

NBER WORKING PAPER SERIES

REEVALUATING AGRICULTURAL PRODUCTIVITY GAPS WITH LONGITUDINAL  
MICRODATA

Joan Hamory Hicks  
Marieke Kleemans  
Nicholas Y. Li  
Edward Miguel

Working Paper 23253  
<http://www.nber.org/papers/w23253>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
March 2017, Revised September 2017

We would like to thank David Albouy, Jorge Alvarez, Lori Beaman, Kevin Donovan, Ben Faber, Fred Finan, Cecile Gaubert, Doug Gollin, Seema Jayachandran, Terence Johnson, Supreet Kaur, David Lagakos, Melanie Morten, Mushfiq Mobarak, Valerie Mueller, Andrés Rodríguez-Clare, Mark Rosenzweig, Todd Schoellman, John Strauss, Chris Udry, and seminar participants at U.C. Berkeley, University of Illinois, Northwestern University, University of Southern California, IMF, the Pacific Development Conference at Stanford, University of Notre Dame, U.C. San Diego, Yale University, MIEDC at University of Wisconsin Madison, Annual World Bank Conference on Africa at Berkeley, CEPR at LSE, the Migration and Development Conference, the Applied and Agricultural Economics Annual Meeting, and SITE for many helpful discussions and suggestions. We are grateful to Brian Feld and Vedika Ahuja for excellent research assistance. All errors remain our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Joan Hamory Hicks, Marieke Kleemans, Nicholas Y. Li, and Edward Miguel. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata  
Joan Hamory Hicks, Marieke Kleemans, Nicholas Y. Li, and Edward Miguel  
NBER Working Paper No. 23253  
March 2017, Revised September 2017  
JEL No. J43,O13,O15,R23

### **ABSTRACT**

Recent research has pointed to large gaps in labor productivity between the agricultural and nonagricultural sectors in low-income countries, as well as between workers in rural and urban areas. Most estimates are based on national accounts or repeated cross-sections of micro-survey data, and as a result typically struggle to account for individual selection between sectors. This paper uses 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 nonagricultural sector (or urban areas), reducing estimated gaps by over 80%. Per capita consumption gaps are also small once individual fixed effects are included. Estimated productivity gaps do not emerge up to five years after a move between sectors. We evaluate whether these findings imply a re-assessment of the conventional wisdom regarding sectoral gaps, discuss how to reconcile them with existing cross-sectional estimates, and consider implications for the desirability of sectoral reallocation of labor.

Joan Hamory Hicks  
CEGA, Center for Effective Global Action  
207 Giannini Hall, Univ of CA, Berkeley  
Berkeley, CA 94720-3310  
jrhamory@berkeley.edu

Nicholas Y. Li  
Evans Hall, #3880  
University of California  
Berkeley, CA 94720-3880  
nicholasli@econ.berkeley.edu

Marieke Kleemans  
Department of Economics  
University of Illinois  
214 David Kinley Hall  
1407 West Gregory Drive  
Urbana, IL 61801  
kleemans@illinois.edu

Edward Miguel  
Department of Economics  
University of California, Berkeley  
530 Evans Hall #3880  
Berkeley, CA 94720  
and NBER  
emiguel@econ.berkeley.edu

An online appendix is available at <http://www.nber.org/data-appendix/w23253>

# 1. Introduction

The shift out of agriculture and into other more “modern” sectors (e.g., manufacturing) has long been viewed as central to economic development. This structural transformation was a focus of influential early scholarship (Rosenstein-Rodan 1943; Lewis 1955; Rostow 1960; Pack 1972; Kuznets 1973; Johnston and Kilby 1978; Schultz 1988) with the issue even stretching back to 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, often using data from national accounts (Gollin, Parente, and Rogerson 2002; Caselli 2005; Restuccia, Yang, and Zhu 2008). This work has documented several patterns that shed light on the sources of cross-country income differences. First, 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 few do in wealthy countries. Importantly, 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 double disadvantage: agricultural work tends to be far less productive in poor countries, yet their workforce is concentrated in this sector.<sup>1</sup> Studies that explore the related gap between urban and rural areas reach similar conclusions.

Several recent studies have examined the extent to which these productivity gaps across sectors can reasonably be viewed as *causal impacts* rather than mainly reflecting *worker selection*. By a causal impact of sector, we 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. In contrast, selection

---

<sup>1</sup> See Appendix Figures A1 and A2, respectively, for an illustration of these differences using cross-country data.

would reflect differences driven by the fact that workers of varying ability and skill are concentrated in certain sectors. This paper disentangles these competing explanations by estimating sectoral wage gaps using unusually long-run individual panel data from two low-income countries, Indonesia and Kenya.

If there are causal impacts of sector, the large share of the workforce employed in the 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 effects from selection, is not solely of scholarly interest: the existence of causal sectoral productivity gaps would imply that the movement of population out of rural agricultural jobs and into other sectors could durably raise living standards in low-income countries, narrowing cross-country differences. The existence of large causal sectoral productivity gaps would also raise questions about the nature of the frictions that limit individual movement into more productive employment, and the public policies that might affect such moves (e.g., see Stren, Halfani, and Malombe 1994 for Tanzania and Au and Henderson 2006 for China).

Gollin, Lagakos, and Waugh (2014, henceforth GLW) and Young (2013) explore this identification issue. 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. In their main contribution, GLW show that accounting for differences in hours worked and average worker schooling attainment across sectors—thus partially addressing worker selection—reduces the average agricultural productivity gap by a third, from roughly 3 to 2. They also find that agricultural productivity and per capita consumption gaps based on household data tend to be somewhat smaller than those estimated using national labor surveys, possibly due to differences in how each source measures economic activity.

GLW remain agnostic regarding the causal interpretation of the large agricultural productivity gaps that they estimate. If individual schooling captures the most important dimensions of worker skill and thus largely addresses selection, 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 to roughly double productivity, a large effect. Of course, to the extent that educational attainment alone fails to capture all aspects of individual human capital, controlling for it would not fully account for selection.

Young (2013) examines the related question of urban-rural differences in consumption (as proxied by measures of household asset ownership, education, and child health), rather than productivity, and similarly finds large cross-sectional gaps.<sup>2</sup> Young's interpretation differs from GLW in emphasizing the role of selective migration across sectors. Using Demographic and Health Surveys that have retrospective information on individual birth district, Young shows that rural-born individuals with more years of 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, shows that this could generate two-way flows of the kind he documents, and argues that he can fully explain urban-rural consumption gaps once he accounts for sorting by education.<sup>3</sup>

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. Use of this data allows us to account for individual fixed effects, capturing all time-

---

<sup>2</sup> While Young (2013) focuses on urban-rural gaps, he sometimes uses data on non-agricultural vs. agricultural differences when urban-rural data is missing; GLW similarly use urban-rural data when they lack data on agriculture.

<sup>3</sup> Porzio (2016) argues that a model of worker sorting can explain a large share (roughly 40%) 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.

invariant dimensions of worker heterogeneity, not just educational attainment (as GLW do). We focus on two country cases – Indonesia and Kenya – that have long-term panel micro data sets with relatively large sample sizes, rich measures of earnings in both the formal and informal sectors, and high rates of respondent tracking over time. The datasets, the Indonesia Family Life Survey and Kenya Life Panel Survey (henceforth “IFLS” and “KLPS”), are described in greater detail below.<sup>4</sup>

For both countries, we start by characterizing the nature of selective migration between 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 are 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 (a Raven’s Progressive Matrices score), to show that those of higher ability in both Indonesia and Kenya are far more likely to move into urban and non-agricultural sectors, even conditional on educational attainment. This is a strong indication that conditioning on completed schooling is insufficient to fully capture differences in average worker skill levels across sectors.

We next estimate sectoral productivity differences, and show that treating the data as a repeated cross-section generates large estimated sectoral productivity gaps, echoing the results in existing work. In our main finding, we show that the inclusion of individual fixed effects reduces estimated sectoral

---

<sup>4</sup> There are other panel data sets where similar approaches could be employed, for instance, the Mexican Family Life Survey; we leave this for future work. It is worth noting that Mexico is a member of the OECD and is considerably richer in per capita terms than Indonesia or Kenya. In related work, Alvarez (2015) finds substantial narrowing of productivity gaps in Brazil with the inclusion of individual fixed effects, albeit only using formal sector wage data, and Herrendorf and Schoellman (2016) employ cross-sectional microdata to assess sectoral differences in human capital.

productivity gaps by over 80%. This pattern is consistent with the bulk of the measured productivity gaps between sectors being driven by worker selection rather than causal impacts.

Specifically, we first reproduce the differences documented by GLW for Indonesia and Kenya, presenting both the unconditional gaps as well as adjusted gaps that account for worker labor hours and education (see Figure 1). These are large for both countries, with raw gaps of around 130 log points, implying roughly a doubling of productivity in the non-agricultural sector. When we treat our data as a series of repeated cross-sections, the gaps remain large, at 70 to 80 log points. These are somewhat smaller than GLW's main estimates, though recall that GLW's estimates using household survey data (like ours) also tend to be smaller. Conditioning on individual demographic characteristics (age and gender) as well as hours worked and educational attainment narrows the gap, but it remains large at 35 to 55 log points. Finally, including individual fixed effects reduces the agricultural productivity gap in wages to 7.8 log points in Indonesia and to 6.1 log points in Kenya. Analogous estimates show that urban productivity gaps are also reduced substantially, to zero in Indonesia and 16.5 log points in Kenya. The estimated gaps in GLW are an order of magnitude larger than our estimates.

We obtain similar results for the gaps in real per capita consumption levels across sectors where this is available for Indonesia, at 3 to 13 log points. This is useful since consumption measures may better capture living standards or total income in less developed economies than earnings or wage measures, given widespread informal economic activity as well as unemployment. Furthermore, we show that the productivity gap is not simply a short-run effect by demonstrating that gaps do not emerge even up to five years after an individual moves to an urban area.

Our methodological approach is related to Hendricks and Schoellman (2017), who use panel data on the earnings of international migrants to the United States, including on their home country

earnings. Mirroring our main results, the inclusion of individual fixed effects in their case greatly reduces the return to international migration (by roughly 60%). Similarly, McKenzie et al. (2010) show that cross-sectional estimates of the returns to international immigration (to New Zealand) 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%) in the sending households of individuals randomly induced to migrate to cities within Bangladesh. Using data from Tanzania and observing individuals at two points in time, Beegle et al. (2011) estimate consumption gains of 36% among those who moved away from their origin area.<sup>5</sup> We improve on the latter study by observing the same individuals at many points in time, allowing us to include time fixed effects to absorb covariate shocks, and by using richer individual-level information on productivity, sector of employment, and cognitive ability. 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. While Bryan and Morten (2017) suggest there are substantial gains to removing migration frictions in Indonesia, gains are associated with moves across region rather than urban residence per se. Related studies on selective migration include Chiquiar and Hanson (2005), Yang (2006), Kleemans (2016), and Rubalcava et al. (2008).

A limitation of the current study is its focus on two countries. This is due to the relative scarcity of long-run individual panel data sets in low-income countries that contain the rich measures necessary for our analysis. That said, the finding of similar patterns in two countries with large populations (250 million in Indonesia, 45 million in Kenya) in different world regions suggests some generalizability.

---

<sup>5</sup> Estimating urban-rural wage gaps is challenging in their setting since only 138 individuals are observed in urban areas.



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 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 we discuss 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, selection will most likely produce an upward bias, leading our estimates to be upper bounds on the true causal impact of moving between sectors. However, absent additional knowledge about the correlation between individual preferences, credit constraints, and unobserved productivity shocks, it is in principle possible that selection could bias our estimates downward instead. Similarly, it is possible 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 return to these important issues of interpretation in the conclusion, including ways to reconcile our estimates with existing empirical findings.

The paper is organized as follows. Section 2 presents a conceptual framework for estimating sectoral productivity gaps, and relates it to the core econometric issue of disentangling causal impacts from worker selection. Section 3 describes the two datasets (IFLS and KLPS); characterizes the distinctions between the non-agricultural and agricultural sectors, and urban vs. rural areas; and presents evidence on individual selection between sectors. Section 4 contains the main empirical results on productivity gaps, as well as the dispersion of labor productivity across individuals by sector, consumption gaps, dynamic effects up to five years after migration, and effects in big cities. The final section presents alternative interpretations of the results, and concludes.

## 2. Theoretical Framework

We present a development accounting framework to disentangle explanations for the aggregate productivity gap across sectors. We consider both observable and unobservable components of human capital, and whether intrinsic worker preferences for sector may bias direct measurement of the productivity gap. A standard model suggests that worker selection is most likely to bias sectoral productivity gaps upward when estimated among those moving into non-agriculture (urban areas) but lead to a downward bias when estimated among those moving into agriculture (rural areas).

### 2A. The Agricultural Productivity Gap through the Lens of an Aggregate Production Function

Following Hendricks and Schoellman (2017), we denote production in sector  $s$  as  $Q_s = K_s^\alpha (A_s H_s L_s)^{1-\alpha}$ . Dropping subscripts for convenience, a representative firm in sector  $s$  solves:

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

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  $\tau_K$  and  $\tau_H$  represent wedges that prevent factors from receiving their marginal product.

Solving the first order condition with respect to the quantity of the labor aggregate yields:

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

An individual's income in sector  $s$  is given by  $Y_{is} = Z_s H_{is} L_{is}$ . Denoting logs in lower case, the average log-income gap across the non-agricultural ( $n$ ) and agricultural ( $a$ ) sectors is:

$$\bar{y}_n - \bar{y}_a = \underbrace{(z_n - z_a)}_{\text{residual income gap}=\beta} + \underbrace{(\bar{l}_n - \bar{l}_a)}_{\text{labor supply gap}} + \underbrace{(\bar{h}_n - \bar{h}_a)}_{\text{human capital gap}} \quad (1)$$

The agricultural productivity gap is comprised of a labor supply gap, a human capital gap, and a productivity residual,  $\beta$ , the key parameter of interest. The residual gap  $\beta$  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 of capital. These wedges are the focus of some models of structural transformation (Restuccia, Yang, and Zhu 2008; Graham and Temple 2006) and may serve as a summary parameter for an economy's degree of underdevelopment. In what follows, we do not take a stand on the specific components contributing to any sectoral productivity gap.

We assume that individual human capital takes the Mincerian form,  $H_i = \exp[\mathbf{x}'_i \mathbf{b} + \eta_i]$  where  $\mathbf{x}_i$  is a vector of observed characteristics (e.g., years of schooling) with corresponding returns  $\mathbf{b}$ , and  $\eta_i$  represents unobserved skill. Substituting into the wage equation, log wages in sector  $s$  are:

$$y_i = z_a + \beta 1[s = n]_i + l_i + \mathbf{x}'_i \mathbf{b} + \eta_i \quad (2)$$

The agricultural productivity gap becomes:

$$\bar{y}_n - \bar{y}_a = \beta + (\bar{l}_n - \bar{l}_a) + (\bar{\mathbf{x}}_n - \bar{\mathbf{x}}_a)' \mathbf{b} + (\bar{\eta}_n - \bar{\eta}_a) \quad (3)$$

It is evident that any differences in unobserved components of worker human capital will be absorbed into the residual wage gap here, and an OLS estimate of  $\beta$  will be biased.<sup>6</sup>

---

<sup>6</sup> This model can be generalized to allow for sector specific human capital with  $h_{ik} = \exp[\mathbf{x}'_i \mathbf{b}_s + \eta_i]$  yielding an urban-rural gap described by  $\bar{y}_n - \bar{y}_a = \beta + (\bar{l}_n - \bar{l}_a) + (\bar{\mathbf{x}}_n - \bar{\mathbf{x}}_a)' \mathbf{b}_a + (\bar{\mathbf{x}}_a)' (\mathbf{b}_n - \mathbf{b}_a) + (\bar{\eta}_n - \bar{\eta}_a)$  which motivates a Oaxaca-style decomposition where  $\mathbf{b}_n - \mathbf{b}_a$  represent different returns paid to observable characteristics in non-agriculture. The main specifications below focus on human capital differences such as those described in equation 2 rather than this more flexible formulation of human capital. We also assume that the agricultural productivity gap does not change over

There are two immediate approaches for obtaining better estimates of  $\beta$ . First, one can obtain a richer set of observable characteristics  $\mathbf{x}_i$ , reducing the scope for unobserved (to the econometrician) ability to determine 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 individual characteristics. While our estimation explores both avenues, our preferred estimates use the second approach, using fixed effects panel data estimation.

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

$$y_{it} = z_a + \beta 1[s = n]_{it} + l_{it} + \theta_i + \omega_{it} \quad (4)$$

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

$$\overline{y_{nt}} - \overline{y_{at}} = \beta + (\overline{l_{nt}} - \overline{l_{at}}) + (\overline{\theta_n} - \overline{\theta_a}) + (\overline{\omega_{nt}} - \overline{\omega_{at}}) \quad (5)$$

Here, the time-varying, sector-specific components of human capital  $\omega_{ist}$ , are potential sources of omitted variable bias. Equation 4 is the key estimation equation; we explore potential limitations and pitfalls to this approach in what follows.

Estimating the agricultural productivity gap via equation 4 allows us to tentatively explore Lagakos and Waugh's (2013) hypothesis that comparative and absolute advantage in a Roy (1951)

---

time, which precludes time-varying frictions or production functions. This contrasts with longer-term perspectives (e.g., Herrendorf, Rogerson, and Valentinyi 2014), but is sensible given the decadal time scale of our analysis.

model of self-selection can explain sectoral productivity gaps within countries. To do so, we also allow for a richer formulation of the dynamic Mincerian human capital equation:  $H_{ist} = \exp[\theta_{is} + \omega_{ist}]$ , where  $\theta_{is} = \mathbf{x}'_i \mathbf{b}_s + \eta_{is}$  allows for different returns to observable human capital and unobserved ability by sector. Correspondingly, using a panel dataset containing multiple individual observations over time, we compute individual time-invariant human capital in each sector ( $\theta_{is}$ ), and examine the joint distribution of these productivities, to explore whether those who have an absolute advantage in both sectors also tend to have a comparative advantage in the non-agricultural sector.

## 2B. Estimation Issues Related to Worker Selection into Sector

Departing from the general equilibrium model specified above, consider an agent  $i$  facing a choice of working in agriculture or non-agriculture. The utility  $v$  obtained by working in sector  $s$  is given by:

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

where  $\xi_{ist}$  is an independent idiosyncratic preference shock for sector  $s$  in time  $t$ . For now, assume these preference shocks are uncorrelated with individual sectoral wage innovations,  $\omega_{ist}$ . We further assume that the non-stochastic component of utility is linearly separable, as  $f(y_{it}, \mathbf{x}_i) = y_{it} + \mathbf{x}'_i \Gamma_s$ .

Substituting in equation 4 for the individual specific productivity term, an individual chooses to work in the non-agriculture sector  $s = n$  if and only if  $v_{int} - v_{iat} > 0$ ; the probability of this occurring (abstracting away from labor supply differences across sectors for parsimony) is given by:

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

The possible bias here is classic simultaneity: wage innovations  $\omega_{ist}$  simultaneously determine the worker's sectoral choice and her wage. In other words, receiving a positive productivity shock in non-

agriculture  $\omega_{int}$  is both positively correlated with the indicator for non-agricultural work and positively correlated with wages in that sector, while a positive productivity shock in agriculture  $\omega_{iat}$  is negatively correlated with an indicator for non-agriculture and positively correlated with wages.<sup>7</sup>

The key threats to econometric identification in this panel data setting are time-varying shocks. The requirements for a convincing instrumental variable to remove selection bias in this context are thus relatively stringent. Such an instrument would ideally affect preferences for migration but be excludable from wage determination; this rules out using local rainfall shocks as an instrument, for instance. Both the IFLS and KLPS provide stated reasons for migration (subsequent to the move), but 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. The dearth of credible natural experiments in migration makes the experimental variation 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 equation 5 (again abstracting away from labor supply differences) is:

$$\overline{y_{nt}} - \overline{y_{at}} = \beta + (\mathbf{E}\{\theta_{in} + \omega_{int} | v_{int} > v_{iat}\} - \mathbf{E}\{\theta_{ia} + \omega_{iat} | v_{iat} > v_{int}\}) \quad (7)$$

The selection expression in equation 6 suggests that we would only observe those employed in non-agriculture who would benefit from it, i.e.,  $\beta + \theta_{in} + \omega_{int} - \theta_{ia} - \omega_{iat} > c_{it}$ , where  $c_{it}$  is the utility cost of moving for individual  $i$  in period  $t$  (captured by the other terms in equation 6). While this may be the average causal effect for this population—analogue to a local average treatment-on-the-treated in the program evaluation literature—extrapolating this effect to the non-movers may be problematic.

---

<sup>7</sup> Explicitly, estimates of the agricultural productivity gap are biased if  $\mathbf{E}\{\omega_{int} | v_{int} > v_{iat}\} \neq \mathbf{E}\{\omega_{iat} | v_{iat} > v_{int}\}$ .

This is especially relevant in fixed effects estimation, which estimates the productivity gap  $\beta$  using the wages of the movers, namely, those with productivity observations in both sectors.

Note that it is also possible that one might observe positive migration flows into non-agricultural employment (or urban areas) even in the case where the true average productivity gap,  $\beta$ , was negative; in such a case, movers would consist of those with particularly large and positive individual returns to urban relative to rural employment in that time period, or perhaps those who face sufficiently large idiosyncratic preferences for the move, say, those with  $c_{it}$  negative.<sup>8</sup>

By this logic, fixed effects estimates (among the movers into urban areas) will be generally larger than the average population treatment effect. This suggests that estimated gaps based on those who were initially in the rural sector are likely to be *upper bounds* on the magnitude of the true average productivity gap in the population as a whole. Hendricks and Schoellman (2017) make the related point that their estimates of the returns to international migration are likely to be upper bounds.

In this study, this will likely be the case with the Kenya data (KLPS) where the entire sample lived in rural areas at baseline. In the Indonesia data (IFLS), which features sorting in both directions (since sample individuals were born in both urban and rural areas), it is conceptually possible to observe a non-agricultural (or urban) premium every time an individual selects into non-agriculture (urban areas) and an agricultural (rural) premium every time an individual selects into agriculture (rural areas). 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 moves among

---

<sup>8</sup> It is in theory possible to observe little or no migration when the true average productivity gap is positive. To generate this, the utility cost  $\xi_{int} - \xi_{iat}$  must be both negatively correlated with the difference in time-varying sector-specific wage innovations  $\omega_{int} - \omega_{iat}$  and sufficiently large to dominate the  $\beta$  term. Note that such a correlation does not include highly productive individuals who simply prefer employment in agriculture, characteristics which are captured in the fixed effects.

those that benefit from working in agriculture (rural areas), i.e.,  $-\beta + \theta_{ia} - \theta_{in} + \omega_{iat} - \omega_{int} > -c_i$ . The resulting estimates would then serve as *lower bounds* on the true average productivity gain.

The IFLS provides an unusually rich testbed to understand the role of these biases, especially in terms of estimating the urban-rural gap. In the spirit of Young's (2013) observation that migrants flow in both directions, the data allow us to condition on individual birth location and measure the dynamic impacts on wages after migration. The bounding 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. We take this prediction to the data and find evidence for it below. This model of selection implies that the true sectoral productivity gap in Indonesia is bounded by these two estimates, generated by movers in each direction.

### 3. Data

This paper uses data from Indonesia (IFLS) and Kenya (KLPS). At 250 million, the Southeast Asian country of Indonesia is the world's fourth most populous, and Kenya is among the most populous Sub-Saharan African countries with 45 million inhabitants. They are fairly typical of other low income countries with respect to their labor shares in agriculture, estimated agricultural productivity gaps using national accounts data, and the relationships between these variables and national income levels.<sup>9</sup>

The high tracking rates of the datasets we employ allow us to construct multiyear panels of individuals' location decisions. Moreover, both datasets include information on both formal and informal sector employment. The latter is difficult to capture in standard administrative data sources yet often accounts for a large share of the labor force in low-income countries.

---

<sup>9</sup> See Appendix Figures A1 and A2 based on data from GLW. In both figures, the values for both Kenya (KEN) and Indonesia (IDN) are close to the best fitting regression line.



### 3A. Indonesia

Data were collected in five rounds of the Indonesia Family Life Survey between 1993 and 2015 (Strauss et al., 2016). The survey is representative of 83% of the country's population who lived in 13 of the 27 provinces that existed in 1993. While the original sample consisted of 22,347 individuals, efforts to track them even when they had moved outside of the original study area, as well as the inclusion of members from split-off households during subsequent rounds (1997-98, 2000, 2007-08, and 2014-2015), ultimately results in a sample of 58,337 individuals. Attrition is often high in panel data; however, with an intensive focus on respondent tracking, the IFLS is unusually well-suited to study migration. In particular, re-contact rates between any two rounds are above 90%, and 87% of the original households were contacted in all five rounds (Strauss et al. 2016).<sup>10</sup>

Detailed employment data were collected during each survey round. In addition to current employment, the survey included questions on previous employment, allowing us to create up to a 28-year annual individual employment panel from 1988 to 2015. Employment status and sector of employment are available for each year, but in the fourth and fifth IFLS round, earnings were collected only for the current job. The 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, work on a family-owned farm or non-farm business, and unpaid family work. In addition to wages and profits, individuals are asked to estimate the value of

---

<sup>10</sup> Thomas et al. (2012) contains a detailed discussion of tracking and attrition in the IFLS.

compensation in terms of share of harvest, meals provided, transportation allowance, housing and medical benefits, and credit; our main earnings measure is the sum of all wages, profits, and benefits.<sup>11</sup>

Individuals are asked to describe the sector of employment for each job. The single largest sector is “agriculture, forestry, fishing, and hunting”: 31% of individuals report it as their primary employment sector, and 50% have secondary jobs in this sector. Agricultural employment is primarily rural: 43% versus 9% of rural and urban individuals, respectively, report working primarily in agriculture (Table 1, Panel A). Other common sectors are wholesale, retail, restaurants, and hotels (22% of main employment); social services (22%); manufacturing (14%); and construction (5%). These non-agricultural sectors are all more common in urban areas. Men are more likely than women to work in agriculture (35 vs. 23%) and less likely to work in wholesale, retail, restaurants, and hotels, in and social services. Smaller male-dominated sectors include construction (7% of male employment vs. 0.7% for females) and transportation, storage, and communications (6% vs. 0.4%).

In the analysis, we employ an indicator variable for non-agricultural employment, which equals 1 if a respondent’s main employment is not in agriculture and 0 if main employment is in agriculture. The main analysis sample includes all individuals who are employed and have positive earnings and positive hours worked to ensure that the main variable of interest, the log wage, is defined. The sample includes 31,843 individuals and 275,600 individual-year observations.<sup>12</sup>

In addition to studying earnings, we explore consumption to get a broader sense of overall welfare and total income. IFLS consumption data were collected by directly asking households the value in Indonesian Rupiah of all food and non-food purchases and consumption in the last month,

---

<sup>11</sup> De Mel, McKenzie, and Woodruff (2009) argue that self-reported profits give a more accurate depiction of firm profits in microenterprises than reconstructed measures.

<sup>12</sup> The panel is unbalanced due to attrition, death, and to limiting observations to respondents at least 16 years old.

similar to consumption data collection in the World Bank’s Living Standards Measurement Surveys.<sup>13</sup> In contrast to the retrospective earnings data in the IFLS, the consumption data are all contemporaneous to the survey. Consumption data were collected at the household level, which we divide by the number of household members to obtain a per capita measure, and are presented in real terms, taking into account prices in rural and urban areas. The consumption sample includes 82,272 individual-year observations from 34,820 individuals in IFLS rounds 1–5. In the consumption analysis, we expand the sample to also include individuals without current earnings data; we also perform a robustness check using the main productivity sample.

Data were collected on the respondent’s location at the time of the survey, and all rounds of the IFLS also collected a full history of migration within Indonesia. All residential moves across sub-districts (“kecamatan”) that lasted at least six months are included, i.e., seasonal migration is excluded. Figure 2, Panel A presents a map of Indonesia with each dot representing an IFLS respondent’s residential location. While many respondents live on Java, we observe considerable geographic coverage throughout the country. The IFLS also asked respondents for the main motivation of each move. Family-related reasons are most common at 50%, especially for women (53%), who are more likely than men to state they migrated for marriage. The second most common reason to migrate is for work (32%), with little difference by gender, while migrating for education is less common. We combine data across IFLS rounds to construct a 28-year panel, from 1988 to 2015 with annual information on the person’s location, in line with the employment panel; refer to Kleemans (2016) and Kleemans and Magruder (2017) for more information on the IFLS employment and migration panel.

---

<sup>13</sup> Note that for a small number of frequently-consumed items, information was collected for the last week, and for a few low-frequency items, data was collected for the last year.

We utilize a survey-based measure of urban residence: if the respondent reports living in a “village”, we define the area to be rural, while they are considered urban if they answer “town” or “city.” We present the correspondence between urban residence and employment in the non-agricultural sector in Table 1, Panel A. In 69% of individual-year observations, people are employed in the non-agricultural sector, and in 35% 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, as expected.

Given the migration focus of the analysis, it is useful to report descriptive statistics both for the main 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 sample period (“Always Rural”), those who were born in a rural area but move to an urban area at some point (“Rural-to-Urban Migrants”), those who are “Always Urban,” and finally, the “Urban-to-Rural Migrants” (born urban but move to a rural area at some point). As discussed above, the fixed effects analysis is driven by individuals who move between sectors during the sample period.

In the main IFLS analysis sample, 87% of adults had completed at least primary education, and more than a third had completed secondary education, while tertiary education remain quite limited, at 11%. Among those who are born in rural areas 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 roughly 0.2 standard

deviation units higher among those who migrate to urban areas, a meaningful effect.<sup>14</sup> Migration rates do not differ substantially by gender.

These relationships are presented in a regression framework in Table 3, Panel A (columns 1 to 5), and the analogous relationships for moves 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 (column 6 of both tables), at 99% confidence. This indicates that sorting on difficult-to-observe characteristics is relevant in understanding sectoral productivity differences.

It is worth noting that if we ignore migrants, individuals who are born and remain in urban areas are far more skilled than those who stay in rural areas. “Always Urban” individuals score over 0.3 standard deviation units higher on Raven’s matrices and have almost triple the rate of secondary schooling and four times the rate of tertiary education relative to “Always Rural” individuals. The urban-to-rural migrants in Indonesia are also negatively selected relative to those who remain urban residents, consistent with Young’s (2013) claim. These patterns emerge in Table 2, Panel A, where the urban-to-rural migrants score lower on all skill dimensions relative to those who remain urban; appendix Tables A1 and A2 report analogous results among those individuals born in urban areas.

### **3B. Kenya**

The Kenya Life Panel Survey (KLPS) includes information on 8,999 individuals who attended primary school in western Kenya in the late 1990s and early 2000s, following them through adolescence and into adulthood. These individuals are a representative subset of participants in two school-based

---

<sup>14</sup> Raven’s Matrices were administered to a subset of individuals in IFLS 3, 4 and 5, namely those 7 to 24 years old. The Raven’s Matrices test is designed to capture fluid intelligence.

randomized interventions: a scholarship program for upper primary school girls that took place in 2001 and 2002 (Kremer, Miguel, and Thornton 2009) and a deworming treatment program for primary school students during 1998–2002 (Miguel and Kremer 2004). In particular, the KLPS sample contains information on individuals enrolled in over 200 rural primary schools in Busia district at the time of these programs’ launch. According to the 1998 Kenya Demographic and Health Survey, 85% of children in Western Province aged 6–15 were enrolled in school at that time, and Lee et al. (2015) show that this area is quite representative of rural Kenya as a whole in terms of socioeconomic characteristics. To date, three rounds of the KLPS have been collected (2003–05, 2007–09, 2011–14).

KLPS data collection was designed with attention to minimizing bias related to survey attrition. Sample individuals who had left the original study area were tracked throughout Kenya (as well as into neighboring Uganda and beyond, although we exclude international migrants from the present analysis).<sup>15</sup> Respondents were sought in two separate “phases” of data collection: the “regular tracking phase” proceeded until over 60% of respondents had been surveyed, at which point a representative subset of approximately 25% of the remaining sample was chosen for the “intensive tracking phase” (and remaining unfound individuals no longer sought). These “intensive” individuals receive roughly four times as much weight in the analysis, to maintain representativeness with the original sample. The effective tracking rate for each KLPS round is roughly 85%.<sup>16</sup>

Similar to the IFLS, the KLPS includes information on educational attainment, labor market participation, and migration choices. Employment data was collected in wage employment and self-employment modules, designed to capture both formal and informal employment. Most individuals were quite young (typically teenagers) during data collection for KLPS Round 1, and few had wage

---

<sup>15</sup> The results presented below are robust to the inclusion of international migrants (not shown).

<sup>16</sup> Baird et al. (2008) describes the motivation behind this methodology and calculates the effective tracking rate.

employment or self-employment to report. Full employment histories, including more detailed questions, were collected during Rounds 2 and 3, and it is from these rounds that we draw the data on individual earnings, hours worked, and wages used in the present analysis.

The Kenya agricultural productivity data deserves detailed discussion. Whenever total household annual agricultural sales were at least moderate, exceeding 40,000 Kenyan Shillings (approximately 400-500 US dollars), full agricultural production and profit information was collected in the self-employment module and included in the present analysis. Agricultural wage employment is also common, and these data are always included. Limited questions on subsistence agricultural production were collected in KLPS rounds 1 and 2, but these are insufficient to create an individual productivity measure; more detailed information on agricultural productivity (in the previous 12 months) is contained in round 3, and this is included in the present analysis. To create a measure of individual productivity comparable with other sectors, we focus on agricultural activities (e.g., growing a particular crop) in which the respondent provided all reported labor hours; we also restrict attention to activities in which the respondent reports being the main decision-maker, since it seems likely that they are most knowledgeable about such activities (although results are not sensitive to this restriction). The profit in an agricultural activity is the sum of all crop-specific production – valued either through actual sales or at the relevant crop price (collected in regular local market price surveys) if consumed directly – minus all input costs and hired labor costs. The individual wage divides this net profit by the labor hours the respondent supplied to the activity. Given possible measurement concerns, we show below that estimates are robust to alternative approaches to constructing individual agricultural productivity, including the exclusion of subsistence agriculture data.

KLPS respondents reported industry for all wage and self-employment. Most individuals are engaged in relatively low-skilled work. The most common industry for wage employment is services,

at 57% overall and 74% for females (with many women employed in domestic services). In rural areas, the most common industries for wage employment are services and agriculture (50 and 21%, respectively), while in urban areas they are services, and manufacturing and construction (62 and 11%). The largest self-employment industries are retail and services (41 and 25%).<sup>17</sup>

KLPS round 3 collected detailed consumption expenditure data for a subset of individuals. However, because it was only collected for this round, we are unable to utilize it in panel estimation. Instead, in the panel analysis we utilize a proxy for consumption, the number of meals eaten in the previous day, which is available in both KLPS rounds 2 and 3. Reassuringly, meals eaten is strongly correlated with our primary measures of labor productivity as well as consumption expenditures per capita (in KLPS-3); see Appendix Table A3. As with Indonesia, in the meal consumption analysis, we are able to expand the sample to also include individuals without current earnings data.

KLPS respondents provide a history of residential locations since their last interview, and this data includes residential district, town, and village, allowing us to classify individuals who lived in towns and cities as urban residents. The KLPS includes information on all residential moves that lasted at least four months in duration, a slightly more permissive definition than in the IFLS, and we are able to construct a monthly residential panel from March 1998 to October 2014.<sup>18</sup> Combined with the retrospective labor productivity data, the main analysis sample is a monthly panel with 134,221 individual-month observations for 4,791 individuals.

---

<sup>17</sup> For wage employment, respondents also report occupation, and these tell a similar story regarding skill level. The most common occupations fall in the “unskilled trades” category (32%), followed by “skilled and semi-skilled trades” (19%), “retail and commercial” (18%), “professionals” (16%), and “agriculture” (15%). Agricultural wage employment is more common for men than for women (20 compared to 6%, respectively), and as expected, agricultural employment is far higher in rural than urban areas (29 vs. 5%). Common urban occupations are “unskilled trades” (37%), “skilled and semi-skilled trades” (22%) and “retail and commercial” (20%).

<sup>18</sup> Similar to the IFLS, the panel is unbalanced due to attrition, death, and inclusion of individuals 16 and older.



Figure 2, Panel B presents a map of Kenya, with each dot representing a respondent residential location during 1998–2014. Most residences in western Kenya are in Busia district (where the sample respondents originally resided), with substantial migration to neighboring areas as well as to cities. Appendix Table A4 presents the list of main towns and cities, and shows that 70% of urban residential moves are to Kenya’s five largest cities, namely, Nairobi, Mombasa, Kisumu, Nakuru, and Eldoret. Men are slightly more likely than women to report migrating for employment reasons (60% of moves compared to 55% for females) while women are more likely to migrate for family reasons, including marriage (13% vs. 1% for men). A smaller share of moves (approximately 6%) are for education.

Summary statistics on employment sector and urban residence for KLPS respondents are presented in Table 1, Panels B and C. Panel B presents data for the main analysis sample; as described above, this contains subsistence agricultural information where available (from KLPS-3). The employment share in agriculture is much higher in rural areas (26.1%) than urban (5.4%), as expected, but the share in rural areas is somewhat lower than expected, likely because subsistence agricultural activities were not captured in earlier KLPS rounds. For a more complete portrait, Panel C focuses on data from the 12 months prior to the KLPS-3 survey, which contains detailed information on subsistence agriculture, and here the agricultural employment share in rural areas is much higher.

Recall that the Kenya sample is all rural at baseline (they were originally attending rural schools). Similar patterns emerge regarding positive selection into urban migration, with educational attainment and normalized Raven’s matrix scores both far higher among those who migrate to cities (Table 2, Panel B). In particular, there is a raw gap of nearly 0.3 standard deviation units in Raven’s matrix scores between urban migrants and those who remain rural. Overall migration rates in Kenya are similar for females and males. Tables 3 and 4 (Panel B) report these patterns in terms of regression estimates, for urban migration and employment in non-agricultural work, respectively. As with

Indonesia, controlling for educational attainment and gender, the Raven's score is strongly positively correlated with urban migration (at 99% confidence).

## 4. Results

### 4A. Main Agricultural and Urban 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 somewhat smaller at 70 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 the IFLS, and which is more likely to capture earnings in informal employment.<sup>19</sup> That said, the raw gap we estimate in the IFLS remains substantial.

Inclusion of control variables similar to those used by GLW to adjust macro data gaps reduces the estimated agricultural productivity gap in the IFLS to 57 and 35 log points (in Table 5, columns 2 and 3). Estimating on the subsample for which we have scores from Raven's matrix tests, the gap is reduced slightly, although note the smaller sample size in this case.

Limiting the analysis to those who have productivity measurements at some point in both agricultural and non-agricultural employment, the productivity gap drops to 25 log points (col. 5), suggesting that selection on unobservable characteristics may be meaningful. Inclusion of fixed effects also reduces the gap (col. 6), and using our preferred labor productivity measure, the log wage (namely, the log of total earnings divided by hours worked), as the dependent variable nearly

---

<sup>19</sup> This comes from log transformed values from the "Average" row of GLW, Table 4, i.e.  $\ln 2.6 - \ln 2.2 = 0.167$ .

eliminates the gap: the coefficient estimate falls to 0.078 (standard error 0.021) in column 7, and further to 0.076 when considering the real log wage (adjusting for higher urban prices, col. 8).

We follow a similar approach for Kenya, where the raw agricultural productivity gap falls from 78 log points to 55 with the inclusion of controls (Table 5, Panel B, columns 1–4), and to 28 log points when including an individual fixed effect. Using the preferred hourly wage measure reduces the gap to 6.1 log points (col. 7), it falls further when adjusted with an urban price deflator (col. 8), and neither fixed effects wage estimate is significant at traditional levels of confidence.

Comparing column 1 (the raw gap) to column 7 (the preferred fixed effects estimate) in Table 5, the agricultural productivity gap is reduced by 88% in Indonesia and by 92% in Kenya. The standard errors are somewhat larger for Kenya, so the upper end of the 95% confidence interval includes a sizable gap of 37 log points, consistent with some non-trivial productivity gains to non-agricultural employment. That said, even this value remains far lower than the 108 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 noted in the introduction, these results for Indonesia and Kenya are presented graphically in Figure 1, Panels A and B and compared to GLW's estimated productivity gaps.<sup>20</sup>

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. To the extent that residential migration is costlier than shifting jobs (but not homes), and

---

<sup>20</sup> Similar patterns are obtained when using alternative definitions of non-agricultural employment, namely, classifying simultaneous work in both sectors as agriculture or as non-agriculture (Appendix Table A5), a point we return to below.

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 the non-agricultural wage premium.

The microdata estimates from Indonesia and Kenya appear to be consistent with this view, at least at first glance: the raw gap reported in column 1 of Table 6 (Panels A and B) are 54 and 86 log points for Indonesia and Kenya, respectively. Similar to the agricultural productivity gap, the urban-rural productivity gap falls when additional explanatory 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 2.8 log points with the inclusion of individual fixed effects in Indonesia (column 6), and 0.2 log points for the preferred log wage measure (column 7). The analogous urban productivity effect estimate for Kenya is slightly larger at 16.5 log points (column 7). Thus, the productivity gap in Indonesia falls by 100% in Indonesia (to zero), and the reduction for Kenya is 81% (from 86.2 to 16.5 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).<sup>21</sup> Urban productivity gaps in real wage terms (that account for higher urban prices) are further reduced in both countries (column 8).

The selection model (in Section 2) predicts that estimated productivity gaps would be higher among rural-to-urban migrants than for urban-to-rural migrants, given plausible patterns of selection bias. Table 7 explores this hypothesis in Indonesia by separately conditioning on birth location; Panel A contains those born rural and Panel B those born in urban areas. The same pattern of declining urban productivity gaps in each subsample is observed as additional controls are included (columns 1-3). In

---

<sup>21</sup> Appendix Table A6 further explores the relationship between these two gaps by conditioning on observations in rural areas. The raw agricultural productivity gap and subsequent decline with the inclusion of controls and fixed effects is quite similar in Indonesia in rural areas (panel A), and even more pronounced in Kenya.

the preferred log wage specification in column 3, productivity gaps are indeed somewhat larger for those born in rural areas, as predicted by the sorting model. The estimated productivity gain to urban employment is 3.7 log points for those born in rural areas (Panel A) and -2.4 log points for those born urban (Panel B). The difference between estimates for those born in rural versus urban areas is significant (p-value=0.03), but is relatively small, suggesting tight bounds around zero.

There are a number of alternative measures of individual agricultural productivity that are worth considering to assess robustness of the main results. Appendix Figure A3 illustrates how each source of agricultural productivity data in both the IFLS and KLPS contributes to the overall sample, and classifies measures into those that are more reliably measured (e.g., hourly wage work), and those that are less reliably measured (e.g., measures based on production in subsistence agriculture). We next assess robustness to different definitions of employment in agriculture, including if the majority of hours are in the sector (our main measure), as well as measures that classify an individual as working agriculture if any hours are in the sector, or alternatively if all hours are in the sector. We additionally explore robustness to the use of both wage earnings and self-employed profits in agriculture (main measure), versus measures that use only one or the other. For both Indonesia (Table 8, Panel A) and Kenya (Panel B), estimated agricultural productivity gaps remain small and positive across five alternative measures, ranging from 1 to 12 log points in Indonesia and 0 to 16 log points in Kenya.<sup>22</sup>

---

<sup>22</sup> Further details and robustness checks are contained in Appendix Tables A7-A10. In Appendix Table A7, we present estimates for Indonesia on a sample of individuals who are at most 30 years old, for greater comparability with the Kenya sample, which consists of young adults; the estimates remain similar. Appendix Tables A8 and A9 report results separately for wage earnings and self-employment earnings, respectively, and generate similar results. Appendix Table A10 reports results for Kenya including subsistence agriculture even when the respondent is not the main decision maker for an activity (Panel A), and excluding subsistence agriculture entirely (Panel B), and results are robust.

#### **4B. Productivity versus Living Standards**

The discussion above establishes at least an 80% reduction in estimated sectoral productivity gaps once individual fixed effects are included in the analysis (Figure 1). The wage measures presented thus far are closely related to the labor productivity 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, amenities, unemployment, and other factors. For instance, there could be considerable individual heterogeneity in the taste for rural versus urban amenities, e.g., comforts of home, ethnic homogeneity, better informal insurance, etc., in rural areas versus cosmopolitan cities’ better public goods and more novelty (but downsides too, such as crime). Moving itself may also impose large utility costs (Kleemans 2016).

Although it is impossible to fully capture these factors and convincingly measure individual welfare, to get somewhat closer to differences in living standards, we draw on consumption data from the IFLS. As described in Section 3, five rounds of the IFLS included questions on the value of household consumption which can be converted to per capita consumption. In the main specification, we include all individuals who have such consumption data, even if they lack earnings measures. Consumption expenditures may also more accurately capture total household income in low income settings like ours with extensive subsistence agriculture, home production, and informal employment, all of which are challenging to measure, making it an attractive alternative to earnings data. The measure should also capture variation in total earnings caused by unemployment or job rationing.

The initial consumption gap between non-agriculture and agriculture is large and similar the productivity gap at 64 log points (Table 9, Panel A). The gap falls considerably when including time fixed effects and control variables in column 2, and falls to only 12.6 log points when also including

individual fixed effects in column 3. A similar pattern is presented for the urban-rural consumption gap in columns 4, 5, and 6: the gap declines from 40 log points to 2.6 log points.<sup>23</sup>

We next explore the estimated urban consumption premium for those born in rural versus urban areas (Table 7, columns 4–6). In the preferred specification with individual fixed effects (col. 6), the urban consumption premium is larger for those born in rural areas (13.3 log points, Panel A) than those born in urban areas (-4.7 log points, Panel B), and the difference is highly significant (p-value<0.001). As with the earnings results, this is consistent with the predictions of the selection model (in Section 2), and suggests the urban premium is bounded rather tightly around zero.

The consumption proxy measure in the KLPS tells a similar story. The raw gap in meals eaten in Kenya between those in non-agriculture versus agricultural employment is positive and statistically significant, though smaller than the earnings gap (Table 9, Panel B); differences in magnitude are difficult to interpret given the different nature of the meals measure, and the possibility that it changes most at very low levels of income. Mirroring the broad pattern observed for labor productivity, this gap falls by almost half when including controls, and is actually slightly negative when including individual fixed effects (columns 1-3); a similar pattern holds for the urban-rural gap (columns 4-6).

Another dimension of welfare relates to patterns of unemployment. Appendix Table A13 explores whether there are differences in unemployment rates and search behavior between urban and rural areas for Kenya, where this data is available. We find that unemployment (measured several ways) is either similar in urban and rural areas (Panel A, column 3) or somewhat higher in urban than

---

<sup>23</sup> Appendix Table A11 shows the gap in both food and non-food consumption (Panels A and B, respectively). The gaps in both components of consumption see reductions of 77-95% when including individual fixed effects. Appendix Table A12 repeats the consumption analyses on the main analysis sample (i.e., those with earnings data) for total consumption (Panel A) and by food and non-food consumption (Panels B and C, respectively), and results are similar.

in rural areas conditional on individual fixed effects (Panel A, column 6, and Panel B), strengthening the main finding that movers to urban areas may not experience large gains in total earnings.<sup>24, 25</sup>

#### 4C. Sector-specific Productivity — Absolute and Comparative Advantage

In the conceptual framework, the richest model of human capital allowed for individual sector-specific productivity  $\theta_{is}$ . Analysis of these productivities has been given renewed focus in Lagakos and Waugh (2013), who argue that self-selection on the basis of comparative advantage could play an important role. In their model, comparative advantage is positively correlated with absolute advantage, meaning that the most productive workers have the most to gain from 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.<sup>26</sup> We recover these estimates, and then normalize the mean of the fixed effects of permanent rural residents (non-movers) to be zero. Figure 3 presents the joint distributions of these estimated individual productivities by sector. Panel A includes Indonesians born in rural areas. It is apparent that rural-to-urban migrants are positively selected relative to non-migrants, with an average rural wage approximately 20 log points higher than non-migrants. These individuals experience only a 6 log point average increase in their wage upon migration to an urban area. Panel B presents the same exercise with Indonesians born in urban areas. Here, there appears to

---

<sup>24</sup> In this paper, we consider *mean differences* in productivity or consumption across sectors, but *variability of outcomes* could also be a determinant of individual wellbeing, as well as of migration choices (Munshi and Rosenzweig 2016). We test whether the variability of earnings in the agricultural (rural) sector is different than variability in the non-agricultural (urban) sector, conditional on individual fixed effects, and find mixed results. There are no statistically significant differences in variability across sectors in the Kenya sample. There is significantly more variability in agricultural (rural) wages and earnings in Indonesia relative to the non-agricultural sector (urban areas), although no significant differences in consumption variability (not shown). We leave additional exploration of these issues for future research.

<sup>25</sup> Results are unchanged when using alternative approaches to accounting for clustering (Bell and McCaffrey 2002, Cameron and Miller (2015), and Young (2016); see appendix Table A14.

<sup>26</sup> This procedure is similar in spirit to the correlated random coefficient models utilized to analyze heterogeneous returns to hybrid seed adoption (Suri 2011) and labor unions' effects on wages (Card 1996, Lemieux 1998), although our approach makes fewer assumptions and is meant to be more descriptive.



be negative selection into rural migration, with the average mover having 20 log points lower wages when still in urban areas, and an increase of only 2 log points in rural wages among moving. Panel C presents results in Kenya (all of whom were rural residents as children) that are analogous to panel A. Compared to Indonesia, there appears to be even more positive selection among urban migrants in Kenya (at 41 log points) as well as a moderate positive urban premium of roughly 16 log points, which is nearly identical to the regression adjusted estimate presented above.

Note that the realizations of roughly half of migrants fall below the 45 degree line in the three panels of Figure 3, which taken literally means that they experience higher earnings in rural than urban areas. This is consistent with the empirical finding of zero or small positive sectoral productivity gaps.

This exercise is meant to be descriptive, and we interpret the relationships between the estimated individual urban and rural productivities with caution here, in part because the estimates are subject to measurement error and thus the fitted regression line may experience attenuation bias. With these caveats in mind, note that all three plots appear to show that absolute advantage plays a role in wage determination: individuals who have high rural productivity tend to have high rural productivity, and vice versa, indicated by the positive slope.

#### **4D. Dynamics of the Productivity Gap and Big City Effects**

In unpacking the main result, we examine if dynamics and experience effects produce productivity gains that do not materialize right away. In particular, while the main specification includes time fixed effects which would account for overall growth of wages as the sample ages or year specific shocks, individuals may begin to earn more after spending time in urban areas. Figure 4 presents event study analyses of whether individuals earn more after migrating, where we estimate regressions of the form:

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

These regressions are estimated on an unbalanced panel of individual-time periods and include individual fixed effects  $\theta_i$ , time fixed effects  $\delta_t$ , squared age as a time-varying covariate  $\mathbf{X}_{it}$ , and additional indicator variables for time periods exceeding five years pre- and post-move,  $\gamma_{pre}$  and  $\gamma_{post}$ , respectively. The Indonesian and Kenyan analyses both condition on individuals being born rural.

The  $\beta_{\tau}$  parameters of primary interest are coefficients on indicators for time periods relative to the individuals' move to an urban area at  $\tau = 0$ . Estimates are relative to the year or month prior to the individuals' move in Indonesia and Kenya, respectively; we exclude an indicator for the period immediately prior to the individuals' move. These coefficients are identified by individuals who have adjacent productivity measures in both the period that they move to urban and the period immediately prior. We do not enforce a requirement that individuals are observed in every period five years prior- and post-move. If the extensive margin decision to exit the labor force entirely or attrit from the sample is correlated with urban labor market experiences, the results may be biased and we thus interpret them with caution. Nonetheless, the richness of the panel dataset is novel and worth exploring.

These parameters represent the difference in mean wages between movers and non-movers net of the difference that existed in the period prior to the urban move. An advantage of this approach is that it also allows us to assess wage dynamics *prior* to the move, which may give some clues about what precipitated the move – e.g., whether rural individuals are more likely to move following a negative earnings shock – and they also allow us to examine whether urban experience leads to gradually rising earnings there.

In Indonesia, urban wages do not change substantially relative to the year prior to moving, and even five years after the urban move, migrants see no average wage gain (Figure 4, Panel A). There are broadly similar results in Kenya relative to the month prior to the move; there is some suggestive indication of slightly rising wages in the first two years of residence in an urban area, but these are small (Panel B). There is no indication of meaningful pre-move trends in either country.

In this analysis, we consider wages for individuals who made an urban move regardless of whether they remained in cities or towns, or later moved back to rural areas. The bottom halves of both panels A and B show a “survival” rate in urban areas of between 50 to 60% after five years (in both countries), suggesting substantial return migration to rural areas. Naturally, one might suspect that those with the worst economic outcomes in urban areas might return home, yet this does not appear to be the case: Appendix Figure A4 separately plots post-move wages for those who remain in urban areas and those who return to rural areas, and we find no evidence of a significant divergence in earnings between these two groups (although note that there is limited statistical power to distinguish between these subgroups). This suggests a direction for future research in uncovering the reasons for these moves, including whether non-economic factors, including family reasons and heterogeneity in the taste for urban living, are often decisive factors.<sup>27</sup>

Other scholars have argued that job experience is particularly valuable in big cities and that residence in these cities may boost individual productivity over time (see de la Roca and Puga 2016 for the Spanish case). We examine this issue, first repeating the main urban productivity gap analysis (from Table 6) but including a breakdown into the five highest population cities in each country, in Table 10. In Indonesia, all five cities are larger than 2 million inhabitants, with the capital Jakarta at 10

---

<sup>27</sup> We carry out an analogous event study of moves to rural areas among those born urban in Indonesia, and similarly find no evidence of significant dynamic impacts (see Appendix Figures A5 and A6).

million. Kenya's capital Nairobi has 3.4 million people, the second largest city (Mombasa) has nearly one million, while the other three cities in Kenya are smaller. The capitals are also the largest destinations for urban migrants in each country.

There is mixed evidence on the extent of big city productivity effects. There is no evidence for significantly larger effects in any of the largest cities in Indonesia, including Jakarta (column 4 of Panel A, Table 10). There is some evidence of significant positive urban productivity gains in the two largest Kenyan cities, Nairobi and Mombasa (Panel B). The total urban effect is moderate and statistically significant in the capital of Nairobi, at 23 log points.<sup>28</sup>

## 5. Conclusion

Several influential recent studies document large sectoral productivity gaps in low-income countries and highlight an apparent puzzle, namely, “why so many workers remain in the agricultural sector, given the large residual productivity gaps with the rest of the economy” (Gollin, Lagakos, and Waugh 2014, p. 941). This study makes two main contributions using data from 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 reduction of over 80% in the estimated gaps. The second main empirical contribution lies in demonstrating that there is extensive individual selection across sectors, both along relatively easily observable

---

<sup>28</sup> While this analysis finds mixed evidence of an overall big city effect in Indonesia and Kenya, we also assess whether effects might manifest over a longer time horizon by repeating the event study analysis over a five year time horizon separately for Jakarta and Nairobi. These figures show no clear evidence of differentially positive dynamic effects in capital cities: differences with other cities are imprecisely estimated and generally not significant (not shown).

dimensions such as educational attainment as well as measures of skill (here, a measure of cognitive ability) that most standard economic datasets lack.

Taken together, the findings point to the importance of individual selection in driving observed sectoral gaps, and call into question strong causal interpretations. As a result, the puzzle of why the share of workers in rural agriculture remains 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.

An historical episode illustrates some of the potential risks of pro-urbanization policies. In the 1970s, Tanzania's authoritarian socialist government sought to move its rural population into larger villages and towns to speed up economic modernization. The underlying idea was that the provision of public services and the shift into non-agricultural work (including manufacturing) would be hastened if households would only leave their traditional homesteads, which were often highly spatially dispersed. After initial rhetorical encouragement and incentives by the government led to few moves, the government resorted 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 the policy in Tanzania since the forced moves were quickly abandoned (within a year) in the face of large-scale resistance, at a minimum, the Tanzanian case indicates that it can sometimes be very costly from a welfare perspective to rapidly induce a large share of the population to move out of traditional rural agriculture.

As noted above, our main 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 small or even zero sectoral gaps with the existing macroeconomic empirical evidence of large average gaps is the possibility that productivity effects among non-movers would be 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 in both sectors for non-movers naturally complicates the rigorous 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, as we do above, in which those with the largest net utility benefits are most likely to move. This could 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 is quite different than those in rural areas, and individuals may have strong and heterogeneous preferences for them, leading to large reductions in utility for some migrants even if wages rise. Poor individuals may also face credit constraints or other financial frictions that prevent them from exploiting wage gaps, and easing these constraints could boost migration rates, as argued for Indonesia by Bazzi (2017) and India by Munshi and Rosenzweig (2016). However, the very long timeframes of both panel datasets employed in this study help to at least partially mitigate this concern: some poor individuals with high returns to migration presumably had access to improved credit at *some* point during 1988-2015 in Indonesia or 1998-2014 in Kenya and managed to move.

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 Bryan et al. (2014) study in

Bangladesh. They find that a modest subsidy did induce a moderate share of recipients (roughly one quarter) to move to towns and cities for temporary work during the agricultural low season; the relatively low rate of migration may indicate that the utility costs of migration are non-trivial. Among movers, there is an estimated increase in per capita consumption among the sending household of roughly 30% over two years, and 25% average gain in earnings (not statistically significant) among those assigned to the subsidy. Overall, the study provides some indication that there are positive returns to temporary seasonal migration among rural workers who are typically non-movers.

Nonetheless, the earnings gains are fairly modest in size and note that they are closer in magnitude to the small gaps we estimate in this paper than to those found in many other recent contributions. It is also worth noting that the subsidy was delivered during times of the year (the agricultural low season) when agricultural productivity was thought to be particularly meager, and targeted to regions thought especially likely to benefit from seasonal moves, suggesting that the 30% consumption gain in Bryan et al (2014) is an upper bound on the return to permanent urban migration in Bangladesh as a whole.

The case of urban-born non-movers is less well understood and raises some intriguing possibilities. Recall (from Table 2) that individuals raised in urban areas have much higher cognitive scores (on a test of fluid intelligence) 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 (e.g., Gertler et al. 2014), growing up in a city might generate higher average adult 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. Anecdotally, many migrants do claim to move in order to improve their children's economic wellbeing more than their own. 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, persistent and real causal urban-rural productivity gaps overall.<sup>29</sup>

The study of sectoral productivity gaps remains an area ripe for further research. Some natural next steps include extending our long-run panel data analysis to new countries and settings (as appropriate data becomes available); conducting more experiments to induce at least partially random selection in migration, thus generating “local” estimates in new sub-populations and improving understanding of the nature of constraints facing potential migrants; and exploration of very long-run and even inter-generational effects of sectoral and residential choice on child ability and productivity.

---

<sup>29</sup> Indeed, Nakamura, Sigurdsson, and Steinsson (2016) study migration induced by a volcanic explosion in Iceland, and show that adult movers gain little from moving out of a rural area but their children earn far more in the long-run.



## References

- Alvarez, J. A. (2015) “The Agricultural Wage Gap: Evidence from Brazilian Micro-data,” Working paper.
- Au, C.-C. and Henderson, J. V. (2006) “How migration restrictions limit agglomeration and productivity in China,” *Journal of Development Economics*, Vol. 80, pp. 350–388.
- Baird, S., Hamory, J., and Miguel, E. (2008) “Tracking, Attrition and Data Quality in The Kenyan Life Panel Survey Round 1 (KLPS-1),” *University of California CIDER*, Working Paper.
- Bazzi, S. (2017) “Wealth Heterogeneity and the Income Elasticity of Migration,” *American Economic Journal: Applied Economics*, Vol. 9, pp. 219–55.
- Bazzi, S., Gaduh, A., Rothenberg, A. D., and Wong, M. (2016) “Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia,” *American Economic Review*, Vol. 106, pp. 2658–2698.
- Beegle, K., Weerdt, J. D., and Dercon, S. (2011) “Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey,” *Review of Economics and Statistics*, Vol. 93, pp. 1010–1033.
- Bell, R. M. and McCaffrey, D. F. (2002) “Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples,” *Survey Methodology*.
- Bryan, G., Chowdhury, S., and Mobarak, A. (2014) “Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh,” *Econometrica*, Vol. 82, pp. 1671–1748.
- Bryan, G. and Morten, M. (2017) “The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia.”
- Cameron, A. C. and Miller, D. L. (2015) “A Practitioner’s Guide to Cluster-Robust Inference,” *Journal of Human Resources*, Vol. 50, pp. 317–372.
- Card, D. (1996) “The Effects of Union on the Structure of Wages: A Longitudinal Analysis,” *Econometrica*, Vol. 64, pp. 957–979.
- Caselli, F. (2005) “Accounting for Cross-Country Income Differences,” in P. Aghion and S. Durlauf eds. *Handbook of Economic Growth*, Vol. 1 of Handbook of Economic Growth: Elsevier, Chap. 9, pp. 679–741.
- Chiquiar, D. and Hanson, G. H. (2005) “International Migration, Self-Selection, and The Distribution of Wages: Evidence from Mexico and the United States,” *Journal of Political Economy*, Vol. 113, pp. 239–281.
- de la Roca, J. and Puga, D. (2016) “Learning by Working in Big Cities,” *Review of Economic Studies*, Forthcoming.

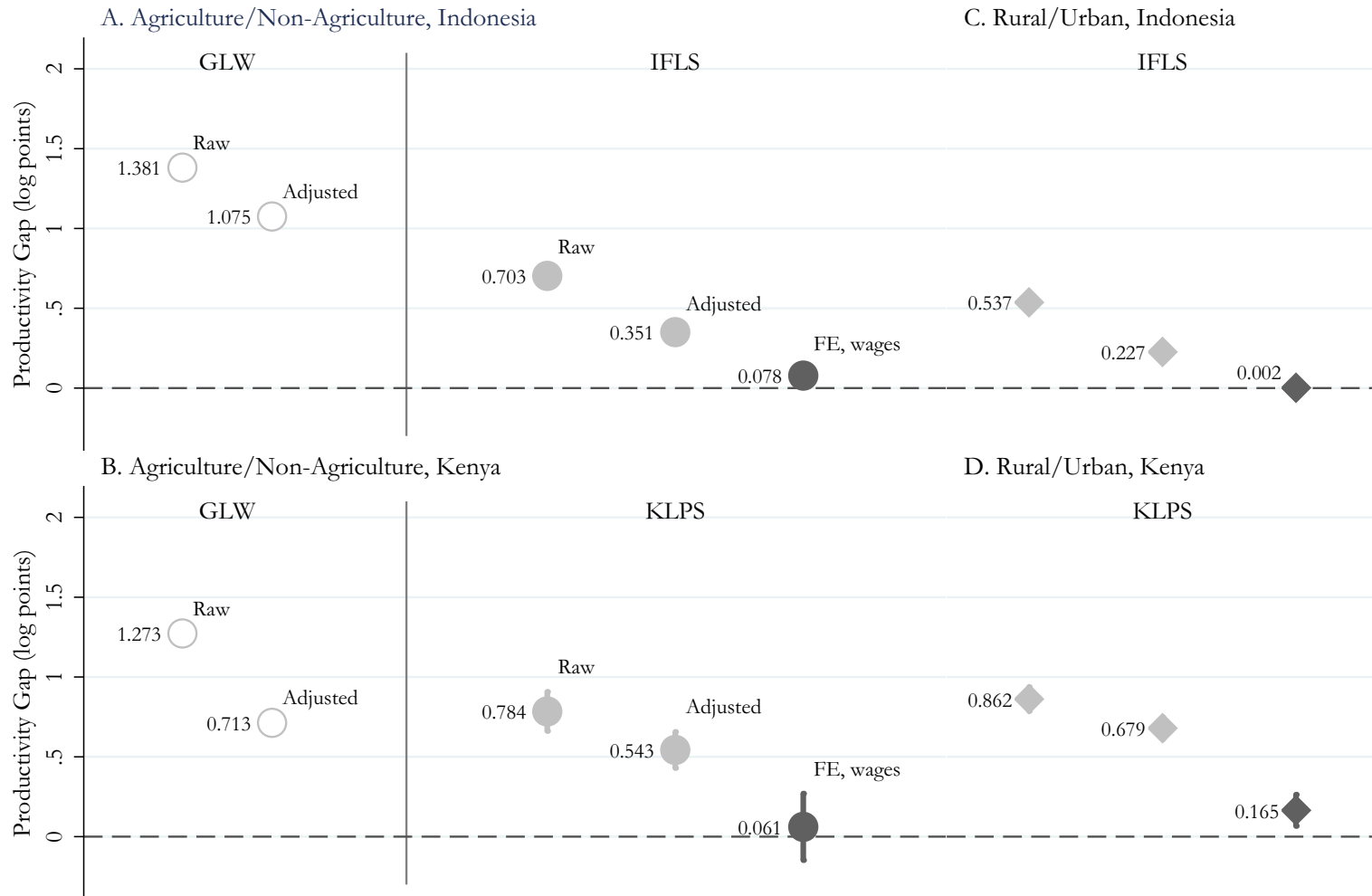
- de Mel, S., McKenzie, D., and Woodruff, C. (2009) “Measuring microenterprise profits: Must we ask how the sausage is made?” *Journal of Development Economics*, Vol. 88.
- Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeerch, C., Walker, S., Chang, S. M., and Grantham-McGregor, S. (2014) “Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica,” *Science*, Vol. 344, pp. 998–1001.
- Gollin, D., Lagakos, D., and Waugh, M. E. (2014) “The Agricultural Productivity Gap,” *Quarterly Journal of Economics*, Vol. 129, pp. 939–993.
- Gollin, D., Parente, S., and Rogerson, R. (2002) “The Role of Agriculture in Development,” *American Economic Review*, Vol. 92, pp. 160–164.
- Graham, B. and Temple, J. (2006) “Rich nations, poor nations: how much can multiple equilibria explain?” *Journal of Economic Growth*, Vol. 11, pp. 5–41.
- Hendricks, L. and Schoellman, T. (2017) “Human Capital and Development Accounting: New Evidence from Wage Gains at Migration,” July.
- Herrendorf, B., Rogerson, R., and Valentinyi, Á. (2014) “Chapter 6 - Growth and Structural Transformation,” in P. Aghion and S. N. Durlauf eds. *Handbook of Economic Growth*, Vol. 2 of Handbook of Economic Growth: Elsevier, pp. 855–941.
- Herrendorf, B. and Schoellman, T. (2016) “Wages, Human Capital, and Structural Transformation,” *Working Paper*.
- Hsieh, C.-T. and Klenow, P. (2009) “Misallocation and Manufacturing TFP in China and India,” *Quarterly Journal of Economics*, Vol. 124, pp. 1403–1448.
- Johnston, B. and Kilby, P. (1978) *Agriculture and Structural Transformation: Economic Strategies in Late-developing Countries*: Oxford University Press.
- Kleemans, M. (2016) “Migration Choice under Risk and Liquidity Constraints,” Working Paper.
- Kleemans, M. and Magruder, J. (2017) “Labor Market Changes In Response To Immigration: Evidence From Internal Migration Driven By Weather Shocks,” *Economic Journal*, accepted.
- Kremer, M., Miguel, E., and Thornton, R. (2009) “Incentives to Learn,” *Review of Economics and Statistics*, Vol. 91, pp. 437–456.
- Kuznets, S. (1973) “Modern Economic Growth: Findings and Reflections,” *American Economic Review*, Vol. 63, pp. 247–258.
- Lagakos, D. and Waugh, M. E. (2013) “Selection, Agriculture, and Cross-Country Productivity Differences,” *American Economic Review*, Vol. 103, pp. 948–980.
- Lee, K., Brewer, E., Christiano, C., Meyo, F., Miguel, E., Podolsky, M., Rosa, J., and Wolfram, C. (2015) “Barriers to Electrification for ”Under Grid” Households in Rural Kenya,” *Development Engineering*, Vol. 1, pp. 26–35.

- Lemieux, T. (1998) “Estimating The Effects of Unions on Wage Inequality in a Panel Data Model with Comparative Advantage and Nonrandom Selection,” *Journal of Labor Economics*, Vol. 16, pp. 261–291.
- Lewis, W. (1955) *The Theory of Economic Growth*, Homewood: R. D. Irwin.
- McKenzie, D., Gibson, J., and Stillman, S. (2010) “How Important Is Selection? Experimental vs. Non-Experimental Measures of The Income Gains from Migration,” *Journal of the European Economic Association*, Vol. 8, pp. 913–945.
- Miguel, E. and Kremer, M. (2004) “Worms: Identifying Impacts on Education and Health in The Presence of Treatment Externalities,” *Econometrica*, Vol. 72, pp. 159–217.
- Munshi, K. and Rosenzweig, M. (2016) “Networks and Misallocation: Insurance, Migration, and The Rural-Urban Wage Gap,” *American Economic Review*, Vol. 106, pp. 46–98.
- Nakamura, E., Sigurdsson, J., and Steinsson, J. (2016) “The Gift of Moving: Intergenerational Consequences of a Mobility Shock,” *Working Paper*.
- Pack, H. (1972) *Employment and productivity in Kenyan manufacturing*: Institute for Development Studies, University of Nairobi.
- 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 Rogerson, R. (2008) “Policy distortions and aggregate productivity with heterogeneous plants,” *Review of Economic Dynamics*, Vol. 11, pp. 707–720.
- Restuccia, D., Yang, D. T., and Zhu, X. (2008) “Agriculture and aggregate productivity: A quantitative cross-country analysis,” *Journal of Monetary Economics*, Vol. 55, pp. 234–250.
- Rosenstein-Rodan, P. N. (1943) “Problems of Industrialisation of Eastern and South-Eastern Europe,” *Economic Journal*, Vol. 53, pp. 202–211.
- Rostow, W. W. (1960) *The Stages of Economic Growth: A Non-Communist Manifesto*, Cambridge: Cambridge University Press.
- Roy, A. D. (1951) “Some Thoughts on the Distribution of Earnings,” *Oxford Economic Papers*, Vol. 3, pp. 135–146.
- Rubalcava, L., Teruel, G., Thomas, D., and Goldman, N. (2008) “The Healthy Migrant Effect: New Findings from the Mexican Family Life Survey,” *American Journal of Public Health*, Vol. 98, pp. 78–84.
- Schultz, T. P. (1988) *Handbook of Development Economics*, Vol. 1, Chap. 13 Education Investments and Returns, pp. 543–630: Elsevier.

- Strauss, J., Witoelar, F., and Sikoki, B. (2016) “The Fifth Wave of the Indonesian Family Life Survey (IFLS4): Overview and Field Report,” *WR-1143/1-NIA/NICHD*.
- Stren, R., Halfani, M., and Malombe, J. (1994) *Beyond Capitalism vs. Socialism in Kenya and Tanzania*, Chap. Coping with Urbanization and Urban Policy, Colorado: Lynne Rienner.
- Suri, T. (2011) “Selection and Comparative Advantage in Technology Adoption,” *Econometrica*, Vol. 79, pp. 159–209.
- Thomas, D., Witoelar, F., Frankenberg, E., Sikoki, B., Strauss, J., Sumantri, C., and Suriastini, W. (2012) “Cutting the costs of attrition: Results from the Indonesia Family Life Survey,” *Journal of Development Economics*, Vol. 98, pp. 108–123.
- Yang, D. (2006) “Why Do Migrants Return to Poor Countries? Evidence from Philippine Migrants’ Responses to Exchange Rate Shocks,” *Review of Economics and Statistics*, Vol. 88, pp. 715–735.
- Young, A. (2013) “Inequality, The Urban-Rural Gap, And Migration,” *Quarterly Journal of Economics*, Vol. 128, pp. 1727–1785.
- (2016) “Improved, Nearly Exact, Statistical Inference with Robust and Clustered Covariance Matrices using Effective Degrees of Freedom Corrections,” January.

# Tables and Figures

Figure 1: Productivity Gap in Total Earnings



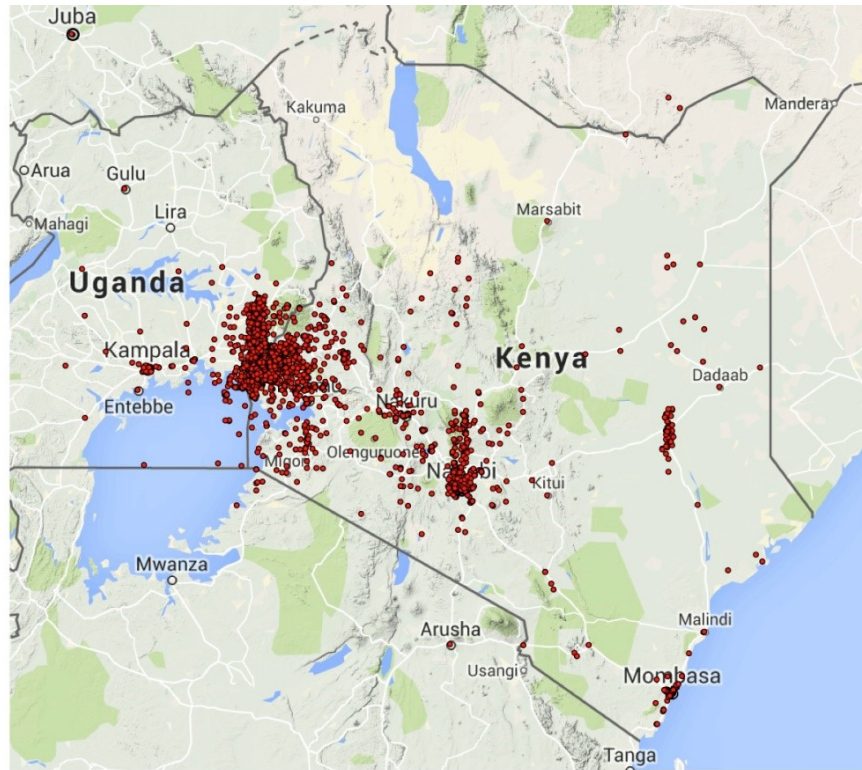
Notes: GLW refers to estimates from Gollin, Lagakos, and Waugh (2014), Online Appendix Table 4. For comparability, the figure reports log transformed numbers from their columns 4 and 5 for Indonesia and Kenya, respectively. Symbols here represent point estimates, and vertical lines represent 95% confidence intervals. Panel A estimates from the 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, wages” is the fixed effects regression estimate of wages on an urban indicator and squared-age from column (7). Corresponding estimates from the 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 the IFLS are smaller than the size of the symbols and are therefore not visible.

Figure 2: Sample Areas

(A) Indonesia Family Life Survey



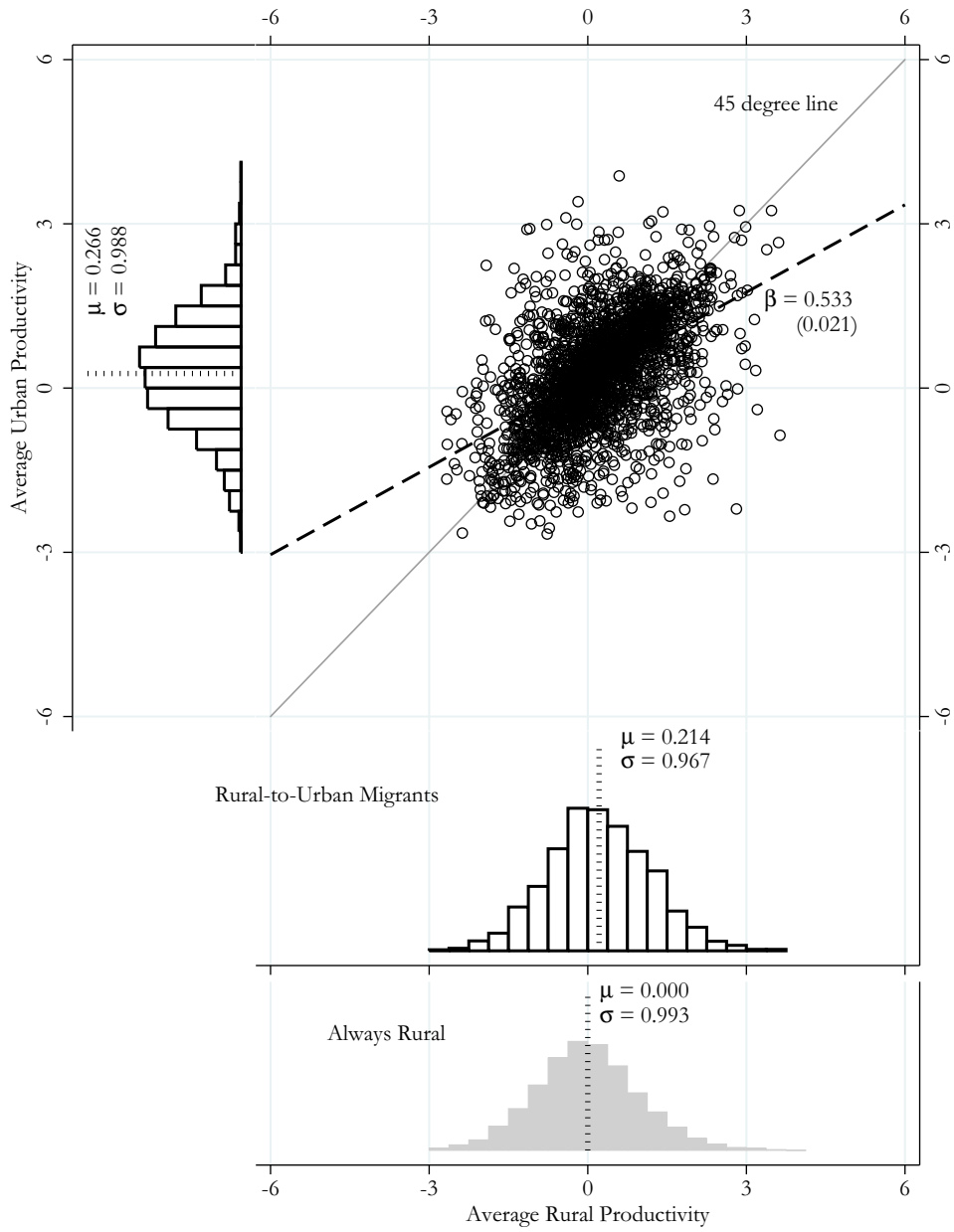
(B) Kenya Life Panel Survey



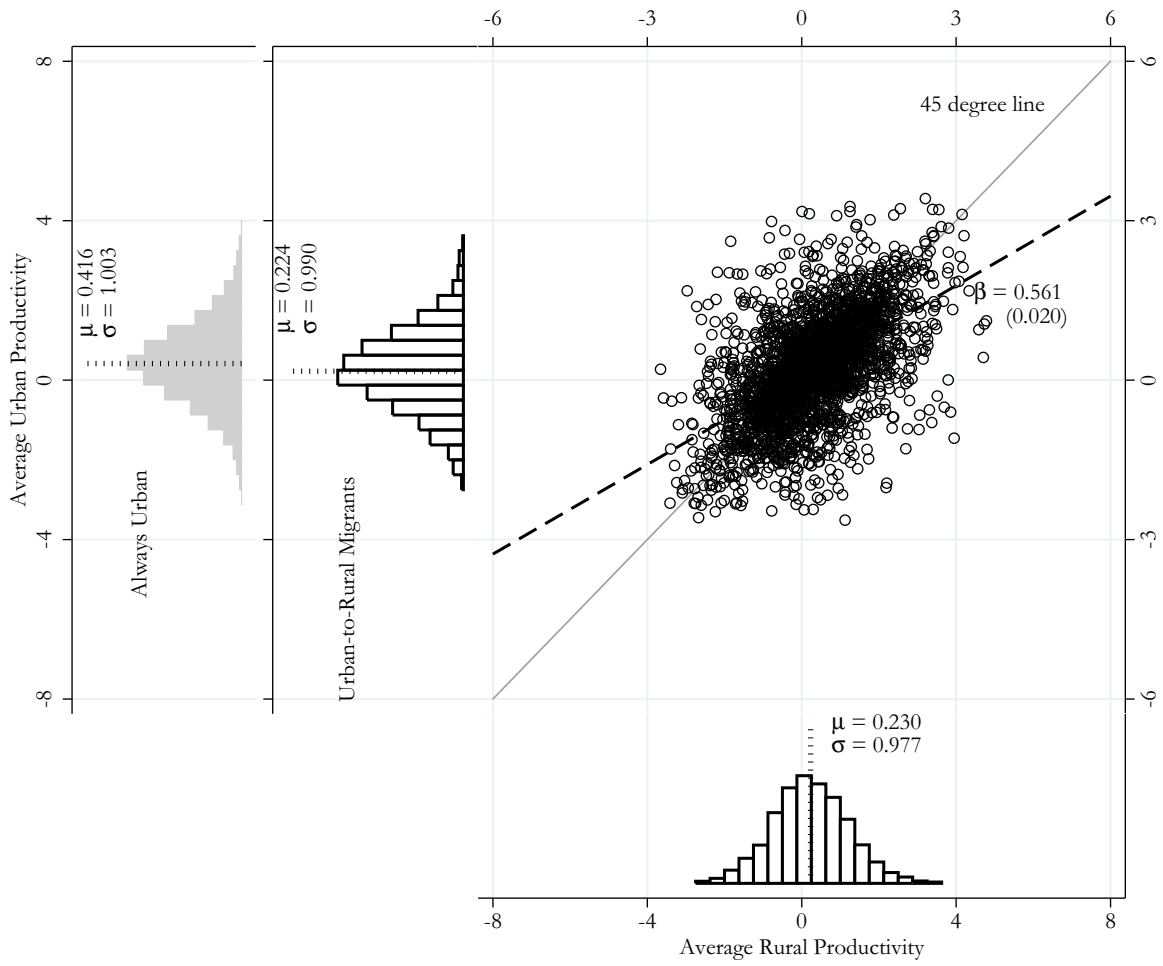
Notes: Panel A shows the residential locations of individuals during the 1988–2008 sample period of rounds 1–4 of the IFLS. For the Kenyan sample, Panel B shows individuals' residential locations during the 1998–2014 sample period that was collected during rounds 2 and 3 of the KLPS. Individuals living outside of Kenya are dropped from the analysis. 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 (Born Rural)

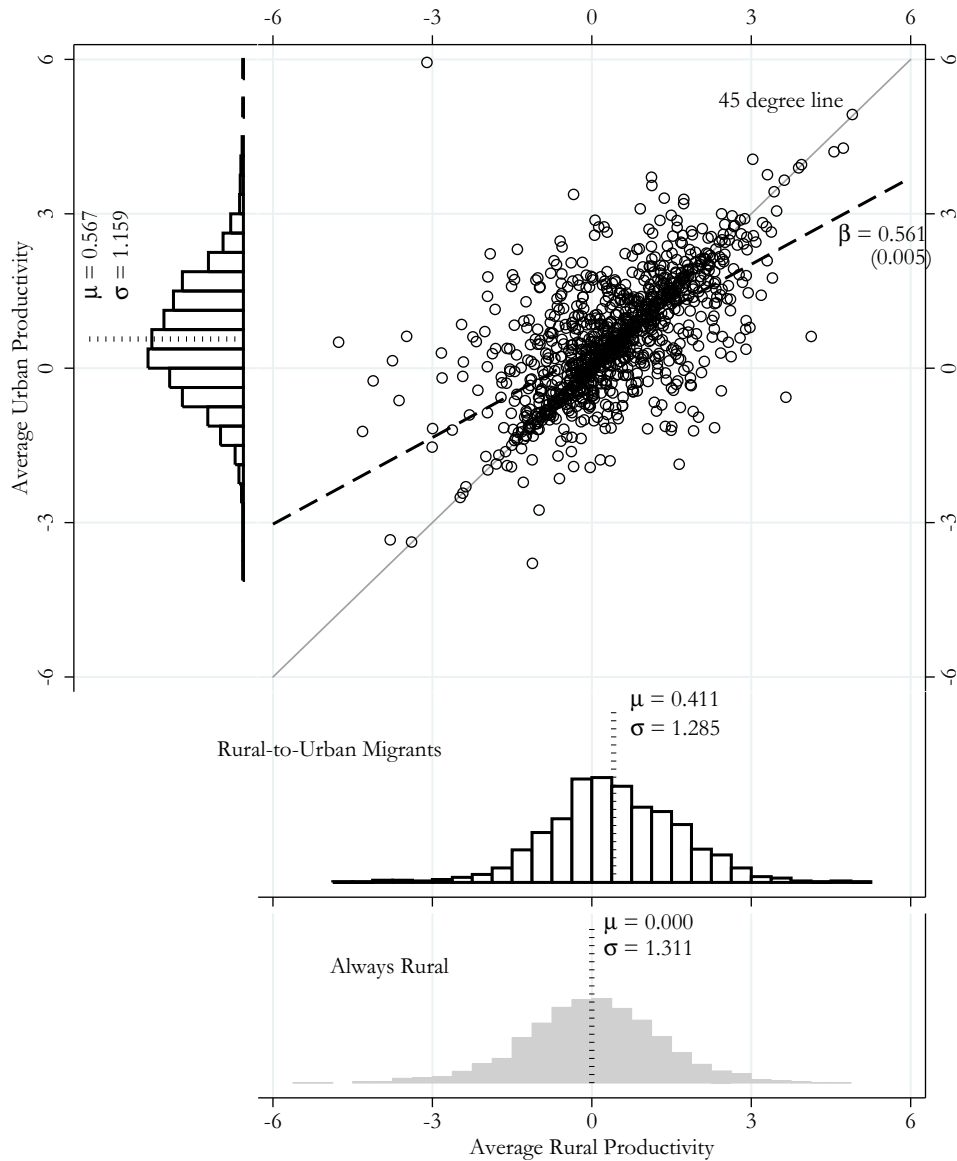


(B) Indonesia (Born Urban)



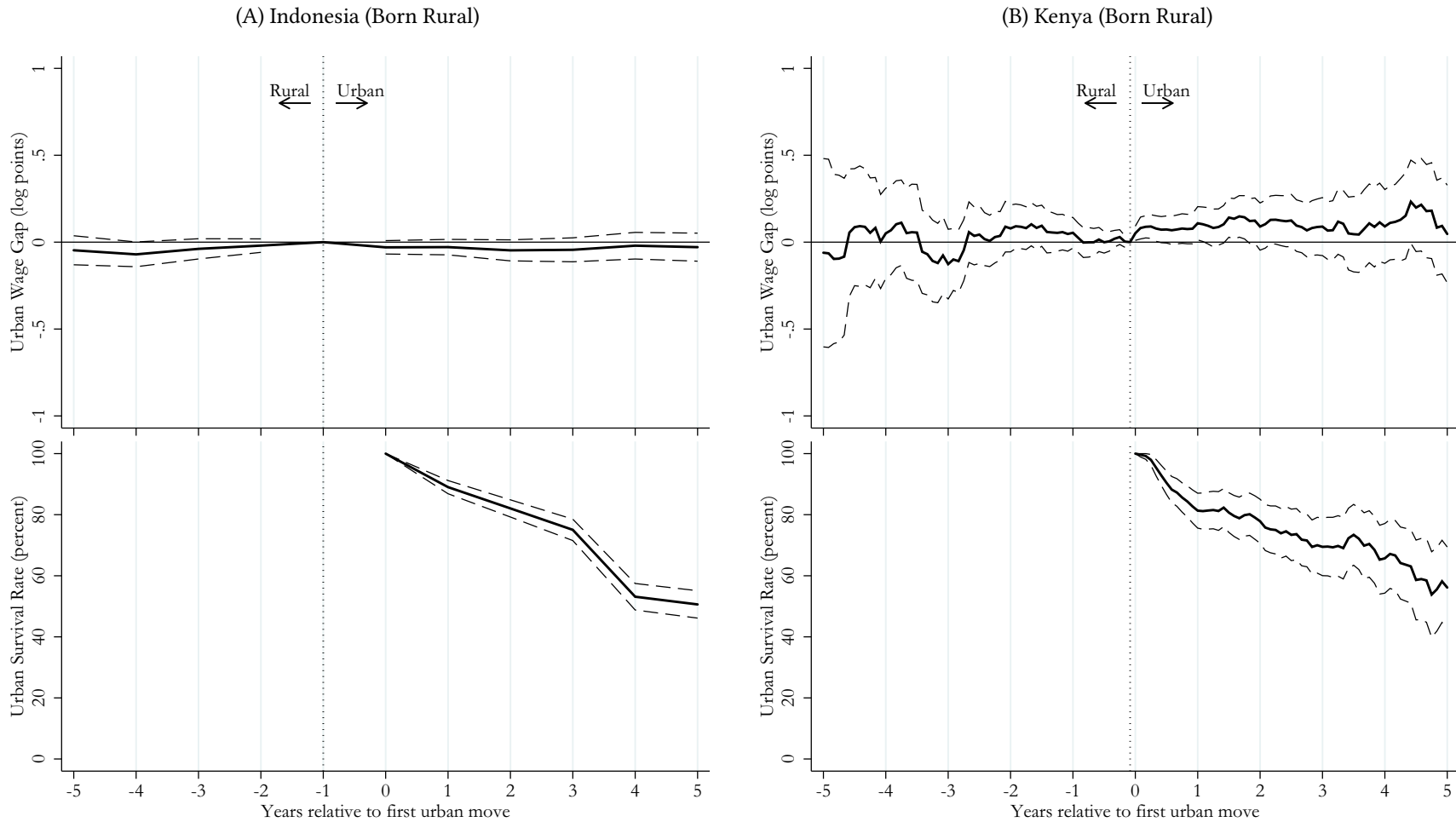


(C) Kenya (Born Rural)



*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 used in Tables 5 and 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 parentheses from 1,000 iterations of block sampling of individuals with replacement. 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. Panel C mimics the format of Panel A except uses data from the KLPS.

Figure 4: Event Study of Urban Migration



*Notes:* Panel A uses data from individuals in the IFLS who are born in rural areas, and Panel B uses data from the KLPS. Please refer to Section 3 for further details on the data. The top half of each panel reports event study coefficients  $\beta_\tau$  from a regression of log wages described in equation 8 (in section 4D). 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. (The estimated fraction of survivors can in principle increase due to sample composition changes as can be seen in the lower half of panel B.) In the IFLS, there are 923 individuals who have observed wages in the year of the move and the year prior; 482 of these individuals report wages 5 years later. In the KLPS, these numbers are 343 and 57, respectively.

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

## (A) Indonesia (Main Analysis Sample)

|                        | Rural   | Urban  | Total   |
|------------------------|---------|--------|---------|
| Agriculture            | 42.6%   | 9.1%   | 30.9%   |
| Non-Agriculture        | 57.4%   | 90.9%  | 69.1%   |
| Number of Observations | 179,756 | 95,844 | 275,600 |

## (B) Kenya (Main Analysis Sample)

|                        | Rural  | Urban  | Total   |
|------------------------|--------|--------|---------|
| Agriculture            | 26.1%  | 5.4%   | 15.2%   |
| Non-Agriculture        | 73.9%  | 94.6%  | 84.8%   |
| Number of Observations | 63,545 | 70,676 | 134,221 |

## (C) Kenya (12 Months with Subsistence Agricultural Module)

|                        | Rural  | Urban  | Total  |
|------------------------|--------|--------|--------|
| Agriculture            | 59.1%  | 9.1%   | 40.6%  |
| Non-Agriculture        | 40.9%  | 90.9%  | 59.4%  |
| Number of Observations | 27,301 | 16,029 | 43,330 |

*Notes:* Panel A reports summary statistics from the Indonesia 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 275,600 individual-year observations, for individuals aged 16 and above for whom earnings measures are available. As described in Section 3, all reported wage employment, self employment, and subsistence agricultural employment is included. Panel B shows the main Kenyan analysis sample of 134,221 individual-month observations of individuals aged 16 and above for whom earnings measures are available. As described in Section 3, wage and self-employment are included for all years, including agricultural wage labor and agricultural self-employment if annual revenues exceeded 40,000 Ksh, and subsistence agricultural data is included from the 12 months preceding the KLPS 3 interview date if the respondent is both the main decision maker and the only supplier of labor in agriculture. Panel C shows data from the 12 months where subsistence agriculture data is available and counts all agricultural activities: including when the person is not the main decision maker and when others work on the agricultural activity; in the case of the latter, the agricultural productivity is weighted by the share of hours that the individual supplies. Each cell reports the percentage of observations by agricultural and non-agricultural sector, and by rural and urban area. In both the IFLS and KLPS, individuals are characterized by the sector of their main employment. The urban indicator from the 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 the KLPS, the urban indicator equals 1 if the person reports living in a large town or city. Please see the text in section 3B for further details on this classification. The list of Kenyan urban areas and frequency of occurrence in the panel are given in Table A4.

Table 2: Summary Statistics

(A) Indonesia

|                 | All<br>N=31843   | Always Rural<br>N=14737 | Rural-to-Urban<br>Migrants<br>N=5287 | Always Urban<br>N=7594 | Urban-to-Rural<br>Migrants<br>N=4218 | Obs   |
|-----------------|------------------|-------------------------|--------------------------------------|------------------------|--------------------------------------|-------|
| Primary Ed.     | 0.865<br>[0.342] | 0.791<br>[0.407]        | 0.939<br>[0.240]                     | 0.957<br>[0.202]       | 0.862<br>[0.344]                     | 31843 |
| Secondary Ed.   | 0.393<br>[0.488] | 0.255<br>[0.436]        | 0.452<br>[0.498]                     | 0.614<br>[0.487]       | 0.400<br>[0.490]                     | 31843 |
| College         | 0.108<br>[0.311] | 0.054<br>[0.226]        | 0.128<br>[0.334]                     | 0.196<br>[0.397]       | 0.116<br>[0.321]                     | 31843 |
| Female          | 0.428<br>[0.495] | 0.420<br>[0.494]        | 0.413<br>[0.492]                     | 0.459<br>[0.498]       | 0.417<br>[0.493]                     | 31843 |
| Raven's Z-score | 0.001<br>[0.925] | -0.143<br>[0.932]       | 0.082<br>[0.904]                     | 0.185<br>[0.881]       | 0.081<br>[0.923]                     | 22899 |

(B) Kenya

|                 | All<br>N=4791    | Always Rural<br>N=1639 | Rural-to-Urban<br>Migrants<br>N=3152 | Always Urban | Urban-to-Rural<br>Migrants | Obs  |
|-----------------|------------------|------------------------|--------------------------------------|--------------|----------------------------|------|
| Primary Ed.     | 0.734<br>[0.442] | 0.637<br>[0.481]       | 0.785<br>[0.411]                     |              |                            | 4791 |
| Secondary Ed.   | 0.353<br>[0.478] | 0.240<br>[0.427]       | 0.412<br>[0.492]                     |              |                            | 4791 |
| College         | 0.035<br>[0.184] | 0.012<br>[0.107]       | 0.047<br>[0.212]                     |              |                            | 4791 |
| Female          | 0.522<br>[0.500] | 0.522<br>[0.500]       | 0.523<br>[0.500]                     |              |                            | 4791 |
| Raven's Z-score | 0.050<br>[0.986] | -0.143<br>[0.982]      | 0.149<br>[0.974]                     |              |                            | 4522 |

*Notes:* Panel A reports summary statistics from the IFLS and panel B reports summary statistics from the 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), “Rural-to-Urban Migrants” are individuals born in rural areas and are observed in urban areas in our sample with data on wages, hours, and sector. “Urban-to-Rural Migrants” are defined similarly. In panel B (Kenya), all individuals are born rural; migrants are those who have subsequent observations with information on income, hours, and sector in urban areas. 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 one.

Table 3: Correlates of Urban Migration

(A) Indonesia (Born Rural)

|                 | (1)                  | (2)                 | (3)                 | (4)                 | (5)                  | (6)                  | (7)                  |
|-----------------|----------------------|---------------------|---------------------|---------------------|----------------------|----------------------|----------------------|
| Primary Ed.     | 0.203***<br>(0.006)  |                     |                     |                     |                      | 0.132***<br>(0.009)  | 0.153***<br>(0.007)  |
| Secondary Ed.   |                      | 0.180***<br>(0.007) |                     |                     |                      | 0.108***<br>(0.009)  | 0.122***<br>(0.008)  |
| College         |                      |                     | 0.211***<br>(0.013) |                     |                      | 0.0695***<br>(0.016) | 0.0921***<br>(0.015) |
| Female          |                      |                     |                     | -0.00565<br>(0.006) |                      | 0.0179*<br>(0.007)   | 0.0169**<br>(0.006)  |
| Raven's Z-score |                      |                     |                     |                     | 0.0509***<br>(0.004) | 0.0252***<br>(0.004) |                      |
| Constant        | 0.0954***<br>(0.005) | 0.209***<br>(0.003) | 0.249***<br>(0.003) | 0.266***<br>(0.004) | 0.271***<br>(0.004)  | 0.101***<br>(0.009)  | 0.0858***<br>(0.006) |
| Observations    | 20024                | 20024               | 20024               | 20024               | 14553                | 14553                | 20024                |

(B) Kenya (Born Rural)

|                 | (1)                 | (2)                 | (3)                 | (4)                 | (5)                  | (6)                  | (7)                 |
|-----------------|---------------------|---------------------|---------------------|---------------------|----------------------|----------------------|---------------------|
| Primary Ed.     | 0.170***<br>(0.016) |                     |                     |                     |                      | 0.0902***<br>(0.019) | 0.111***<br>(0.018) |
| Secondary Ed.   |                     | 0.169***<br>(0.014) |                     |                     |                      | 0.0969***<br>(0.017) | 0.111***<br>(0.016) |
| College         |                     |                     | 0.237***<br>(0.025) |                     |                      | 0.121***<br>(0.027)  | 0.133***<br>(0.027) |
| Female          |                     |                     |                     | 0.000781<br>(0.014) |                      | 0.0168<br>(0.014)    | 0.0136<br>(0.013)   |
| Raven's Z-score |                     |                     |                     |                     | 0.0674***<br>(0.007) | 0.0308***<br>(0.008) |                     |
| Constant        | 0.533***<br>(0.014) | 0.598***<br>(0.009) | 0.650***<br>(0.007) | 0.657***<br>(0.010) | 0.657***<br>(0.007)  | 0.543***<br>(0.017)  | 0.525***<br>(0.016) |
| Observations    | 4791                | 4791                | 4791                | 4791                | 4522                 | 4522                 | 4791                |

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. Both panels are estimated on individuals who are born rural. Columns 6 and 7 report coefficients from multiple regressions with corresponding rows as included covariates. Column 7 omits the Raven's matrix exam to preserve sample size. Robust standard errors reported below in parentheses, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 4: Correlates of Employment in Non-Agriculture

## (A) Indonesia (Born Rural)

|                 | (1)                 | (2)                 | (3)                 | (4)                  | (5)                  | (6)                  | (7)                  |
|-----------------|---------------------|---------------------|---------------------|----------------------|----------------------|----------------------|----------------------|
| Primary Ed.     | 0.289***<br>(0.009) |                     |                     |                      |                      | 0.208***<br>(0.013)  | 0.241***<br>(0.010)  |
| Secondary Ed.   |                     | 0.226***<br>(0.005) |                     |                      |                      | 0.129***<br>(0.007)  | 0.157***<br>(0.006)  |
| College         |                     |                     | 0.225***<br>(0.006) |                      |                      | 0.0481***<br>(0.008) | 0.0622***<br>(0.007) |
| Female          |                     |                     |                     | 0.0533***<br>(0.006) |                      | 0.0801***<br>(0.007) | 0.0872***<br>(0.006) |
| Raven's Z-score |                     |                     |                     |                      | 0.0633***<br>(0.004) | 0.0348***<br>(0.004) |                      |
| Constant        | 0.512***<br>(0.009) | 0.683***<br>(0.004) | 0.736***<br>(0.003) | 0.730***<br>(0.004)  | 0.790***<br>(0.003)  | 0.519***<br>(0.013)  | 0.462***<br>(0.009)  |
| Observations    | 20024               | 20024               | 20024               | 20024                | 14553                | 14553                | 20024                |

## (B) Kenya (Born Rural)

|                 | (1)                 | (2)                 | (3)                  | (4)                 | (5)                  | (6)                  | (7)                  |
|-----------------|---------------------|---------------------|----------------------|---------------------|----------------------|----------------------|----------------------|
| Primary Ed.     | 0.142***<br>(0.012) |                     |                      |                     |                      | 0.106***<br>(0.014)  | 0.117***<br>(0.013)  |
| Secondary Ed.   |                     | 0.101***<br>(0.007) |                      |                     |                      | 0.0419***<br>(0.008) | 0.0523***<br>(0.008) |
| College         |                     |                     | 0.0912***<br>(0.007) |                     |                      | 0.0178*<br>(0.008)   | 0.0262***<br>(0.008) |
| Female          |                     |                     |                      | 0.0209*<br>(0.008)  |                      | 0.0326***<br>(0.009) | 0.0300***<br>(0.008) |
| Raven's Z-score |                     |                     |                      |                     | 0.0442***<br>(0.004) | 0.0205***<br>(0.005) |                      |
| Constant        | 0.802***<br>(0.011) | 0.870***<br>(0.006) | 0.903***<br>(0.004)  | 0.895***<br>(0.006) | 0.905***<br>(0.004)  | 0.794***<br>(0.013)  | 0.785***<br>(0.012)  |
| Observations    | 4791                | 4791                | 4791                 | 4791                | 4522                 | 4522                 | 4791                 |

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 who are born in rural areas, whereas panel B (Kenya) includes the full sample subject to previously defined sample restrictions. Please see Appendix Table A1 for analogous regressions of individuals born urban in Indonesia. Columns 6 and 7 report coefficients from a multiple regression with corresponding rows as included covariates. Column 7 omits the Raven's matrix exam to preserve sample size. Robust standard errors reported below in parentheses, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 5: Non-Agricultural/Agricultural Gap in Earnings

(A) Indonesia

|                             | Dependent variable: Log Earnings |                      |                      |                      |                      |                     |                     |                      |
|-----------------------------|----------------------------------|----------------------|----------------------|----------------------|----------------------|---------------------|---------------------|----------------------|
|                             | (1)                              | (2)                  | (3)                  | (4)                  | (5)                  | (6)                 | (7)<br>Log Wage     | (8)<br>Log Real Wage |
| Non-agricultural employment | 0.703***<br>(0.013)              | 0.572***<br>(0.012)  | 0.351***<br>(0.012)  | 0.349***<br>(0.014)  | 0.249***<br>(0.018)  | 0.252***<br>(0.018) | 0.078***<br>(0.021) | 0.076***<br>(0.021)  |
| Log hours                   |                                  | 0.566***<br>(0.017)  | 0.437***<br>(0.015)  | 0.460***<br>(0.017)  | 0.279***<br>(0.035)  | 0.355***<br>(0.020) |                     |                      |
| Log hours squared           |                                  | -0.022***<br>(0.003) | -0.011***<br>(0.003) | -0.014***<br>(0.003) | 0.018***<br>(0.007)  | -0.006<br>(0.004)   |                     |                      |
| Female                      |                                  |                      | -0.468***<br>(0.011) | -0.465***<br>(0.012) | -0.530***<br>(0.028) |                     |                     |                      |
| Years of education          |                                  |                      | 0.015***<br>(0.004)  | -0.002<br>(0.005)    | 0.028***<br>(0.009)  |                     |                     |                      |
| Years of education squared  |                                  |                      | 0.004***<br>(0.000)  | 0.005***<br>(0.000)  | 0.003***<br>(0.001)  |                     |                     |                      |
| Normalized Ravens           |                                  |                      |                      | 0.065***<br>(0.007)  |                      |                     |                     |                      |
| Normalized Ravens squared   |                                  |                      |                      | 0.014***<br>(0.005)  |                      |                     |                     |                      |
| Individual fixed effects    | N                                | N                    | N                    | N                    | N                    | Y                   | Y                   | Y                    |
| Time fixed effects          | N                                | Y                    | Y                    | Y                    | Y                    | Y                   | Y                   | Y                    |
| Switchers only              |                                  |                      |                      |                      | Y                    |                     |                     |                      |
| Number of observations      | 275600                           | 275600               | 275600               | 201699               | 55802                | 275600              | 275600              | 275600               |
| Number of individuals       | 31843                            | 31843                | 31843                | 22899                | 4208                 | 31843               | 31843               | 31843                |

## (B) Kenya

|                             | Dependent variable: Log Earnings |                     |                      |                      |                      |                     |                  |                      |
|-----------------------------|----------------------------------|---------------------|----------------------|----------------------|----------------------|---------------------|------------------|----------------------|
|                             | (1)                              | (2)                 | (3)                  | (4)                  | (5)                  | (6)                 | (7)<br>Log Wage  | (8)<br>Log Real Wage |
| Non-agricultural employment | 0.784***<br>(0.063)              | 0.548***<br>(0.061) | 0.543***<br>(0.058)  | 0.559***<br>(0.059)  | 0.310***<br>(0.101)  | 0.283***<br>(0.090) | 0.061<br>(0.106) | 0.049<br>(0.106)     |
| Log hours                   |                                  | 0.078<br>(0.196)    | 0.021<br>(0.172)     | 0.071<br>(0.174)     | 0.197<br>(0.377)     | 0.228<br>(0.258)    |                  |                      |
| Log hours squared           |                                  | 0.041*<br>(0.021)   | 0.040**<br>(0.019)   | 0.035*<br>(0.019)    | 0.019<br>(0.045)     | 0.013<br>(0.028)    |                  |                      |
| Female                      |                                  |                     | -0.562***<br>(0.038) | -0.533***<br>(0.039) | -0.708***<br>(0.137) |                     |                  |                      |
| Years of education          |                                  |                     | -0.012<br>(0.038)    | -0.025<br>(0.039)    | -0.041<br>(0.120)    |                     |                  |                      |
| Years of education squared  |                                  |                     | 0.006***<br>(0.002)  | 0.006***<br>(0.002)  | 0.005<br>(0.007)     |                     |                  |                      |
| Normalized Ravens           |                                  |                     |                      | 0.083***<br>(0.023)  | 0.034<br>(0.070)     |                     |                  |                      |
| Normalized Ravens squared   |                                  |                     |                      | -0.046**<br>(0.020)  | -0.153**<br>(0.070)  |                     |                  |                      |
| Individual fixed effects    | N                                | N                   | N                    | N                    | N                    | Y                   | Y                | Y                    |
| Time fixed effects          | N                                | Y                   | Y                    | Y                    | Y                    | Y                   | Y                | Y                    |
| Switchers only              |                                  |                     |                      |                      | Y                    |                     |                  |                      |
| Number of observations      | 134221                           | 134221              | 134221               | 128215               | 14922                | 134221              | 134221           | 134221               |
| Number of individuals       | 4791                             | 4791                | 4791                 | 4522                 | 341                  | 4791                | 4791             | 4791                 |

Notes: Panel A uses data from rounds 1–5 of the Indonesia 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. 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 main employment is in the non-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. Each regression in columns 2–8 include quadratic controls for age. All regressions are clustered at the individual level. Robust standard errors are in parentheses, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.



Table 6: Urban/Rural Gap in Earnings

(A) Indonesia

|                            | Dependent variable: Log Earnings |                      |                      |                      |                      |                     |                  |                      |
|----------------------------|----------------------------------|----------------------|----------------------|----------------------|----------------------|---------------------|------------------|----------------------|
|                            | (1)                              | (2)                  | (3)                  | (4)                  | (5)                  | (6)                 | (7)<br>Log Wage  | (8)<br>Log Real Wage |
| Urban                      | 0.537***<br>(0.012)              | 0.462***<br>(0.011)  | 0.227***<br>(0.010)  | 0.217***<br>(0.012)  | 0.067***<br>(0.015)  | 0.028**<br>(0.013)  | 0.002<br>(0.014) | -0.094***<br>(0.014) |
| Log hours                  |                                  | 0.536***<br>(0.017)  | 0.417***<br>(0.015)  | 0.441***<br>(0.017)  | 0.465***<br>(0.039)  | 0.347***<br>(0.020) |                  |                      |
| Log hours squared          |                                  | -0.012***<br>(0.003) | -0.003<br>(0.003)    | -0.007**<br>(0.003)  | -0.011<br>(0.007)    | -0.003<br>(0.004)   |                  |                      |
| Female                     |                                  |                      | -0.423***<br>(0.011) | -0.419***<br>(0.013) | -0.370***<br>(0.023) |                     |                  |                      |
| Years of education         |                                  |                      | 0.023***<br>(0.004)  | 0.006<br>(0.005)     | 0.030***<br>(0.008)  |                     |                  |                      |
| Years of education squared |                                  |                      | 0.004***<br>(0.000)  | 0.004***<br>(0.000)  | 0.004***<br>(0.000)  |                     |                  |                      |
| Normalized Ravens          |                                  |                      |                      | 0.070***<br>(0.007)  |                      |                     |                  |                      |
| Normalized Ravens squared  |                                  |                      |                      | 0.012**<br>(0.005)   |                      |                     |                  |                      |
| Individual fixed effects   | N                                | N                    | N                    | N                    | N                    | Y                   | Y                | Y                    |
| Time fixed effects         | N                                | Y                    | Y                    | Y                    | Y                    | Y                   | Y                | Y                    |
| Switchers only             |                                  |                      |                      |                      | Y                    |                     |                  |                      |
| Number of observations     | 275600                           | 275600               | 275600               | 201699               | 62944                | 275600              | 275600           | 275600               |
| Number of individuals      | 31843                            | 31843                | 31843                | 22899                | 5086                 | 31843               | 31843            | 31843                |

## (B) Kenya

|                            | Dependent variable: Log Earnings |                     |                      |                      |                      |                     |                     |                      |
|----------------------------|----------------------------------|---------------------|----------------------|----------------------|----------------------|---------------------|---------------------|----------------------|
|                            | (1)                              | (2)                 | (3)                  | (4)                  | (5)                  | (6)                 | (7)<br>Log Wage     | (8)<br>Log Real Wage |
| Urban                      | 0.862***<br>(0.039)              | 0.752***<br>(0.038) | 0.679***<br>(0.035)  | 0.680***<br>(0.036)  | 0.367***<br>(0.055)  | 0.262***<br>(0.047) | 0.165***<br>(0.050) | 0.087*<br>(0.050)    |
| Log hours                  |                                  | 0.106<br>(0.180)    | 0.050<br>(0.163)     | 0.084<br>(0.166)     | 0.506<br>(0.362)     | 0.259<br>(0.262)    |                     |                      |
| Log hours squared          |                                  | 0.033*<br>(0.020)   | 0.034*<br>(0.018)    | 0.031*<br>(0.018)    | -0.009<br>(0.039)    | 0.009<br>(0.029)    |                     |                      |
| Female                     |                                  |                     | -0.514***<br>(0.037) | -0.485***<br>(0.038) | -0.368***<br>(0.069) |                     |                     |                      |
| Years of education         |                                  |                     | -0.025<br>(0.035)    | -0.032<br>(0.037)    | -0.104*<br>(0.060)   |                     |                     |                      |
| Years of education squared |                                  |                     | 0.006***<br>(0.002)  | 0.006***<br>(0.002)  | 0.010***<br>(0.003)  |                     |                     |                      |
| Normalized Ravens          |                                  |                     |                      | 0.072***<br>(0.022)  | 0.078*<br>(0.040)    |                     |                     |                      |
| Normalized Ravens squared  |                                  |                     |                      | -0.027<br>(0.019)    | -0.014<br>(0.033)    |                     |                     |                      |
| Individual fixed effects   | N                                | N                   | N                    | N                    | N                    | Y                   | Y                   | Y                    |
| Time fixed effects         | N                                | Y                   | Y                    | Y                    | Y                    | Y                   | Y                   | Y                    |
| Switchers only             |                                  |                     |                      |                      | Y                    |                     |                     |                      |
| Number of observations     | 134221                           | 134221              | 134221               | 128215               | 39338                | 134221              | 134221              | 134221               |
| Number of individuals      | 4791                             | 4791                | 4791                 | 4522                 | 1037                 | 4791                | 4791                | 4791                 |

*Notes:* Panel A uses data from the IFLS and Panel B uses data from the 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 the 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 the KLPS, the urban indicator equals 1 if a person lives in a large town or a city. Please see the text in section 3B for further details on this classification. Column 5 only includes switchers, who are defined as individuals with at least one observation in both an urban and rural area. Each regression in columns 2–8 include quadratic controls for age. All regressions are clustered at the individual level. Robust standard errors are in parentheses, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

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

(A) Indonesian individuals born in rural areas

|                               | Dependent Variable: Log Earnings |                     |                   | Dependent Variable: Log Consumption |                     |                     |
|-------------------------------|----------------------------------|---------------------|-------------------|-------------------------------------|---------------------|---------------------|
|                               | (1)                              | (2)                 | (3)               | (4)                                 | (5)                 | (6)                 |
|                               |                                  |                     | Log Wage          |                                     |                     |                     |
| Urban                         | 0.635***<br>(0.018)              | 0.296***<br>(0.015) | 0.037*<br>(0.020) | 0.580***<br>(0.018)                 | 0.306***<br>(0.013) | 0.133***<br>(0.020) |
| Individual fixed effects      | N                                | N                   | Y                 | N                                   | N                   | Y                   |
| Control variables and time FE | N                                | Y                   | Y                 | N                                   | Y                   | Y                   |
| Number of observations        | 179158                           | 179158              | 179158            | 53840                               | 53840               | 53840               |
| Number of individuals         | 20010                            | 20010               | 20010             | 22240                               | 22240               | 22240               |

(B) Indonesian individuals born in urban areas

|                               | Dependent Variable: Log Earnings |                     |                   | Dependent Variable: Log Consumption |                     |                      |
|-------------------------------|----------------------------------|---------------------|-------------------|-------------------------------------|---------------------|----------------------|
|                               | (1)                              | (2)                 | (3)               | (4)                                 | (5)                 | (6)                  |
|                               |                                  |                     | Log Wage          |                                     |                     |                      |
| Urban                         | 0.262***<br>(0.020)              | 0.091***<br>(0.016) | -0.024<br>(0.020) | 0.206***<br>(0.020)                 | 0.034***<br>(0.012) | -0.047***<br>(0.015) |
| Individual fixed effects      | N                                | N                   | Y                 | N                                   | N                   | Y                    |
| Control variables and time FE | N                                | Y                   | Y                 | N                                   | Y                   | Y                    |
| Number of observations        | 96249                            | 96249               | 96249             | 28363                               | 28363               | 28363                |
| Number of individuals         | 11784                            | 11784               | 11784             | 12524                               | 12524               | 12524                |

Notes: Columns 1 - 3 of Panels A and B repeat the analyses of Table 6A for those born in rural and urban areas, respectively, and columns 4 - 6 repeat the analyses of Table 9A for those born in rural and urban areas. Please refer to Section 3 for further details on the data and to the notes of Table 6 and 9 for additional information on the variables. Control variables include log hours worked, log hours worked squared, 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 log hours worked, log hours worked squared, and age squared because the others are absorbed by the individual fixed effects. When using log wage as the dependent variable in column 3, only age squared is included as a control variable. All regressions are clustered at the individual level. Robust standard errors are in parentheses, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 8: Robustness to Alternative Agricultural Productivity Measures

(A) Indonesia

| Definition of Agriculture               | Productivity Measure Includes... |                       | Dependent variable: |
|---|----------------------------------|-----------------------|---------------------|
|   | Formal Wages                     | Self-Employed Profits | Log Wage            |
| <b>Majority of hours in agriculture</b> |                                  |                       |                     |
| <b>Main Estimation</b>                  | ✓                                | ✓                     | 0.078***<br>(0.021) |
| Any hours in agriculture                | ✓                                | ✓                     | 0.013<br>(0.018)    |
| All hours in agriculture                | ✓                                | ✓                     | 0.123***<br>(0.020) |
| Majority of hours in agriculture        | ✓                                |                       | 0.021<br>(0.026)    |
| Majority of hours in agriculture        |                                  | ✓                     | 0.084***<br>(0.032) |

(B) Kenya

| Definition of Agriculture               | Productivity Measure Includes... |                       | Dependent variable: |
|---|----------------------------------|-----------------------|---------------------|
|   | Formal Wages                     | Self-Employed Profits | Log Wage            |
| <b>Majority of hours in agriculture</b> |                                  |                       |                     |
| <b>Main Estimation</b>                  | ✓                                | ✓                     | 0.061<br>(0.106)    |
| Any hours in agriculture                | ✓                                | ✓                     | 0.096<br>(0.097)    |
| All hours in agriculture                | ✓                                | ✓                     | 0.064<br>(0.108)    |
| Majority of hours in agriculture        | ✓                                |                       | 0.157<br>(0.121)    |
| Majority of hours in agriculture        |                                  | ✓                     | 0.002<br>(0.187)    |

*Notes:* Panel A uses data from the IFLS and Panel B uses data from the KLPS. Each row shows the robustness results of a regression of log wages (calculated as earnings per hour) on a non-agricultural indicator, age squared, and time and individual fixed effects. In each panel, the estimate in row 1 can be found in Table 5, column 7; row 2 can be found in Table A5, column 4; row 3 in Table A5, column 8; row 4 in Table A8, column 4; and row 5 in Table A9, column 4. All regressions report standard errors clustered at the individual level. Robust standard errors are in parentheses, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 9: Gaps in Consumption

(A) Indonesia

|                               | Dependent variable: Log Consumption |                     |                     |                     |                     |                    |
|-------------------------------|-------------------------------------|---------------------|---------------------|---------------------|---------------------|--------------------|
|                               | (1)                                 | (2)                 | (3)                 | (4)                 | (5)                 | (6)                |
| Non-agricultural employment   | 0.636***<br>(0.011)                 | 0.250***<br>(0.008) | 0.126***<br>(0.012) |                     |                     |                    |
| Urban                         |                                     |                     |                     | 0.403***<br>(0.011) | 0.139***<br>(0.008) | 0.026**<br>(0.012) |
| Individual fixed effects      | N                                   | N                   | Y                   | N                   | N                   | Y                  |
| Control variables and time FE | N                                   | Y                   | Y                   | N                   | Y                   | Y                  |
| Number of observations        | 82272                               | 82272               | 82272               | 82272               | 82272               | 82272              |
| Number of individuals         | 34820                               | 34820               | 34820               | 34820               | 34820               | 34820              |

(B) Kenya

|                               | Dependent variable: Log Meals Eaten |                     |                    |                     |                     |                   |
|-------------------------------|-------------------------------------|---------------------|--------------------|---------------------|---------------------|-------------------|
|                               | (1)                                 | (2)                 | (3)                | (4)                 | (5)                 | (6)               |
| Non-agricultural employment   | 0.078***<br>(0.016)                 | 0.059***<br>(0.017) | -0.090*<br>(0.049) |                     |                     |                   |
| Urban                         |                                     |                     |                    | 0.029***<br>(0.010) | 0.030***<br>(0.011) | -0.023<br>(0.040) |
| Individual fixed effects      | N                                   | N                   | Y                  | N                   | N                   | Y                 |
| Control variables and time FE | N                                   | Y                   | Y                  | N                   | Y                   | Y                 |
| Number of observations        | 4203                                | 4203                | 4203               | 4203                | 4203                | 4203              |
| Number of individuals         | 3601                                | 3601                | 3601               | 3601                | 3601                | 3601              |

*Notes:* Panel A uses data on total consumption from the IFLS, and Panel B uses data on meals eaten in the last day from the KLPS. Unlike previous tables, the sample includes individuals with and without earnings measures. Consumption data in the IFLS are obtained by adding up the value of food and non-food consumption in Indonesian Rupiah at the household level and dividing this by the number of household members. The data was collected for each of the five waves so each household has five observations at most. Separate analyses by food and non-food consumption in Indonesia can be found in Appendix Table A11, and Appendix Table A12 provides consumption analyses when using the sample with positive earnings measures. Data on meals eaten in Kenya are available from KLPS rounds 2 and 3 and refer to the day prior to the survey date. In the analysis sample, 0.6% of individual-time observations ate no meals in the prior day, 10.9% ate one meal, 53.2% ate two meals, 34.0% ate three meals, and 1.3% ate four or more. Control variables in both panels 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 parentheses, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 10: Urban/Rural Gap in Wages for Top 5 Cities

(A) Indonesia

|                                   | Dependent variable: Log Wages |                     |                     |                   |
|-----------------------------------|-------------------------------|---------------------|---------------------|-------------------|
|                                   | (1)                           | (2)                 | (3)                 | (4)               |
| Urban                             | 0.356***<br>(0.012)           | 0.299***<br>(0.012) | 0.081***<br>(0.011) | 0.002<br>(0.015)  |
| Jakarta (population 10 million)   |                               | 0.269***<br>(0.020) | 0.243***<br>(0.018) | -0.033<br>(0.034) |
| Surabaya (population 2.8 million) |                               | 0.017<br>(0.058)    | -0.004<br>(0.047)   | 0.096<br>(0.094)  |
| Bandung (population 2.6 million)  |                               | 0.262***<br>(0.065) | 0.110**<br>(0.054)  | 0.229*<br>(0.125) |
| Medan (population 2.5 million)    |                               | 0.303***<br>(0.047) | 0.251***<br>(0.044) | -0.069<br>(0.104) |
| Bekasi (population 2.5 million)   |                               | 0.628***<br>(0.063) | 0.426***<br>(0.056) | 0.112<br>(0.080)  |
| Individual fixed effects          | N                             | N                   | N                   | Y                 |
| Control variables and time FE     | N                             | N                   | Y                   | Y                 |
| Number of observations            | 275600                        | 275600              | 275600              | 275600            |
| Number of individuals             | 31843                         | 31843               | 31843               | 31843             |

(B) Kenya

|                                  | Dependent variable: Log Wages |                     |                     |                     |
|----------------------------------|-------------------------------|---------------------|---------------------|---------------------|
|                                  | (1)                           | (2)                 | (3)                 | (4)                 |
| Urban                            | 0.574***<br>(0.040)           | 0.373***<br>(0.060) | 0.320***<br>(0.055) | 0.071<br>(0.063)    |
| Nairobi (population 3.4 million) |                               | 0.324***<br>(0.063) | 0.303***<br>(0.057) | 0.156***<br>(0.060) |
| Mombasa (population 1.2 million) |                               | 0.267***<br>(0.079) | 0.265***<br>(0.075) | 0.307***<br>(0.094) |
| Kisumu (population 0.4 million)  |                               | 0.014<br>(0.153)    | 0.090<br>(0.142)    | 0.099<br>(0.244)    |
| Nakuru (population 0.3 million)  |                               | 0.252**<br>(0.118)  | 0.156*<br>(0.094)   | 0.136<br>(0.155)    |
| Eldoret (population 0.3 million) |                               | 0.148<br>(0.163)    | 0.105<br>(0.163)    | -0.078<br>(0.139)   |
| Individual fixed effects         | N                             | N                   | N                   | Y                   |
| Control variables and time FE    | N                             | N                   | Y                   | Y                   |
| Number of observations           | 134221                        | 134221              | 134221              | 134221              |
| Number of individuals            | 4791                          | 4791                | 4791                | 4791                |

*Notes:* Panel A uses data from the IFLS and Panel B uses data from the 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, and 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 parentheses, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.