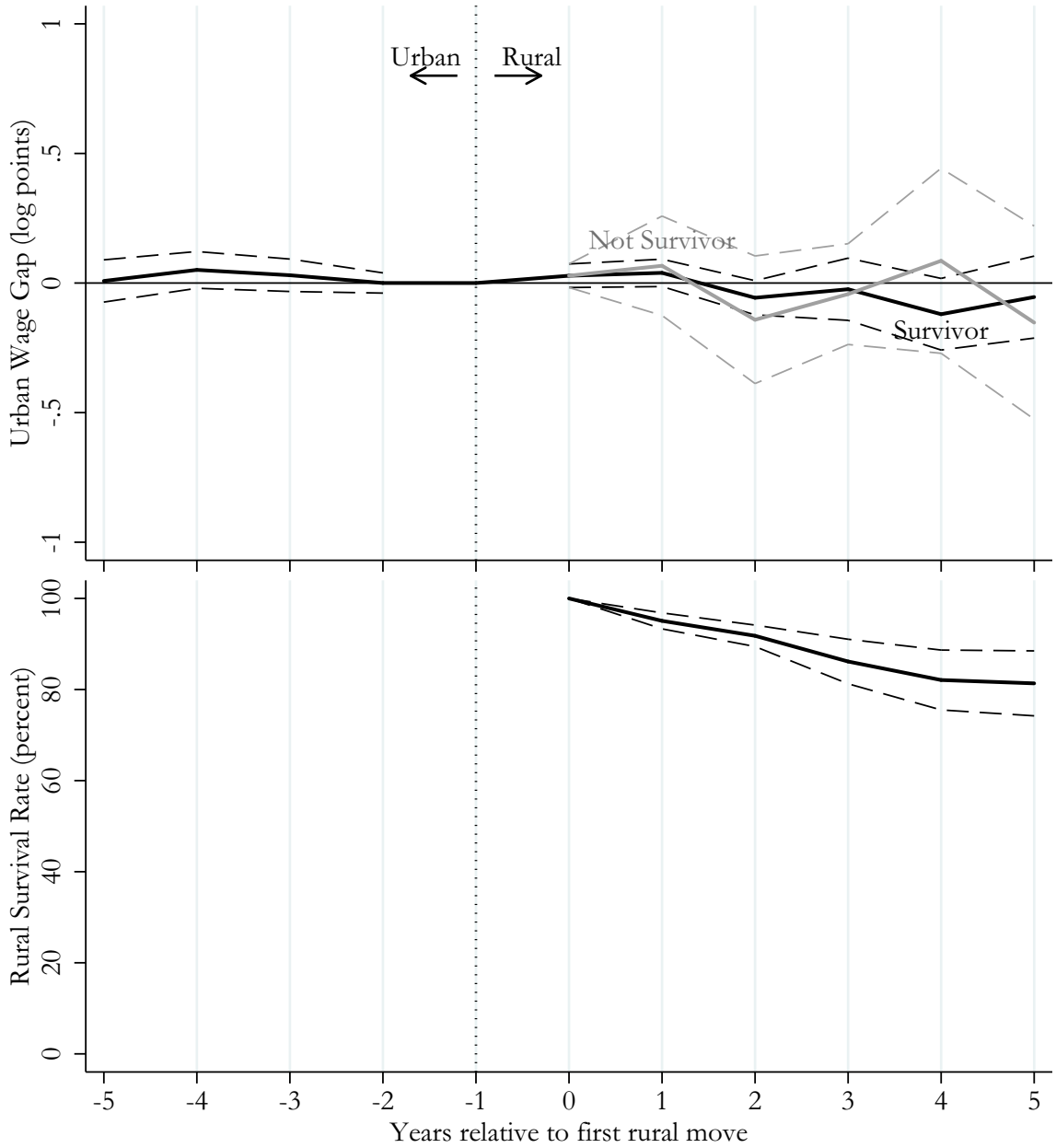
Notes: Event study coefficients reported in top half of figure separately for "survivors" and "not-survivors." "Survivor" status is defined as having no rural observations from period zero (when the individual moved an urban area) to the period of interest, corresponding exactly to the survivor rate graph on the lower half of the figure. Survivor coefficients (black line in the top half) obtained by interacting a survivor indicator with post-event time indicators described in Section IV.D; "not-survivor" coefficients (grey line in the top half) is the event time indicator interacted with one minus the survivor indicator. Panel A reports results for Indonesia, and Panel B reports results for Kenya. Please refer to Figure 4 notes for additional details on included control variables and computation of survivor rates.

Figure A2: Event Study of Rural Migration



Notes: Figure uses data on individuals in IFLS who are baseline urban. Event time indicator variables defined analogously to Figure 4 except with respect to individuals' first rural move. Coefficients multiplied by negative 1 to interpret difference in earnings as an urban premium. Sample includes 710 movers with wage observations at the time of move and one period prior; 118 individuals report wages five years later. Please refer to Figure 4 notes for additional details on included control variables and computation of survivor rates.

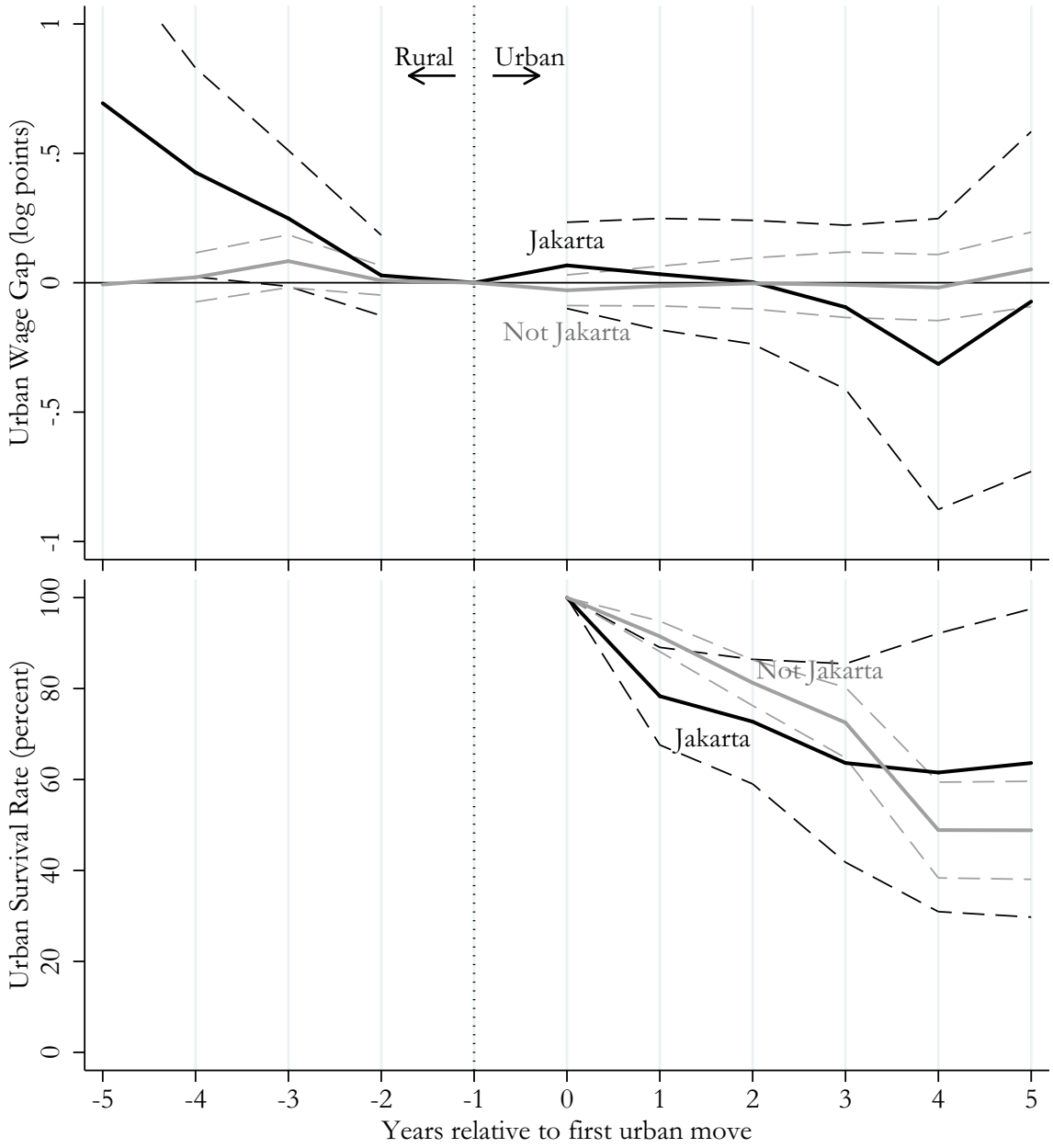
Figure A3: Event Study of Rural Migration for Survivors



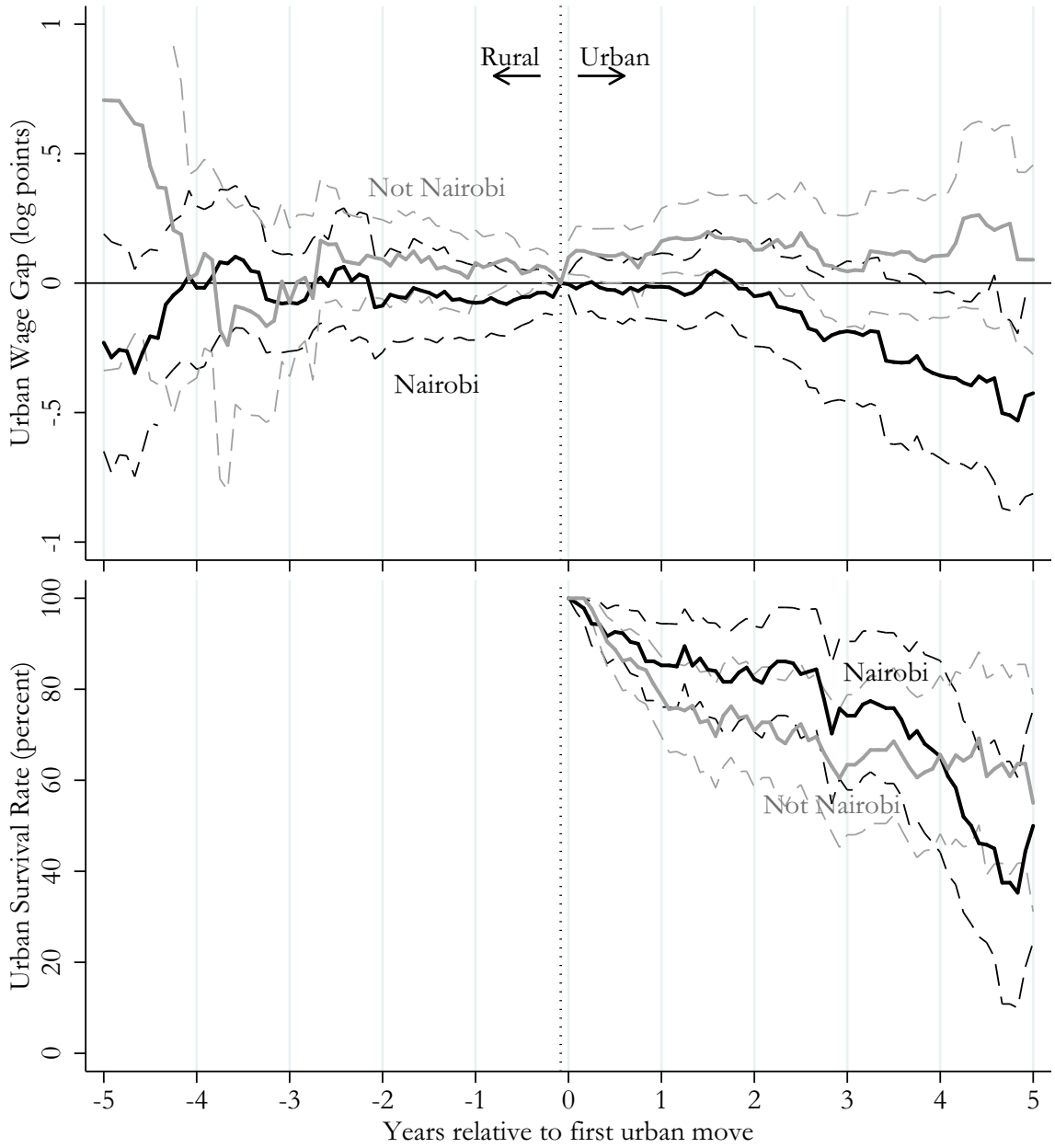
Notes: Figure uses data on individuals in IFLS who are baseline urban. Event study coefficients reported in top half of figure separately for "survivors" and "not-survivors." "Survivor" status is defined as having no urban observations from period zero (when the individual moved a rural area) to the period of interest, corresponding exactly to the survivor rate graph on the lower half of the figure. Survivor coefficients (black line in the top half) obtained by interacting a survivor indicator with post-event time indicators described in Section IV.D; "not-survivor" coefficients (grey line in the top half) is the event time indicator interacted with one minus the survivor indicator. Panel A reports results for Indonesia, and Panel B reports results for Kenya. Please refer to Figure 4 notes for additional details on included control variables and computation of survivor rates.

Figure A4: Event Study of Migration to Capital

(A) Indonesia



(B) Kenya



Notes: Event study coefficients reported in top half of panels A and B separately for moves to the Jakarta and Nairobi, respectively (black line) or other urban areas (grey line). Estimates obtained by interacting a capital city indicator with all event time indicators described in Section IV.D; "Not Jakarta" and "Not Nairobi" is the event time indicator interacted with one minus the respective capital city indicator. In the sample, 75 initial moves were to Jakarta, and 136 initial moves were to Nairobi. Please refer to Figure 4 notes for additional details on included control variables and computation of survivor rates.

Table A1: Correlates of Rural Migration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	-0.0950***					-0.0735**	-0.0798
	(0.026)					(0.027)	(0.049)
Secondary Ed.		-0.0637***				-0.0560***	-0.0779***
		(0.011)				(0.012)	(0.017)
College			-0.0377**			-0.000402	0.0175
			(0.015)			(0.016)	(0.023)
Female				-0.0297**		-0.0361**	-0.0437**
				(0.011)		(0.011)	(0.015)
Raven's Z-score					-0.0167*		-0.00556
					(0.008)		(0.009)
Constant	0.260***	0.203***	0.177***	0.184***	0.180***	0.284***	0.312***
	(0.026)	(0.008)	(0.006)	(0.008)	(0.008)	(0.027)	(0.049)
Observations	4549	4549	4549	4549	2517	4549	2517

Notes: Please see notes from Table 3. Estimation sample is for individuals in Indonesia whose first person-year observation is urban.

Table A2: Correlates of Employment in Non-Agriculture (Indonesia, Baseline Urban)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.0175					0.0194	0.0227
	(0.012)					(0.012)	(0.022)
Secondary Ed.		0.00674				0.00846	0.00811
		(0.004)				(0.005)	(0.007)
College			-0.00488			-0.0103	-0.0230*
			(0.006)			(0.007)	(0.011)
Female				0.0197***		0.0209***	0.0243***
				(0.004)		(0.004)	(0.006)
Raven's Z-score					0.00128		0.00274
					(0.003)		(0.004)
Constant	0.962***	0.975***	0.979***	0.970***	0.976***	0.948***	0.941***
	(0.011)	(0.003)	(0.002)	(0.003)	(0.003)	(0.012)	(0.022)
Observations	4549	4549	4549	4549	2517	4549	2517

Notes: Please see notes from Table 4. Estimation sample is for individuals in Indonesia whose first person-year observation is urban.

Table A3: Unemployment and Search Behavior, Kenya

(A) Unemployment

	Dependent Variable: Unemployment or Subsistence Agriculture			Dependent Variable: Unemployment		
	(1)	(2)	(3)	(4)	(5)	(6)
Urban	−0.020	−0.007	0.032	0.194***	0.189***	0.226***
	(0.013)	(0.014)	(0.030)	(0.011)	(0.011)	(0.021)
Time FE	N	Y	Y	N	Y	Y
Individual FE	N	N	Y	N	N	Y
Control variables	N	Y	Y	N	Y	Y
Mean dependent variable	0.297	0.297	0.297	0.073	0.073	0.073
Number of observations	10504	10504	10504	10527	10527	10527
Number of individuals	6624	6624	6624	6625	6625	6625

(B) Search Behavior

	Dependent variable: Total Hours Job Search		
	(1)	(2)	(3)
Urban	1.461***	1.545***	2.110***
	(0.184)	(0.189)	(0.401)
Time FE	N	Y	Y
Individual FE	N	N	Y
Control variables	N	Y	Y
Number of observations	10502	10502	10502
Number of individuals	6624	6624	6624

Notes: Panel A reports urban gaps in unemployment. The first three columns define an individual as being unemployed if they are searching for work and have no income from wage or salary employment. The second three columns define an individual as being unemployed if they are searching for work and have no income from wage, salary, or proceeds from subsistence agriculture reported in the agricultural model. Sample sizes differ from analysis of wage gaps because questions about job search are contemporaneous to the time of the survey and are not retrospective. The dependent variable in Panel B is the number of hours a person reports to be searching for work; this variable equals 0 if the person is not searching for work. The sample size is smaller than in previous tables because data on search behavior was collected only contemporaneous to when KLPS survey was conducted, and no recall data on this behavior was collected. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Gap in Food and Non-Food Consumption, Indonesia

(A) Food Consumption

	Dependent variable: Log Food Consumption (in IDR)					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.333***	0.157***	−0.001			
	(0.009)	(0.009)	(0.017)			
Urban				0.353***	0.193***	0.036
				(0.012)	(0.011)	(0.023)
Time FE	N	Y	Y	N	Y	Y
Individual FE	N	N	Y	N	N	Y
Control variables	N	Y	Y	N	Y	Y
Number of observations	37471	37338	37338	37471	37338	37338
Number of individuals	19552	19498	19498	19552	19498	19498

(B) Non-Food Consumption

	Dependent variable: Log Non-Food Consumption (in IDR)					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.830***	0.462***	0.029			
	(0.015)	(0.014)	(0.026)			
Urban				0.853***	0.462***	0.037
				(0.017)	(0.016)	(0.033)
Time FE	N	Y	Y	N	Y	Y
Individual FE	N	N	Y	N	N	Y
Control variables	N	Y	Y	N	Y	Y
Number of observations	37479	37346	37346	37479	37346	37346
Number of individuals	19553	19500	19500	19553	19500	19500

Notes: Both panels use data from IFLS. Panels A and B repeat the consumption analyses shown in Table 7, broken down by food and non-food consumption respectively. Please refer to Table 7 for further details. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Gap in Consumption (Main Analysis Sample), Indonesia

	Dependent variable: Log Consumption (in IDR)					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.498***	0.240***	−0.022			
	(0.012)	(0.011)	(0.021)			
Urban				0.507***	0.265***	0.034
				(0.013)	(0.012)	(0.025)
Time FE	N	Y	Y	N	Y	Y
Individual FE	N	N	Y	N	N	Y
Control variables	N	Y	Y	N	Y	Y
Number of observations	31838	31838	31838	31838	31838	31838
Number of individuals	17018	17018	17018	17018	17018	17018

Notes: Both panels use data from IFLS. Panels A and B repeat the consumption analyses shown in Table 7, broken down by food and non-food consumption respectively. Please refer to Table 7 for further details. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Gap in Food and Non-Food Consumption (Main Analysis Sample), Indonesia

(A) Food Consumption

	Dependent variable: Log Food Consumption (in IDR)					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.328***	0.144***	−0.025			
	(0.010)	(0.010)	(0.019)			
Urban				0.334***	0.178***	0.039
				(0.012)	(0.011)	(0.025)
Time FE	N	Y	Y	N	Y	Y
Individual FE	N	N	Y	N	N	Y
Control variables	N	Y	Y	N	Y	Y
Number of observations	31821	31821	31821	31821	31821	31821
Number of individuals	17016	17016	17016	17016	17016	17016

(B) Non-Food Consumption

	Dependent variable: Log Non-Food Consumption (in IDR)					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.796***	0.424***	−0.007			
	(0.016)	(0.015)	(0.029)			
Urban				0.803***	0.432***	0.034
				(0.018)	(0.017)	(0.035)
Time FE	N	Y	Y	N	Y	Y
Individual FE	N	N	Y	N	N	Y
Control variables	N	Y	Y	N	Y	Y
Number of observations	31827	31827	31827	31827	31827	31827
Number of individuals	17017	17017	17017	17017	17017	17017

Notes: All regressions use data from IFLS. This table repeats the analyses shown in Table 7 using the main analyses sample, which excludes individual-year observations without earnings measures. Thus, the sample size is smaller than in Table 7. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Gap in Wage Earnings

(A) Indonesia

	Dependent variable: Log Wage Earnings (in IDR)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	0.800***	0.415***	0.119***	−0.090***				
	(0.021)	(0.019)	(0.026)	(0.035)				
Urban					0.665***	0.301***	0.001	−0.047**
					(0.020)	(0.016)	(0.017)	(0.021)
Time FE	N	Y	Y	Y	N	Y	Y	Y
Individual FE	N	N	Y	Y	N	N	Y	Y
Control variables	N	Y	Y	Y	N	Y	Y	Y
Number of observations	64125	64125	64125	63883	64125	64125	64125	63883
Number of individuals	12316	12316	12316	12299	12316	12316	12316	12299

(B) Kenya

	Dependent variable: Log Wage Earnings (in KSh)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log wage				Log wage
Non-agricultural employment	0.809***	0.565***	0.408***	0.224*				
	(0.081)	(0.074)	(0.099)	(0.135)				
Urban					0.699***	0.582***	0.273***	0.245**
					(0.055)	(0.048)	(0.084)	(0.098)
Time FE	N	Y	Y	Y	N	Y	Y	Y
Individual FE	N	N	Y	Y	N	N	Y	Y
Control variables	N	Y	Y	Y	N	Y	Y	Y
Number of observations	92578	92244	92244	92244	92578	92244	92244	92244
Number of individuals	3849	3838	3838	3838	3849	3838	3838	3838

Notes: Both panels use data from IFLS. This table repeats the analyses shown in Table A6 but using the main analyses sample, which excludes individual-year observations without earnings measures. Thus, the sample size is smaller than in Table A6. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Gap in Self-Employment Earnings

(A) Indonesia

	Dependent variable: Log Self-Employment Earnings (in IDR)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log Wage				Log Wage
Non-agricultural employment	0.391***	0.340***	0.048	−0.115**				
	(0.021)	(0.021)	(0.040)	(0.050)				
Urban					0.620***	0.414***	0.023	0.031
					(0.029)	(0.027)	(0.039)	(0.048)
Time FE	N	Y	Y	Y	N	Y	Y	Y
Individual FE	N	N	Y	Y	N	N	Y	Y
Control variables	N	Y	Y	Y	N	Y	Y	Y
Number of observations	65962	65962	65962	65027	65962	65962	65962	65027
Number of individuals	10275	10275	10275	10177	10275	10275	10275	10177

(B) Kenya

	Dependent variable: Log Self-Employment Earnings (in KSh)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Log wage				Log wage
Non-agricultural employment	−0.400**	−0.269	0.246	0.160				
	(0.182)	(0.174)	(0.239)	(0.236)				
Urban					0.361***	0.234*	0.049	0.026
					(0.137)	(0.128)	(0.110)	(0.141)
Time FE	N	Y	Y	Y	N	Y	Y	Y
Individual FE	N	N	Y	Y	N	N	Y	Y
Control variables	N	Y	Y	Y	N	Y	Y	Y
Number of observations	37567	36860	36860	36860	37567	36860	36860	36860
Number of individuals	1289	1248	1248	1248	1289	1248	1248	1248

Notes: Panel A uses data from IFLS and Panel B uses data from KLPS. The table repeats some of the analyses shown in Tables 5 and 6, but instead of using the sum of earnings from both wage employment and self-employment as the dependent variable, this table only includes earnings from wage employment. For columns 4 and 8, the dependent variable is earnings from wage employment divided by hours worked. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table A9: Non-Agricultural/Agricultural Gap in Earnings using Alternative Definition of Agriculture

(A) Indonesia

	Dependent variable: Log Earnings (in IDR)			
	(1)	(2)	(3)	(4) Log Wage
Any non-agricultural employment	0.527***	0.318***	0.181***	0.159***
	(0.030)	(0.028)	(0.050)	(0.060)
Time FE	N	Y	Y	Y
Individual FE	N	N	Y	Y
Control variables	N	Y	Y	Y
Number of observations	86340	86340	86340	86340
Number of individuals	15337	15337	15337	15337

(B) Kenya

	Dependent variable: Log Earnings (in KSh)			
	(1)	(2)	(3)	(4) Log wage
Any non-agricultural employment	0.543***	0.376***	0.334***	0.142
	(0.080)	(0.076)	(0.112)	(0.149)
Time FE	N	Y	Y	Y
Individual FE	N	N	Y	Y
Control variables	N	Y	Y	Y
Number of observations	127254	127254	127254	127254
Number of individuals	4439	4439	4439	4439

Notes: Panel A uses data from IFLS and Panel B uses data from KLPS. The table repeats some of the analyses shown in Tables 5 and 6, but instead of using the sum of earnings from both wage employment and self-employment as the dependent variable, this table only includes earnings from self-employment. For columns 4 and 8, the dependent variable is earnings from self-employment divided by hours worked. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table A10: Kenya Urban Counties

	Population	Density (Pop per square km)	Central City Size	Fraction of Urban Individual-Months
Nairobi	**3,138,369**	**4,515**	**3,133,518**	45.93
Mombasa	939,370	**4,292**	**938,131**	20.93
Nakuru	**1,603,325**	214	**307,990**	9.08
Kisumu	968,909	465	**409,928**	7.01
Kakamega	**1,660,651**	544	116,358	4.92
Bungoma	**1,375,063**	453	94,927	3.87
Kiambu	**1,673,785**	638	238,858	2.24
Uasin Gishu	894,179	267	**289,380**	2.2
Machakos	**1,098,584**	177	218,557	0.97
Vihiga	554,622	**1,045**	118,696	0.74
Kilifi	**1,109,735**	88	207,253	0.72
Kisii	**1,263,559**	875	83,460	0.58
Migori	917,170	353	**256,086**	0.35
Kitui	**1,012,709**	33	155,896	0.24
Mandera	**1,025,756**	39	87,692	0.12
Meru	**1,591,533**	196	53,627	0.09

Notes: Table presents list of urban counties, where urban is defined as a county that has (a) population greater than 1,000,000, (b) population density greater than 1,000 per square km, and/or (c) central city population greater than 250,000. Numbers bolded when county meets threshold described in a particular column. Column 4 lists fraction of individual months in analysis sample that are urban. Source for populations and densities: 2009 Kenya Census.