

Inequality, the Urban-Rural Gap and Migration *

Abstract

Using population and product consumption data from the Demographic and Health Surveys I construct comparable measures of inequality and migration for 65 countries, including some of the poorest countries in the world. I find that the urban-rural gap accounts for 40% of mean country inequality and much of its cross-country variation. One out of every four or five individuals raised in rural areas moves to urban areas as a young adult, where they earn much higher incomes than non-migrant rural permanent residents. Equally, one out of every four or five individuals raised in urban areas moves to rural areas as a young adult, where they earn much lower incomes than their non-migrant urban cousins. These flows and relative incomes are suggestive of a world where the population sorts itself geographically on the basis of its human capital and skill. I show that a simple model of this sort explains the urban-rural gap in living standards.

Alwyn Young
Department of Economics
London School of Economics

*I would like to thank Elhanan Helpman, Ho Veng-Si and anonymous referees for many helpful comments and Measure DHS (www.measuredhs.com) for making their data publicly available.

I. Introduction

Inequality, whether within countries or across countries, can be construed as coming from either differences in the ownership of quantities of factors of production or from differences in earnings per unit ownership of factors of production. These competing perspectives have dramatically different implications for academic research and public policy. Differences in the ownership of factors focus attention on the distribution of resources and the incentives for accumulation. In contrast, differences in earnings per factor suggest the existence of “wedges”, unexploited differences that must reflect barriers or market failures of some sort and whose removal could give rise to low cost welfare improvements. In performing decompositions of the causes of inequality, the analyst, after removing observable differences in the ownership of factors, is always left with a residual. The temptation is to treat this residual variation as reflecting wedges. Since the residual variation usually accounts for a large share of total inequality, this interpretation has profound implications. In this paper I present evidence that the large residual gaps between urban and rural living standards in developing countries, while accounting for much of the inequality within those countries, reflect selection based upon unobserved skill and human capital, i.e. unobserved ownership of factors of production.

I use data in 170 Demographic and Health Surveys (DHS) for 65 countries to develop a new set of internationally comparable measures of inequality and migration. My technique involves using Engel curves estimated off of household educational attainment combined with household consumption random effects to calculate the components of consumption inequality in educational equivalent units. I separately estimate the contribution to overall inequality of inequality within urban and rural areas and the urban-rural gap in living standards, as well as decomposing total inequality into educational and residual (net of education) inequality. The

DHS disproportionately samples the poorest (principally sub-Saharan) countries of the world, but also includes observations on lower middle income and middle income countries in Africa, Latin America, South and Southeast Asia and Eastern Europe, all within a standardized survey framework. Thus, my approach produces comparable estimates for a sample covering the earliest stages of development up through the lower-middle range of the world income distribution.

I find that the urban-rural gap in living standards is a major source of inequality, accounting for 40% of average inequality and for much of the cross-country variation in levels of inequality. Countries with unusually high levels of inequality are countries where the urban rural gap is unusually large. There is no significant correlation between the magnitude of the urban-rural gap in living standards and urbanization or GDP per capita. While the gaps between mean urban and rural living standards are dramatic, the average levels of consumption dispersion or inequality within urban and rural areas are about equal. I find inequality in educational attainment to be a comparatively minor source of inequality, on average accounting for only 19% of total inequality and explaining very little of its cross-country variation.

The DHS collects data on individuals' childhood and current place of residence, allowing for a detailed analysis of the number, characteristics and consumption of migrants. About one out of every four or five individuals raised in rural areas migrates to urban areas as a young adult. Surprisingly, it is also true that one out of every four or five individuals raised in urban areas migrates to rural areas as a young adult. Rural to urban migrants are typically better educated than rural permanent residents and urban to rural migrants are typically less educated than urban permanent residents. I find that migrants enjoy consumption levels which, corrected for educational attainment, are quite close to those of permanent residents in their destination region. With residual urban-rural consumption gaps equivalent, on average, to the earnings from 9 years

of education (or 1.16 in log money metric terms), this translates into seemingly large “gains” for rural to urban migrants and equally large “losses” for urban to rural migrants. While a model of migration as an attempt to exploit urban-rural differences in factor returns can motivate the observed rural to urban movement of labour, it cannot explain the large flow of urban residents to rural areas where they receive vastly lower earnings per unit of education than their non-migrant cousins.

I develop a model of integrated factor markets with sorting on the basis of unobservable skill and show it matches features of the DHS data. Production uses both skilled and unskilled workers and urban industries are more skill intensive, i.e. have a higher relative demand for skilled workers. Observable education and unobservable skill are imperfectly correlated. While education increases the probability an individual acquires skill, the relation is in no way deterministic. Due to the higher relative demand for skill, in equilibrium workers observed in urban areas are more likely to have skill than comparably educated workers in rural areas. This produces a residual, educationally adjusted, gap in urban-rural living standards. Since education is positively correlated with skill, better educated rural workers are more likely to move to urban areas, while less educated urban workers are more likely to move to rural areas.

The model produces empirical predictions that are supported by the data. The urban-rural gap is produced solely by the relative skill intensity of production in urban and rural areas. Since the probability of acquiring skill is increasing in the educational attainment of workers, the urban/rural residence probabilities of highly and poorly educated workers provide proxy information on the regional skill intensity of production. I show that these residence probabilities completely explain average urban-rural differences, leaving no room for the constant or any other variable for that matter. In other words, when the urban residence probabilities of highly and

poorly educated workers are about the same, reflecting very little difference in the skill intensity of production in urban and rural areas, the urban-rural gap completely disappears.

This paper draws upon a rich literature in a number of areas. Methodologically, my consumption estimation methods build on the work of Filmer and Pritchett (2001), who were the first to suggest the use of product consumption data to calculate inequality in surveys where the requisite total expenditure data are unavailable. They proposed to use a principal components analysis of the variation in the ownership of products or household conditions to construct measures of relative household wealth, an approach that has since been widely implemented by DHS programmers.¹ The principal components approach is, however, unit-less and devoid of economic content, as the consumption measures are standardized by their mean and standard deviation. Consequently, the inequality calculated in one survey cannot be compared to another, or to conventional measures of inequality. By using the correlation of consumption with educational attainment I produce inequality measures that are motivated by Engel curves and demand theory, weight products by their correlation with an observable determinant of relative incomes, and are internationally comparable. Comparing my results with other studies for my sample countries, I show that my methods produce estimates of country inequality and urban-rural gaps in living standards that are quite consistent with those calculated using conventional methods.

The simultaneous existence in the poorest economies of the world of two sectors, urban and rural (or nearly equivalently, agricultural and non-agricultural), producing vastly different average living standards, has attracted the attention of economists since the Second World War.

¹While Filmer and Pritchett and the DHS use the terminology “wealth,” I prefer to use the word “consumption”, as most of the measures they include (such as the ownership of consumer durables and housing conditions) provide a flow value of consumption. A proper measure of wealth would include assets which do not yield direct consumption services, such as financial instruments, physical capital and land. Theory, of course, suggests that consumption should vary with overall wealth, so practically the distinction may be largely moot.

Early efforts focused on labour market distortions. Thus, Lewis (1954), in a paper that sparked an entire literature on “dual” economies, argued that workers in rural areas, in deciding to migrate to urban areas, compared their average product in rural family output (which they shared) with their marginal product in urban output, producing a situation with excess and surplus rural labour. More recent analysts, however, have looked for competitive explanations. Thus, for example, Gollin, Parente and Rogerson (2004) argue that rural areas offer greater opportunities for (unmeasured) home production, while Lagakos and Waugh (2011) argue that workers sort themselves into urban and rural areas based upon their intrinsic abilities and comparative advantage.

The model presented in this paper is similar to Lagakos and Waugh (2011) and owes much to their insightful paper. Lagakos and Waugh posit that workers are endowed with productivity draws in agricultural and non-agricultural activities and, by assuming that the productivity draws are positively correlated and non-agricultural productivity draws have higher variance, produce a situation where workers selecting into non-agricultural industry have higher productivity than those selecting into agricultural industry. This paper follows their emphasis on urban-rural sorting, but motivates urban-rural living standard differences by appealing to unobserved skill which is correlated with educational attainment. The correlation with educational attainment allows me to test the model using the residence characteristics of highly and poorly educated households.

The results of this paper complement those of recent cross-national empirical studies of urban and rural industry. Caselli (2005) and Restuccia, Yang and Zhu (2008), using PWT and FAO data, document that the ratio of non-agricultural to agricultural productivity in fixed international prices falls with GDP per capita. In the DHS I find that the urban-rural gap in real

consumption has no relation to GDP per capita once one introduces a dummy for sub-Saharan Africa (where the gap appears to be somewhat larger). I show that in Caselli's data the relationship between relative non-agricultural to agricultural productivity and GDP is largely a rich country phenomenon and in poorer countries disappears with the addition of a sub-Saharan dummy. Thus, although there is difference in the measures, productivity vs. real consumption, there is basic agreement that urban-rural gaps are not correlated with GDP amongst poorer countries.

Gollin, Lagakos & Waugh (2012), working carefully with census sources and 10 household expenditure surveys, show that large gaps between agricultural and non-agricultural average value products in local prices remain even after careful consideration of sectoral differences in human capital and hours worked. This study is similar to theirs in its use of household surveys and focus on living standards. My estimates adjust for human capital, but lack the detail of their exploration of the factors behind the residual differences. At this price, I gather a larger international sample and relate the urban-rural gap to overall income inequality and the relative consumption of migrants and permanent residents. Finally, I should note that Herrendorf and Schoellman (2012) argue that much of the reported difference in agricultural/non-agricultural value added per worker in the United States is a consequence of mismeasurement of agricultural value added in the national accounts. This paper does not rely on national accounts assessments, but instead uses direct measures of the consumption of urban and rural households in LDCs.

The paper proceeds as follows. Section II provides a short description of the DHS product consumption and migration data, while Section III outlines the inequality estimation methodology. Section IV shows that my methods, applied to the DHS data, produce estimates of Gini coefficients and urban-rural gaps that are broadly consistent with other sources. I also show

that the urban-rural gap plays a major role and educational inequality a comparatively minor role in explaining both the mean level of inequality and its cross-country variation. Finally, I document the absence of a relationship between the urban-rural consumption gap and overall urbanization or GDP per capita and show that migrants, by and large, enjoy education adjusted consumption levels that are close to those of permanent residents in their destination region. Section V presents the model of correlated unobservable skill and observable educational attainment and applies it to the DHS data. In contrast with overall urbanization and GDP per capita, I document the sharp significance of the urban residence probabilities of individuals with low and high educational attainment in explaining the residual (education adjusted) urban-rural consumption gap. Section VI concludes.

II. Demographic and Health Survey Data

The Demographic Health Survey and its predecessor the World Fertility Survey, both supported by the U.S. Agency for International Development, have conducted irregular but in-depth household level surveys of fertility and health in developing countries since the late-1970s. Over time the questions and topics in the surveys have evolved and their coverage has changed, with household and adult male question modules added to a central female module, whose coverage, in turn, has expanded from ever married women to all adult women. I make use of all DHS associated surveys that are freely available (i.e. do not require the permission of national authorities), have household member educational attainment data (as this is used in all of my estimation equations), and include data on either (a) at least four of my measures of durable goods or housing consumption, (b) migration, or (c) individual wages. In all, I make use of 170 surveys covering 2.1 million households in 65 developing countries since 1990, as listed in

Appendix A. The occasional nature of the DHS surveys means that I have an unbalanced panel with fairly erratic dates.

The raw data files of the DHS surveys are distributed as standardized "recode" files. Unfortunately, this standardization and recoding has been performed, over the years, by different individuals using diverse methodologies and making their own idiosyncratic errors. This produces senseless variation across surveys as, to cite two examples, individuals with the same educational attainment are coded as having dramatically different years of education or individuals who were not asked education attendance questions are coded, in some surveys only, as not attending. In addition, there are underlying differences in the coverage of the surveys (e.g. children less than 5 years vs. children less than 3 years) and the phrasing and number of questions on particular topics (e.g. employment) which produce further variation. Working with the original questionnaires and supplementary raw data generously provided by DHS programmers, I have recoded all of the individual educational attainment data, corrected coding errors in some individual items, recoded variables to standardized definitions and, as necessary, restricted the coverage to a consistent sample (e.g. married women, children less than 3 years) and removed surveys with inconsistent question formats (in particular, regarding labour force participation). Appendix A lists the details.

I use the DHS data to derive 23 measures of real consumption distributed across four areas: (1) ownership of durables; (2) housing conditions; (3) household time and family economics; and (4) children's health. Table I details the individual variables and sample means. All of these variables are related to household demand and expenditure, broadly construed and I have found them (Young 2012) to be very significantly correlated with household real incomes, as proxied by adult educational attainment. I have selected these variables on the basis of their

availability and with an eye to providing a sampling of consumption expenditures that would, through material durables, household time and health (which is related to nutrition), cover much of the budget of households in the developing world. My list of consumption “goods” includes negative outcomes, such as diarrhea, but this is accounted for in my estimation procedure (described further below) which uses the correlation of consumption with rises in household educational attainment, e.g. the absence of diarrhea, as the metric in the calculation of an educational equivalent household random effect for consumption of all goods.²

By including health and family economics, I follow Becker, Philipson and Soares (2005) and Jones and Klenow (2011) and take a broader view of consumption than is typically used in the national accounts. However, this does not drive my results. First, to keep my estimates grounded in traditional measures, I only make use of the 167 surveys which have data on at least four measures of durable goods or housing consumption. Second, the health and family economics variables are individually coded. Because of the large idiosyncratic individual variation in these products within households, they tend not to dominate the estimation of household random effects. Finally, household level inequality in these outcomes moves with household level inequality in the consumption of durables and housing, so the inclusion of these non-traditional measures of consumption ultimately lowers standard errors without much influence on point estimates. Thus, on the theoretical grounds that health and family economics are equally part of household consumption and on the practical grounds that larger samples are always preferable, I use these non-traditional products to supplement my traditional measures of

²Relative to Young (2012), which analyzed growth in living standards, I drop three continuous measures of household consumption, namely ln rooms per capita and the ln height and weight of young children. The likelihood for the normal regression model is intrinsically much more concave than the likelihood for the discrete choice logit model used to represent the household decision to consume the remaining 0/1 dichotomous variables. Thus, these variables, when included in a common random effects specification, dominate the estimates. This problem was less acute in Young (2012), where I estimated each equation separately to calculate a weighted average of growth rates.

consumption, while showing the reader that they do not influence the results.

In estimation, I drop a product from a survey's sample if it is present in more than .99 or less than .01 of either urban or rural households as, practically speaking, calculation of random effects is both meaningless and computationally problematic when virtually everyone or no one consumes a product. I have also made the decision to break measures of household time into different age groups to account for different demand patterns at different ages as the possibilities for substitution between home production, human capital accumulation and market labour evolve. Thus, for example, in richer households young women are more likely to be in school and less likely to be working in the late schooling years (ages 15-24), but, consequently, are more likely to be working as young adults (ages 25-49). Although males are included in the schooling and children's health variables, I do not include separate time allocation measures for adult males because male questionnaire modules are much less consistently available and male participation behavior, when recorded, is less strongly related to household education and, hence, by my methodology, would play little role in estimating inequality.

Turning to migration, the DHS contains two questions that provide information on the migrant status of household members. First, adult men and women are often asked, in their interview modules, what type of region (i.e. capital, other city, town or countryside) they lived in prior to the age of 12. Second, adult men and women are often asked if they have always lived in the current locale and, if not, when they moved there and from what type of region (again, capital, other city, town or countryside).³ Using the first of these questions, I classify individuals

³Both types of questions occasionally allow for an individual's earlier residence to be "abroad". In such cases, I treat these as unknown and drop them from the analysis. I compare across survey years for a given country and make sure that where foreign origin accounts for a significant share of the population it is considered in all surveys. Thus, I drop the Jordan 1990 survey, as it does not allow for foreign origin but a large segment of the population lists themselves as international migrants in other survey years. Aside from this, I have also corrected a number of coding errors that have arisen in translating and standardizing original questionnaire files.

as rural (countryside) to urban (capital, city or town) migrants or urban to rural migrants if there is a discrepancy between the characteristics of their childhood locale and their current residence,⁴ and as lifelong urban or rural residents if there is not. Using the second of these questions, I classify individuals who claim to have always lived in their locale as lifelong urban or rural residents and otherwise, based upon their region prior to moving in, as rural to urban, rural to rural, urban to urban or urban to rural migrants. I focus on individuals aged 25 to 49 as the women's questionnaire only covers women 15 to 49 and most early non-African surveys only cover ever-married women. As most women over 25 are ever married (Table I), focusing on individuals aged 25 to 49 produces consistent and comparable samples. I focus on the same age group for the male data which, in any case, are available much less frequently.

Table II presents the basic characteristics of the DHS migrant data. As shown in the table, rural to urban migrants (row 6), as defined by childhood residence, account for an average of 34.3 to 37.8 % of adult urban members in the sample countries. When measured using information on whether the respondent has recently moved, this ratio falls to 22.1 to 23.9 %. This reflects the possibility of multiple moves and the emphasis on recent moving in the latter measure, as many currently urban individuals who claim to have been in rural areas before the age of 12 (the first measure) indicate that they recently moved in from another urban area (the second measure). Of greater significance, because of the lower consumption levels in rural areas described later on, is the fact that by any measure about 13 to 17 % of rural residents are urban to rural migrants (row 3). Measured in terms of the population in their originating area, an average of 22.1 to 22.5 % of individuals who lived in rural areas prior to the age of 12 reside in urban

⁴In all DHS surveys the authorities executing the survey report whether the *current* household residence is urban or rural and in many surveys they also indicate whether it is the capital, other city, a town or countryside (both as defined by national statistical authorities). I only use surveys where both measures are reported and the agreement between the two (following my categorization of countryside as rural and capital, city and town as urban) is greater than 99%.

areas, while 23.0 to 27.7 % of individuals who lived in urban areas prior to the age of 12 reside in rural areas as adults. Thus, as a fraction of the originating population, urban to rural migration is at least on par with rural to urban migration.⁵ This same pattern is repeated when the individual's origin is measured by the last locale they lived in, with rural to urban migrants representing 15.1 to 16.7 % of the population originating in rural areas, but urban to rural migrants representing 20.2 to 26.3 % of the population originating in urban areas.

With regards to educational attainment, it is apparent in Table II that rural to urban migration draws from the better educated part of the rural population, as the typical rural to urban migrant is much better educated than the typical permanent rural resident (although less educated than urban permanent residents). Conversely, urban to rural migration draws from the less educated part of the urban population, as the typical urban to rural migrant is less educated than urban permanent residents (although better educated than rural permanent residents). These differences in educational attainment are not due to migrants arriving young and completing their education in the destination area. As shown in the table, the typical migrant over 25 is about 35 and arrived 10 to 13 years earlier, i.e. in their early to mid 20s, when their education was long completed.⁶ While rural to rural migrants share the educational attainment of rural permanent residents, urban to urban migrants appear to be slightly better educated than urban permanent residents.

In preparing Table II, I have taken pains to remove countries and surveys from the sample that might produce misleading or spurious results. Thus, internal conflict in many developing

⁵The difference relative to shares of destinations arises because of the smaller average urban population share (.41 vs. .59 for rural). Overall, net migration is in favour of urban areas with, on average, .126 of the aggregate young adult female population moving to urban areas and only .070 to rural areas

⁶Of 118,000 male or female urban to rural and rural to urban migrants between the ages of 25 to 49 in the sample, only 14,000 claim to have arrived prior to the age of 15.

countries has forced individuals to abandon their homes and move to other regions. To focus on voluntary migration alone, Table II excludes any country in the DHS sample which is reported by the Norwegian Refugee Council as having more than 1% of the population internally displaced by conflict or violence in any year between 2001 and 2010 or reported by the UNHCR as having more than 1% of its population as protected or assisted internally displaced persons in any year between 1993 and 2010.⁷ Some of the excluded countries which have experienced large internal displacements show very large urban to rural migrations (e.g. 52% of the women originating in urban areas in Cambodia), but by and large this effect is quite small for the typical country. When calculated using the entire sample for which I have migration data, the shares of originating populations involved in rural to urban or urban to rural migration (e.g. 22.0 and 25.1 %, respectively, for the 51 countries with women's migration data) are quite close to those reported with the restricted sample in the panels of Table II above.

The skeptical reader might still wonder whether measurement error at the level of the individual respondent is responsible for the results reported in Table II. If respondents don't quite know whether they grew up in urban or rural areas, the discrepancy between their current residence and their random responses will produce the appearance of large bi-directional population flows. There are three reasons why this is unlikely to be true. First, as noted above, I do not code an individual's migrant status on the basis of a response to a vague question about rural or urban origin, which many individuals might find difficult to understand, but on the basis of a response to a question which lists capital, other city, town or countryside as alternatives. These terms are much easier for respondents to interpret. Second, as Table II shows, there are

⁷The choice of years is given by the availability of data on internally displaced persons from the two sources (Norwegian Refugee Council, "Internal displacement caused by conflict and violence: Estimated numbers of internally displaced people from 2001 to 2010", www.internal-displacement.org and UNHCR Statistical Online Population Database, "IDPs protected/assisted by UNHCR," 1993-2010, www.unhcr.org), but this time period corresponds to that covered by my DHS data. I take country population numbers from Penn World Tables 7.0.

large systematic educational differences between respondent types that are intuitively consistent with the differing schooling opportunities in their place of origin, with rural to urban migrants being less educated than urban permanent residents and urban to rural migrants being better educated than rural permanent residents.

The most important evidence against a preponderance of measurement error lies in the fact that the DHS data are consistent with the well known tendency of migrants to live close to each other. Table III reports logits of an individual's migrant status (1 = yes, 0 = no) based upon current vs childhood place of residence on the average migrant status of other adults outside of their household in their survey cluster (MigMean)⁸ and a complete set of survey fixed effects (dummies). As shown, the average migrant status of an individual's neighbours is both statistically and quantitatively an incredibly significant determinant of the probability they themselves are migrants. In urban areas, .287 of the sample women are migrants. The data predict that in the absence of any migrant neighbours that fraction would be .123, while if surrounded by migrants it would rise to .774. In rural areas, where urban to rural migrants are to be found, the movement from the complete absence to the complete presence of migrant neighbours moves the probability thirteen-fold, from .061 to .818. In the relatively limited male data, the effects are smaller but still dramatic, with the predicted probability an individual is an urban to rural migrant rising six-fold, from .076 to .429, as the characteristics of their neighbours move from one extreme to another. If migrant status were heavily determined by measurement error in an individual's report of their childhood residence, an individual's migrant status would be largely uncorrelated with that of their neighbours. This is clearly not the case in the DHS data, where migrants, following the worldwide pattern, generally live close to other migrants.

⁸Thus, if 75% of adults outside of an individual's household in their cluster are migrants, MigMean is .75 for that individual.

There is measurement error in all surveys, and there is little doubt that, by creating discrepancies between an individual's current locale and their self-reported childhood or earlier residence, this exaggerates true population flows. However, the wording of the DHS questionnaire, the substantial differences between the educational characteristics of migrants and non-migrants within regions, and the tendency of self-reported migrants to cluster close to other self-reported migrants, all suggest that the population movements recorded in the DHS are by and large a genuine feature of the data. The voluntary movement of young adults from poor rural to rich urban regions is a well-known feature of life in less developed countries. The voluntary movement of young adults from rich urban areas to poor rural areas appears to be a characteristic of life in these countries as well.

III. Methods: Product Sampling and the Measurement of Inequality

Let real consumption in household h of group g in country c at time t be given by:

$$(1) \ln(C_{hgt}^R) = \ln(C_{gt}^{R-E}) + R_E^{ct} E_{hgt} + u_{hgt}$$

where E_{hgt} is household adult educational attainment, R_E^{ct} the educational profile of \ln consumption in the country (motivated, say, by the relationship between earnings and education), $\ln(C_{gt}^{R-E})$ mean group \ln consumption at zero educational attainment, and u_{hgt} a mean zero orthogonal error term. Consumption inequality within a country is driven by two factors: (1) educational inequality; and (2) "residual inequality" as determined by the variance of u and inter-group variation in consumption net of education $\ln(C^{R-E})$. It will be convenient to define residual inequality in units of equivalent educational inequality. Thus, one can think of u_{hgt} as having a group standard deviation $R_E^{ct} \sigma_{gt}(u)$ and $\ln(C_{gt}^{R-E})$ as being given by $R_E^{ct} D_{gt}$, so that $\sigma_{gt}(u)$ and D_{gt} represent the standard deviation and mean of group residual (net of education)

consumption measured in units of education. Groups can be defined at any level that allows consistent and comparable aggregation and comparison over time. In my main analysis I divide the population on the basis of their residence into two groups, urban or rural, but later on I differentiate groups by both their location and migrant status.

Say the real demand for product p is given by:

$$(2) \ln(Q_{phgt}) = \alpha_{pct} + \eta_{pct} \ln(C_{hgt}^R) + \vec{\beta}'_{pct} \vec{X}_{hgt} + \varepsilon_{phgt}$$

where the α_{pct} are constants, η_{pct} the slopes of quasi-Engel curves, \vec{X}_{hgt} and $\vec{\beta}_{pct}$ vectors of demographic characteristics and their associated coefficients, and ε_{phgt} a product x household error term driven by variation in preferences and local prices. I use the term quasi in describing the consumption elasticities η_{pct} because $\ln(Q_{phgt})$ need not be actual \ln quantity demanded but only some measure related to that quantity, such as the index in a probability model. Variation in local conditions may result in country x time variation in the level of product consumption (α) and its responsiveness to real consumption and household demographic characteristics (η and β), as indicated by the “ct” subscripts on these terms in the equation.

Consider next the use of survey micro data for country c at time t to simultaneously estimate a set of product demand equations of the form:

$$(3) \ln(Q_{phgt}) = a_{pct} + b_{pct} (E_{hgt} + d_{gt} + u_{hgt}) + \vec{c}'_{pct} \vec{X}_{hgt} + e_{phgt}$$

where, on the right hand side, data on household educational attainment and demographic characteristics (E_{hgt} and \vec{X}_{hgt})⁹ identify the coefficients b_{pct} and \vec{c}_{pct} , the co-movement of the consumption levels of all products across regions and within households identifies the group

⁹I use the mean educational attainment of household members aged 25 to 65 for E . As for the demographic controls (X), these are: (a) household durables and housing conditions: \ln number of household members; (b) youths' school attendance: individual's age, age² and sex; (c) women's work, fertility, and marital status: individual's age and age²; (d) infants' diarrhea, fever and cough: individual's sex, $\ln(1+\text{age in months})$ and $\ln(1+\text{age in months})^2$; (e) infants' survival: individual's sex and $\ln(1+\text{age in months})$.

dummies and household random effects d_{gt} and u_{hgt} , and the a_{pct} and e_{phgt} are product means and residual error, respectively. Estimation is carried out survey by survey to calculate the degree of inequality implied by consumption patterns in each survey and also allow for country x time variation in the coefficients, but all product equations within a survey are estimated simultaneously so as to identify the group dummies and random effects. One of the group dummies must be set equal to zero (as the base), and in my main analysis I take that to be the rural population of the country. Using the subscripts U and R to denote the urban and rural groups, from (1) and (2) above it can be seen that asymptotically the coefficient estimates converge to:

$$(4) \quad \hat{b}_{pct} = \eta_{pct} R_E^{ct} \quad \hat{c}_{pct} = \bar{\beta}_{pct} \quad \hat{a}_{pct} = \alpha_{pct} + \eta_{pct} \ln(C_{Rt}^{R \sim E})$$

$$\hat{\sigma}_{U_t}(u) = \sigma_{U_t}(u) \quad \hat{\sigma}_{R_t}(u) = \sigma_{R_t}(u) \quad \hat{d}_{U_t} = D_{U_t} - D_{R_t}$$

where, as noted earlier, group differences in mean ln consumption at zero education and residual consumption variation are measured in education equivalent units.

At the theoretical level, the model described above is an extension of the product sampling techniques used to measure real consumption levels in Young (2012) to the measurement of inequality. While the earlier paper provides additional intuitive and technical detail, a few comments can highlight some of the approach's strengths and weaknesses. Real consumption depends upon nominal expenditure divided by prices. Gathering data on both of these can be challenging, particularly as there may be substantial variation in local prices that is difficult to measure. Thus, the implicit cost of attaining a given household standard of living depends not merely on nominal prices, but also on factors such as government provided infrastructure, the proximity of markets for goods, labour and education, and the disease

environment, factors which probably vary considerably across locales in developing countries. As the real consumption of individual products is related, through the Engel curve, to total real consumption, a sampling of real product consumption provides a direct measure of total real consumption, bypassing the need to measure nominal expenditure and (explicit and implicit) nominal prices. This is the strength of the approach.

The weakness of my method lies in its dependence on product sampling. Any sample is subject to sampling error. Differences in relative prices induce substitution between products. Depending upon how the relative price effects for a given sample of products are correlated with total real consumption differences across households and locales, the sample might exaggerate or underestimate inequality. Relative price effects raise the consumption of one product at the expense of another, so their expected value in a sample of products can legitimately be described as zero.¹⁰ Nevertheless, in a small finite sample of products the possibility of bias always exists and, at a minimum, there is always random error variation induced by the random sample. With regards to bias, a particular concern, given this paper's emphasis on the role of urban-rural gap, might be that the DHS products exaggerate urban-rural differences.¹¹ To address this, in the pages below I compare my estimates with conventional data sources and show that there does not appear to be any particular exaggeration in my measures of urban-rural differences relative to my measures of aggregate inequality. Recognizing that there is substantial error variation, not least because the sample of available products varies somewhat survey by survey, I focus on

¹⁰Anything that raises consumption of all products in a particular area is most appropriately described as a rise in total real consumption.

¹¹Thus, one might argue that publicly provided infrastructure lowers the relative price of electricity and electrical appliances in urban areas. One might respond, however, that a lower frequency of diarrhea, fevers and coughs is easier (cheaper) to attain in less densely populated rural areas and that the demand for transport vehicles such as bicycles is increased by the lack of alternative public transport. The issue in this debate does not lie in whether or not it is desirable to account for the explicit and implicit differences in the prices associated with various outcomes in urban and rural areas, but whether there is a systematic bias in my sample that exaggerates (or understates) urban-rural differences relative to, say, within region differences.

systematic patterns in the entire sample without placing undue credence in the estimates for individual countries.

In its empirical implementation, the model described in equation (3) is a simple extension of Butler and Moffitt's (1982) discrete choice random effects model to include constraints (the b_{pct}) on the influence of dummy variables and random effects on the consumption of each good.¹² These constraints produce an implicit weighting by the tightness of the relation between product consumption and household educational attainment. In products where the correlation between household consumption and educational attainment is very high, the estimation of the quasi-income elasticity b_{pct} is driven by this relation and the estimated group dummies and random effects ($D_i, \sigma_i(u)$), in trying to match cross-group and within group differences in average consumption levels, conform to it. In products where the correlation between household consumption and educational attainment is weak, the likelihood is very flat in b_{pct} , and the estimated value of b_{pct} is made to conform to the estimated group dummies and random effects. To illustrate this, I have estimated each of the 3286 product x survey consumption combinations appearing in my estimation of (3) as a separate standard discrete choice model on mean household educational attainment, a country constant, and the demographic variables used in (3). Regressing the absolute difference between the quasi-elasticities b_{pct} estimated in these equations with the b_{pct} estimated in the joint product likelihoods of (3) on the standard error of the individually estimated b_{pct} , I get a coefficient (s.e.) of .787 (.133). Thus, in moving from the individual product equation to the joint model, b_{pct} moves less where the individual product relation between education and consumption is strong (standard error is small) and, consequently,

¹²Thus, I use maximum likelihood techniques, modeling the household random effects as normally distributed and using 20-point Gauss-Hermite quadrature to integrate the joint logit probability distribution of each household's consumption bundle.

b_{pct} in these circumstances has a stronger influence in determining the magnitude of the group dummies and random effects.

All of the coefficient estimates described above are in units of equivalent education. To convert these into money metric units, one needs estimates of R_E , the proportionate change in monetary consumption associated with an additional year of education. Under the assumption that savings rates do not differ systematically by educational attainment, I estimate R_E using data on individual labour income present in 27 DHS surveys for 14 sub-Saharan and 11 non sub-Saharan countries. I run Mincerian regressions of the ln income of individuals aged 25 to 65 working for others on their years of educational attainment, sex, age and age squared, with cluster fixed effects, and in separate regressions arrive at estimates of .113 (.003) for sub-Saharan Africa and .087 (.002) for the non-African countries. There is likely to be considerable measurement error in individual attainment and, indeed, when I instrument the worker's educational attainment with the educational attainment of other household members the point estimates rise to .139 (.009) for sub-Saharan Africa and .104 (.005) for the non-African countries.¹³ These results compare favourably with those of other studies. Thus, for example, Psacharopolous (1994) in his oft-cited survey of Mincerian regressions finds an average marginal return of .134 in 7 sub-Saharan Africa and .107 in 37 non-OECD non sub-Saharan countries.

In what follows, where an estimate of R_E is necessary I extrapolate the coefficient estimates of the 27 survey Mincerian IV regressions described above to the entire DHS sample, using an R_E of .139 for sub-Saharan Africa and .104 for the non-African countries. Thus, I use this estimate of R_E in the next section to show that my method produces Gini coefficients and

¹³The OLS sample sizes are 8041 sub-Saharan individuals, 15513 non-African individuals, while the corresponding IV sample sizes (reduced by the need to find individuals in households with other adult family members) are 5897 and 13054. Further details on these regressions are provided in Young (2012), so I do not reproduce them here. Relative to that earlier paper, the only change is that I've added data on Jordan to the non-African sample, but this has virtually no effect on the point estimates.

urban-rural gaps that are similar to those found in other sources. However, the strength of the DHS lies in its consistent cross-national data on product consumption, not in its limited information on labour earnings. Consequently, in presenting the main results of the paper I focus on relative magnitudes, decompositions or regressions where R_E plays no role, as it simply scales all components equally or is incorporated into the regional fixed effects of a ln regression, so the analysis is just as easily executed in terms of years of equivalent education.

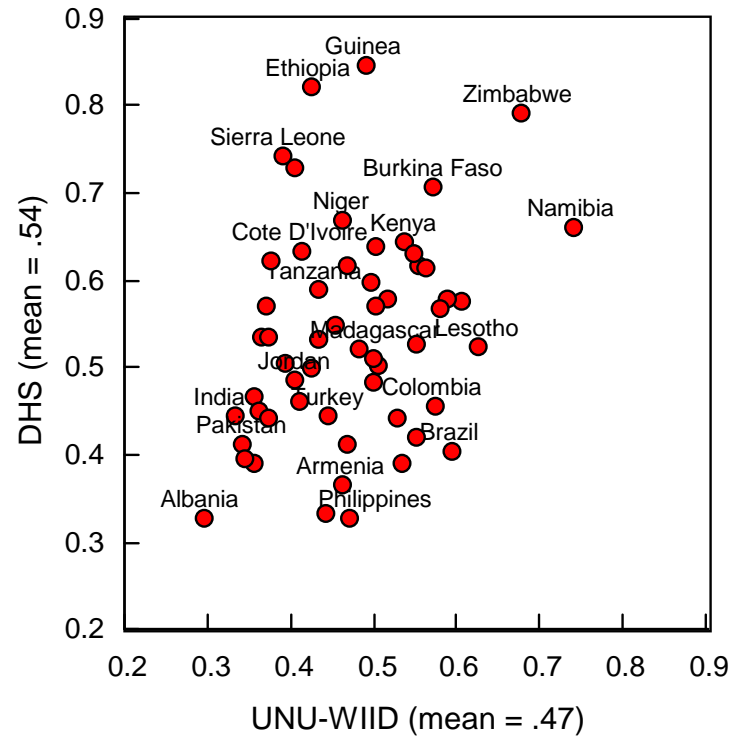
IV. Results: Patterns of Inequality

(a) A Comparison of Ginis & Urban-Rural Gaps

I begin by establishing that my methodology produces reasonable estimates of aggregate inequality and urban-rural gaps. From the United Nations University World Income Inequality Database I take the country average of all estimates of the Gini coefficient since 1990 for my sample countries. There are 437 such observations, of which only 12 are deemed grade A quality by the UNU, for 56 of my 65 countries. Most of these estimates come from World Bank sources. Against these I graph, in Figure I, the country average Gini implied by my DHS estimates, using the R_{ES} of .139 for sub-Saharan Africa and .104 for the remainder of the sample estimated using the IV Mincerian regressions of individual labour income described above.¹⁴ I find that the two sets of Ginis have a modest correlation of .304, with strong disagreement on countries such as Guinea and Ethiopia counterbalanced by basic agreement on the high levels of inequality in Zimbabwe and Namibia and the comparatively equitable distributions of countries such as Albania and Pakistan. The average Gini of .54 found in my DHS estimates is somewhat higher than the .47 found in the typical micro-data based calculation. I note, for the purposes of

¹⁴As, under the assumptions of the estimation framework, residual consumption in each group is distributed ln normally, I use the formula for the Gini coefficient for a mixture of ln-normals developed in Young (2011).

Figure I: UNU-WIID & DHS Ginis
(country averages)



$\rho = .304$

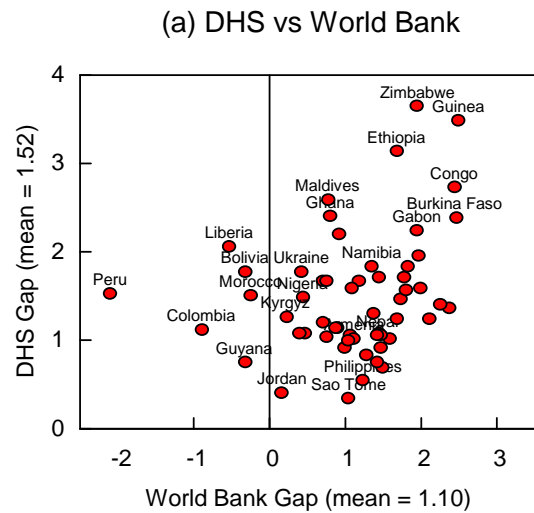
discussion further below, that multiplying my estimates of R_E by .85 (i.e. reducing them by 15%) brings the mean DHS Gini coefficient in line with the mean UNU-WIID observation.

Figure II compares my estimates of the urban-rural gap with macro data reported by international agencies. In Figure II (a) I graph the money metric values of the DHS urban-rural gaps (calculated using the baseline R_E s mentioned above and including the contribution of both residual and educational differences), against the average 1990-2010 ln relative GDP per worker of non-agricultural to agricultural workers in my sample economies, as reported in the World Bank's Databank. The two measures use distinct concepts (relative real consumption per household vs. relative nominal output per worker) and sectoral definitions (urban/rural vs. agricultural/non-agricultural),¹⁵ but are nevertheless modestly correlated. However, the mean DHS ln urban/rural consumption gap of 1.52 is 42 percent higher than the 1.10 difference in non-agricultural/agricultural output per worker suggested by World Bank GDP data. In panel (b), however, I substitute the FAO's measures of the agricultural/non-agricultural economically active population for the World Bank's numbers, while continuing to use the World Bank GDP figures. The average macro estimate of the sectoral gap rises to 1.27 and the correlation with the DHS data improves. In the DHS, the ratio of total household members to individuals of working age (between 16 and 65) is 14 percent higher in rural areas. This suggests that an adjustment from per worker to per capita might easily add .14 to the FAO and World Bank numbers in Figure II. Multiplying my estimates of R_E by .93 and .82 eliminates the remaining difference with the FAO and World Bank means, respectively.

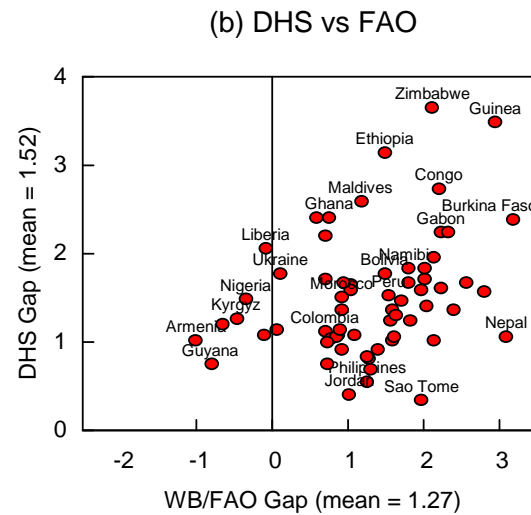
In this paper I emphasize the contribution of the urban-rural gap to aggregate within country inequality. Figures I and II show that my estimates of inequality are somewhat higher

¹⁵The DHS do not report industry of employment, while the ISIC codes dividing GDP and employment are based upon industry rather than urban/rural residence, so it is not possible to make the two sources completely comparable.

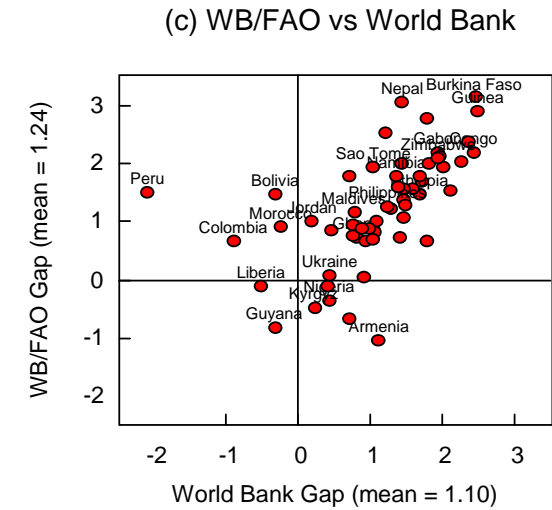
Figure II: Urban-Rural Gaps Estimated Using Alternative Sources (country averages)



rho = .275



rho = .314



rho = .568

than conventional sources, but this discrepancy is not concentrated in my measures of urban-rural differences. The reduction in R_E needed to bring my estimates of urban-rural differences in line with alternative sources is smaller or equivalent to that needed to equalize my measures of aggregate inequality, which include my estimates of within region inequality, with conventional data. This suggests that there is no relative bias in my measurement of the urban-rural gap relative to within region sources of inequality.

In evaluating their differences with the DHS based estimates, it is perhaps worth recognizing that the discrepancies between and uncertainty within the conventional data sources listed in Figures I and II is nothing short of staggering. As shown in panel (c) of Figure II, the correlation between the World Bank and World Bank/FAO estimates is surprisingly weak, given that the two sets of data share the same numerators (the World Bank estimate of sectoral GDP) and differ only in the denominators (the sectoral population). As regards within uncertainty, regressing the country x year observations of each source on country dummies, one gets an estimated observation standard error of .711 for the World Bank data (466 observations in 56 countries) and .315 for the WB/FAO data (1231 observations in 64 countries). These standard errors imply, for example, that the 95% confidence interval for the typical World Bank ln observation is ± 1.39 , i.e. the true (anti-ln) value is somewhere between $\frac{1}{4}$ and 4 times the reported value. Similarly, regressing the 437 UNU Gini's on 56 country dummies, one gets an estimated observation standard error of .064, which is quite large for a Gini coefficient. As for my DHS estimates, to their estimated standard errors¹⁶ must be added the inaccuracies created by imposing, due to a lack of country level data, two values of R_E on the entire sample.

¹⁶Regressing the 167 DHS survey urban-rural gaps and Gini coefficients on country dummies produces observation standard errors of .220 and 036, respectively. This, however, understates the true standard error as it does not take into account the unmeasured correlation across observations brought about by the use of a common sample of products and estimates of R_E .

There is fairly little and often highly inaccurate and inconsistent data regarding income inequality and urban-rural differences in the less developed countries which constitute my sample. The figures above show that my methods produce estimates that are broadly consistent with alternative sources and do not have any obvious relative bias in the measurement of urban-rural differences relative to within region differences. These money metric comparisons, however, rely upon estimates of R_E based upon the data of a small subset of DHS surveys, as the collection of data on labour income has never been a central objective of the DHS. To finesse this obvious weakness, for the remainder of the paper I focus on educational equivalent measures where I am able to make use of the information present in most of the DHS surveys to explore patterns of inequality within countries.

(b) Basic Characteristics of the Estimates

Table IV summarizes the DHS-based estimates of the components of inequality, in years of equivalent education. The top panel provides the mean estimate of each component across the 167 surveys, its standard deviation, and its mean estimated standard error, while the bottom panel provides the same statistics for the country means (across all available surveys for each country) of each component. While the standard errors of the estimates are substantial, they are small relative to the observed coefficient variation across surveys and countries. There is a great deal of persistence in the measures for individual countries, as shown by the fact that the standard deviation of the mean country coefficient estimates is generally slightly larger than the standard deviation of the survey estimates.¹⁷

The DHS data suggest that on average urban residual inequality is slightly lower than rural residual inequality, while urban educational inequality is substantially higher. Across the

¹⁷Since on average there are 2.6 surveys per country, if the variation across surveys was all iid the standard deviation of the country means would be about .62 of the standard deviation of the survey coefficients.

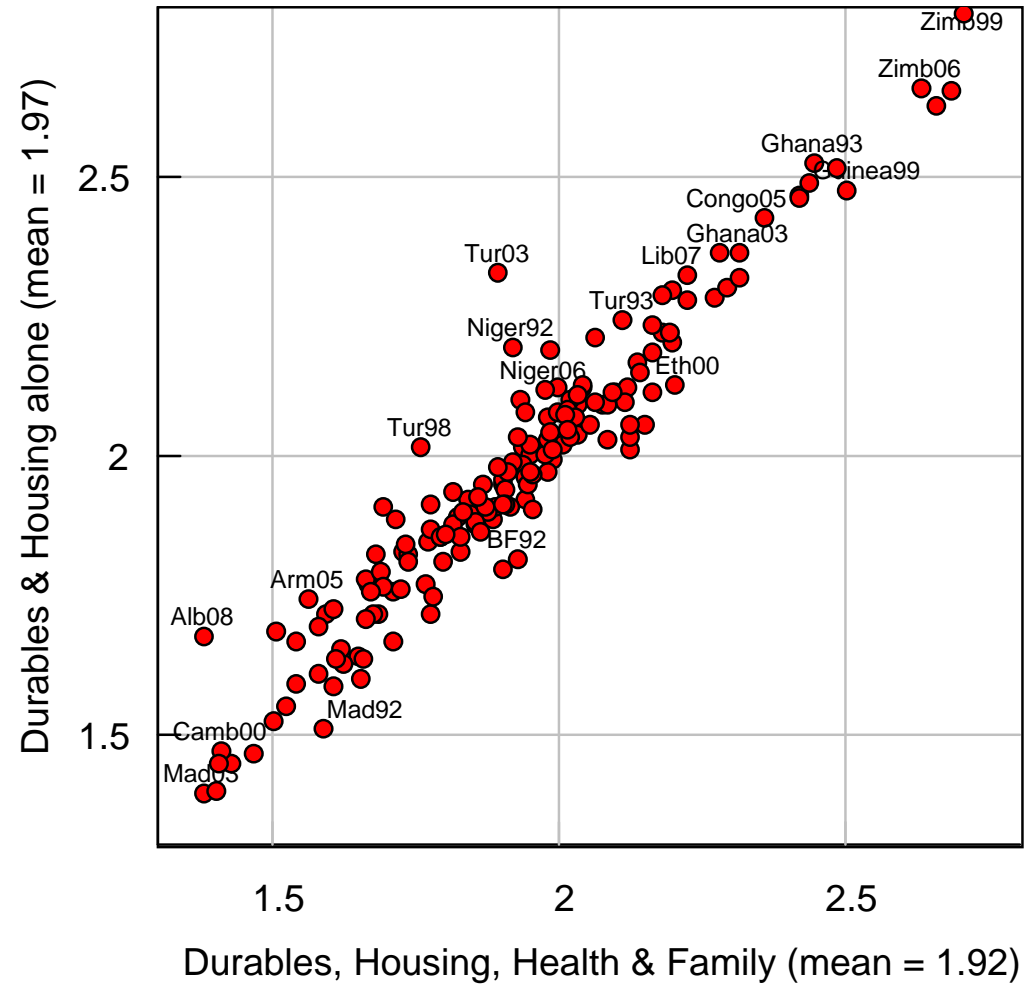
167 surveys, the standard deviation of total within region inequality, equal to the square root of the sum of the residual and educational variances, averages 6.88 for urban areas and 6.55 for rural areas, with a correlation of .692. Within region inequality pales next to the urban-rural gap, which averages more than 12 years of equivalent education, composed of 3 years of educational inequality and more than 9 years of residual inequality. The variation in the urban-rural gap across countries is also much greater than the variation in the other components of inequality. The greater magnitude and variation of the urban-rural gap means that it accounts for most of within country inequality and most of the variation across countries in inequality, as shown further below.

As noted earlier, I follow Becker, Philipson & Soares (2005) and Jones & Klenow (2011) in taking a broader view of consumption than the usual expenditure survey by including measures of health and family economics. None of the results reported in this paper depend upon this broader concept of consumption. Figure III graphs the estimated ln standard deviation of residual inequality, i.e. total inequality in the absence of any educational inequality,¹⁸ measured using only durables and housing against that measured using all available products. The correlation across the 167 survey observations is .961 and the mean measures are only 5 percent apart. Idiosyncratic variation in individual outcomes tends to make the estimates of household random effects in health and family economics less precise,¹⁹ but because these products are broadly consistent with durables and housing they lower the average standard error of the ln standard deviation of residual inequality by 4.6 percent, without much influence on the coefficient point

¹⁸The formula for this is given shortly below. Educational inequality is always the same, regardless of the products used in the analysis, as it is based upon the survey population characteristics.

¹⁹As discrete choice models assume a given variance for the underlying latent variable, additional idiosyncratic variation results in a proportional downward scaling of all of the coefficient estimates. In my model this does not influence the estimates of the components of inequality, which are measured relative to the coefficient on educational attainment (i.e. scaling keeps these ratios intact). However, it does raise the standard error of the components as the residual variation rises relative to model variation.

Figure III: In Standard Deviation of Residual Inequality (yrs of equivalent education - 167 survey observations)



rho = .961

estimates. In sum, on the theoretical grounds that health and the allocation of family time are, along with durables and housing, all equally part of a household's consumption decision, and on the practical grounds that the larger sample lowers estimation error without offending skeptical readers by unduly influencing the average results, I include these products in my measures of consumption inequality.

(c) Decompositions

The model estimated in this paper allows for some interesting decompositions of the aggregate variance of ln consumption, namely:

$$\begin{aligned}
(5) \quad \sigma^2 &= R_E^2 \left[\underbrace{S_U(\sigma_U^2(E) + \sigma_U^2(u))}_{\text{urban inequality}} + \underbrace{S_R(\sigma_R^2(E) + \sigma_R^2(u))}_{\text{rural inequality}} + \underbrace{S_U S_R (E_U + D_U - E_R - D_R)^2}_{\text{urban-rural gap}} \right] \\
&= R_E^2 \left[\underbrace{S_U \sigma_U^2(E) + S_R \sigma_R^2(E)}_{\text{within region educational}} + \underbrace{S_U \sigma_U^2(u) + S_R \sigma_R^2(u)}_{\text{within region residual}} + \underbrace{S_U S_R (E_U + D_U - E_R - D_R)^2}_{\text{urban-rural gap}} \right] \\
&= R_E^2 \left[\underbrace{S_U \sigma_U^2(E) + S_R \sigma_R^2(E) + S_U S_R (E_U - E_R)^2}_{\text{educational inequality} = \sigma^2(E)} + \underbrace{S_U \sigma_U^2(u) + S_R \sigma_R^2(u) + S_U S_R (D_U - D_R)^2}_{\text{residual inequality} = \sigma^2(u)} \right. \\
&\quad \left. + \underbrace{2S_U S_R (E_U - E_R)(D_U - D_R)}_{\text{interaction}} \right]
\end{aligned}$$

where S_U and $S_R = 1 - S_U$ are the urban and rural population shares, and E_i , D_i , $\sigma_i(E)$, and $\sigma_i(u)$ the corresponding mean regional educational attainment, educational equivalent mean ln consumption at zero education, standard deviation of educational attainment, and educational equivalent standard deviation of residual regional household consumption, respectively.²⁰ The first line of (5) decomposes inequality into the contribution of urban inequality, rural inequality and the urban-rural gap. Since I find that overall urban and rural inequality are generally quite

²⁰For notational simplicity, I drop the "ct" notation, but each of these variables is calculated separately for each country x time period (i.e. survey) combination.

similar, the second line rearranges terms to focus on the separate contribution of within region educational and residual inequality. Finally, the third line decomposes total inequality into the inequality due to education in the absence of any residual inequality $\sigma^2(E)$, the converse $\sigma^2(u)$, and the interaction between mean urban and rural differences in education and residual consumption levels. Because R_E enters each line of the equation multiplicatively, the percentage contribution of each of these elements to overall inequality can be analyzed independently of R_E . This is not true for other measures of inequality, such as the Gini coefficient or population quantiles, and is the reason why I use this measure for the decompositions further below.

Figure IV graphs the country average of the share of the variance of ln consumption coming from residual and educational inequality. On average, educational inequality accounts for only 19% of total inequality, while residual inequality accounts for 66%, with the interaction between the two (not drawn) making up the remainder. As shown, the share of residual inequality is higher and that of educational inequality lower in more unequal countries.²¹ These results stem from the greater mean value and cross-country variation of the residual components of inequality ($\sigma_U(u)$, $\sigma_R(u)$, D_U-D_R) relative to the educational components of inequality ($\sigma_U(E)$, $\sigma_R(E)$, E_U-E_R), as shown earlier in Table IV.

Figure V graphs the country averages of the shares of the variance of ln consumption attributable to urban-rural differences (both educational and residual), within region residual inequality and within region educational inequality. On average, urban-rural differences account for 40% of total inequality, and within region residual accounts for 43%, while within region educational only accounts for 17%. The share of urban-rural differences is much higher in more unequal countries, while the share of within region residual and educational inequality is

²¹The share of the interaction term has no particular association with the overall level of inequality.

Figure IV: Shares of Variance of Consumption
(65 countries)

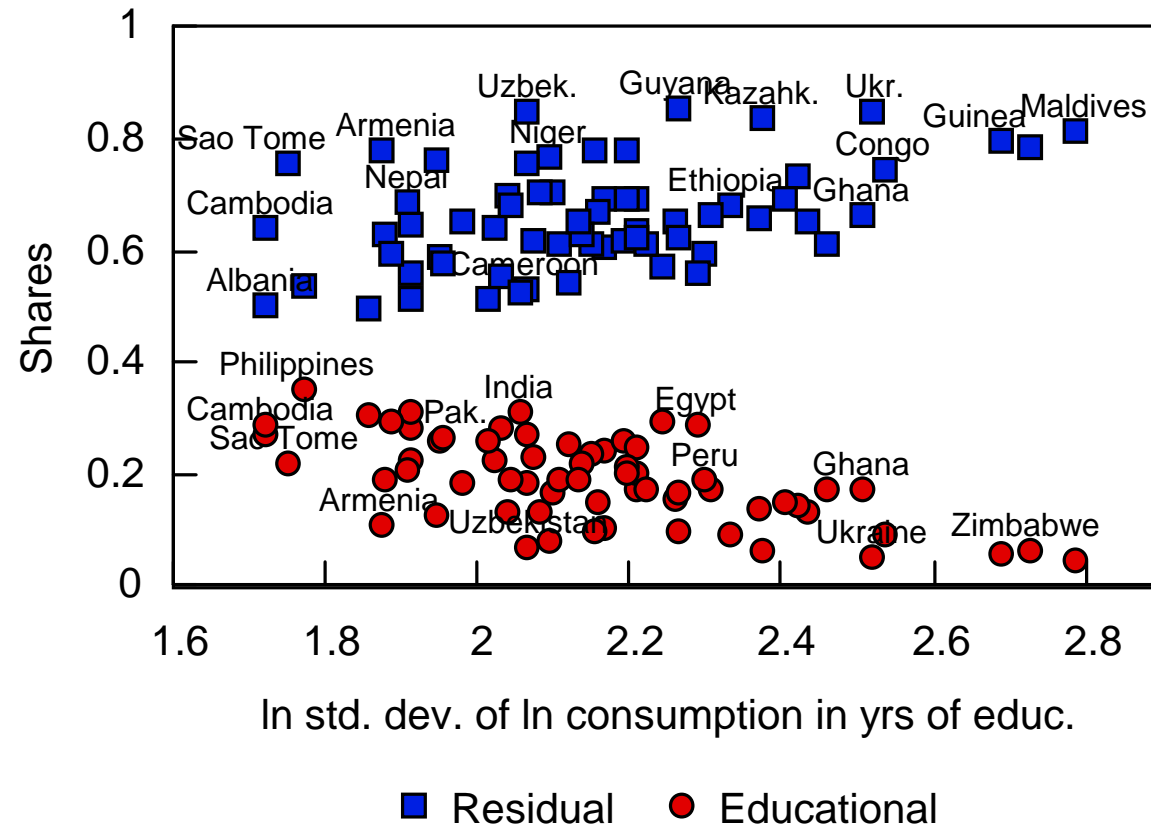
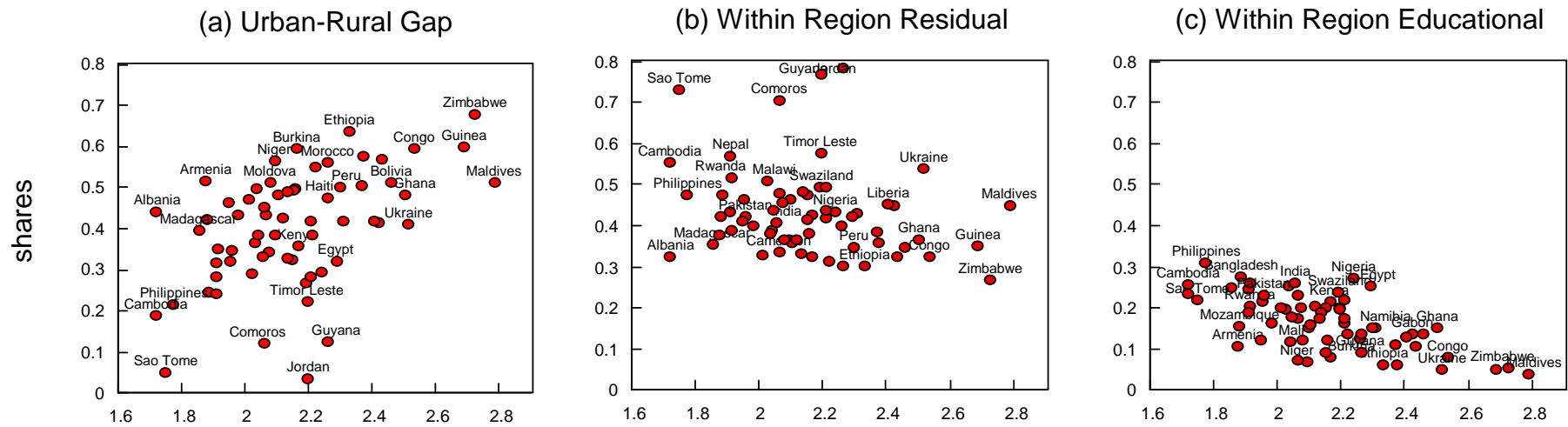


Figure V: Shares of Variance of Consumption (65 countries)



In std. dev. of ln consumption in yrs of educ.

correspondingly lower. This result stems from the large magnitude and cross-country variation of the urban-rural gap, as described earlier in Table IV.

Table V provides a quantitative re-expression of the patterns observed in Figures IV and V, focusing on the standard deviation of ln consumption under various scenarios. In the first line, we see that the average *within country* standard deviation of ln consumption measured in years of education is 8.85. The standard deviation of this measure *across countries* is 2.22. Absent any educational inequality, the average standard deviation of within country inequality would fall by 1.6 years, to 7.25, and its cross-country variation would be largely unchanged. However, absent residual inequality, that is if the only source of inequality were educational, the standard deviation of within country inequality would fall by more than 5 yrs and its cross country standard deviation would fall to less than 1/3 of its original value.²² As shown in the table, the removal of all urban-rural differences, maintaining within region residual and educational inequality, would on average lower the standard deviation of within country ln consumption by 2 years to 6.69 and reduce its cross-country variation by almost a year. The elimination of residual urban-rural differences clearly plays a much bigger role in this reduction than the elimination of urban-rural educational differences. The elimination of within (urban/rural) region residual differences brings the average within country standard deviation of ln consumption down to 6.67, on par with the impact of the elimination of urban rural differences, but it has a much smaller effect on the cross-national variation in this measure, which only falls to 1.96 years. Thus, as indicated earlier in Figure V, on average urban-rural inequality and within region residual inequality account for about the same share of total within country inequality, but urban-rural

²²These decompositions are not additive for two reasons. First, as shown earlier in (5), there is an interaction between residual and educational inequality in determining the total variance of ln consumption. Second, in order to provide a slightly different take on the data, in this table I focus on the average standard deviation of ln consumption, rather than the component shares of the variance of ln consumption presented in Figures IV and V.

differences are a much bigger determinant of variation across countries in overall inequality. Finally, as shown in the bottom line of the table, the elimination of within region educational differences reduces the average standard deviation of ln consumption by less than a year, without much impact on its cross-country variation.

(d) Correlations with Urbanization and GDP

This section explores the correlations, in levels and differences, between the components of inequality and urbanization and GDP. The first six columns of the top row of Table VI below list the estimated components of inequality in units of equivalent education: urban and rural residual inequality ($\sigma_U(u)$, $\sigma_R(u)$), the residual urban-rural gap ($D_U - D_R$), urban and rural educational inequality ($\sigma_U(E)$, $\sigma_R(E)$), and the urban-rural educational gap ($E_U - E_R$). The next four columns report aggregates calculated from these components, namely overall urban or rural inequality ($\sigma_i = \sqrt{(\sigma_i(u))^2 + (\sigma_i(E))^2}$), the overall urban-rural gap ($D_U - D_R + E_U - E_R$), and the national standard deviation of ln consumption, as given by equation (5) earlier. Each cell of the table represents a different regression of the 65 country mean values of the variable listed in the column heading on the independent variable in the row heading, plus a sub-Saharan dummy. Since I use the ln of the dependent variables, were I to measure these in money metric units (i.e. include R_E in their calculation), the (unreported) sub-Saharan dummy would change, but the coefficients I focus on would remain the same.

My main purpose in presenting the results of Table VI is to serve as a contrast for the significant results presented later in the paper. Whether measured using the DHS's urban household population share or the World Bank's estimate of the population's urbanization rate, the Penn World Tables measure of real gdp per capita or the United Nations' measure of constant dollar real GDP, the urbanization rate and real GDP per capita have virtually no significant

relation with any of the components of inequality. Rural educational inequality does appear to rise with the level of GDP per capita, perhaps simply because, given that educational attainment is bounded from below by zero, the expansion of education at very low levels of attainment cannot but raise its standard deviation. Levels of aggregate inequality also appear to be higher in more urbanized countries. Overall, however, there are precious few significant relations.²³ In particular, the residual urban-rural gap, shown to be one of the most important determinants of inequality in the tables and figures above, is quite independent of the urbanization rate and GDP.

The results regarding the urban-rural gap in Table VI seemingly contradict Caselli (2005) and Restuccia, Yang and Zhu (2008) who, using PWT and FAO data, show that the ratio of non-agricultural to agricultural productivity falls with real GDP per capita. The negative relation they find between relative sectoral productivity and GDP per capita exists primarily amongst rich countries, which are absent from my sample. Taking Caselli's publicly posted data, a regression of the ln of his estimate of relative gdp per worker in non-agriculture to agriculture on his ln PWT 6.1 gdp per worker measure produces a coefficient of $-.799$ (.085). However, restricting the regression to the poorest $\frac{1}{2}$ of his sample changes the coefficient to $-.397$ (.182). If one then adds a sub-Saharan dummy to this regression, the coefficient is an insignificant $-.059$ (.232). Similarly, restricting the regression to the overlap of Caselli's data with the countries in my DHS sample (37 countries) and adding a sub-Saharan dummy produces a coefficient of $-.262$ (.200). The results in Table VI are quite consistent with the stylized facts presented by Caselli and Restuccia, Yang and Zhu, given which countries account for the patterns they describe.²⁴

²³Things do not improve if I enter the urban population shares along with the PWT or UN measure of GDP, while regressions on the variable listed in each row plus its square (i.e. quadratics) always find the two terms to be jointly insignificant.

²⁴There is also the issue that they are focusing on relative productivity, whereas my measures are relative real living standards. If q is output per worker, P_Q the price of output, and P_C the price of consumption, then their measure relates to a comparison of q across sectors, whereas my measure, roughly speaking, relates to a comparison

(e) Relative Consumption of Migrant Households

In this section I re-estimate the consumption inequality model treating migrant status (rather than simple urban/rural residence) as the defining characteristic of groups. This allows me to explore the relative mean consumption and variance of consumption by migrant status. As described earlier in Section II, I use data on an individual's current residence and residence prior to the age of 12 to classify households into permanent rural or urban residents and rural to urban or urban to rural migrants, while data on when and from where an individual moved into an area allow me to classify households into permanent urban or rural residents and rural to urban, rural to rural, urban to rural and urban to urban migrants. I determine the migrant status of households on the basis of the migrant status of each sex, dropping households where there is disagreement between the migrant status of individuals aged 25 to 49 within the household. As male questionnaires are available much less frequently than female questionnaires, I consider separate classifications of households based upon the status of either adult male or female members.

Table VII reports the relative consumption, after accounting for educational attainment, of the different migrant groups. As shown, urban permanent residents enjoy residual mean consumption levels that depending upon the sample are about 8 to 10 years higher, in units of equivalent education, than those of permanent rural residents. This is consistent with the mean 9 year gap in urban-rural residual consumption levels for the full DHS sample noted earlier in Table IV.²⁵ On average, rural to urban migrants appear to enjoy living standards that are slightly lower than urban permanent residents, while urban to urban migrants enjoy residual living

of P_Q^*q/P_C (nominal income divided by consumption prices). As noted by Lagakos and Waugh (2011), relative non-agricultural to agricultural output prices rise with GDP per capita, and this offsets the fall in relative non-agricultural productivity.

²⁵The estimates in this table, as in Table II earlier, exclude surveys where there is less than a .99 consistency between the urban/rural and city/town/village classifications used by local statistical authorities, as well as any country which is reported by the Norwegian Refugee Council as having more than 1% of the population internally

standards that are slightly higher. For the most part, however, these differences are not statistically significant, so whatever difference exists on average it is not very precisely estimated survey by survey.²⁶ As shown in the table, urban to rural migrants enjoy living standards that are a little over two years of equivalent education greater than those of rural permanent residents, while rural to rural migrants have an advantage of half a year or less. Again, while these differences appear to exist on average, they are often insignificant in individual surveys.²⁷ Comparing the living standards of migrants to permanent residents in their region of origin, in money metric terms (using the R_{ES} mentioned earlier) rural to urban migrants appear to enjoy a ln consumption gain ranging from .89 to 1.10, depending upon the measure used to determine migrant status, while urban to rural migrants appear to suffer a ln consumption loss ranging from -.66 to -1.00. Finally, I note that the consumption standard deviations of all the groups are extraordinarily similar. In particular, there are rarely any significant differences between the variance of outcomes of interregional migrants relative to their stay-at-home cousins.²⁸ Thus, it is difficult to motivate the interregional movement of labour with a mean-variance tradeoff.

V. Unobserved Human Capital and the Urban-Rural Gap

There are large urban-rural differences in consumption. Moreover, rural to urban migrants appear to enjoy great improvements in their living standards with no change in the

displaced by conflict or violence in any year between 2001 and 2010 or reported by the UNHCR as having more than 1% of its population as protected or assisted internally displaced persons in any year between 1993 and 2010.

²⁶Thus, looking at the consumption of rural to urban migrants relative to urban permanent residents, the difference is significant at the 1% level in 18 of the 33 countries in the upper left-hand panel, but in only 13 of 38, 6 of 25 and 6 of 30 countries in the other panels. Regarding urban to urban migrants and urban permanent residents, the differences are significant in 23 of 38 and 8 of 30 countries in the two relevant panels.

²⁷Thus, the urban to rural versus permanent rural difference is significant in 29 of 33, 30 of 38, 11 of 25 and 14 of 30 countries, while the rural to rural versus permanent rural difference is only significant in 10 of 38 and 6 of 30 countries.

²⁸Of the 252 such comparisons possible in the various panels of Table VII, only 28 are statistically significant at the 1% level.

variance of their outcomes. These facts naturally lead one to the conclusion that the urban-rural gap represents an incompletely exploited arbitrage opportunity, a “wedge” between urban and rural factor returns whose removal would dramatically increase overall living standards.

However, it is apparently also true that there are large numbers of urban to rural migrants who seemingly “enjoy” great reductions in their living standards, again with no change in the variance of their outcomes. Since this population flow appears to be completely voluntary,²⁹ one can only conclude that these individuals would not actually have had better outcomes had they remained in urban areas, i.e. that they have unobserved characteristics that firmly place them in the lower tail of the consumption distribution of the urban born. This viewpoint then allows one to reconsider rural to urban flows as representing, possibly, the outward movement of individuals whose unobserved characteristics firmly place them in the upper tail of the consumption distribution of the rural born. In sum, the urban-rural gap need not represent a gap between urban and rural factor returns; it might simply be the empirical manifestation of the geographic sorting of the population on the basis of unobserved attributes.

In this section I present a model of observable and unobservable human capital and the demand for skills that produces an urban-rural sorting of the type described above. The model is similar in spirit to the model of unobservable skills and agricultural-non-agricultural productivity differences developed by Lagakos and Waugh, which in turn is an application of Roy’s (1951) model of worker heterogeneity. Relative to Lagakos and Waugh, the principal innovations are geographically distinct relative factor demands and the introduction of education as an observable indicator of unobservable skills. These elements end up producing testable predictions of the urban-rural gap as a function of residence probabilities which are supported by the DHS data.

²⁹At the risk of annoying the reader, I once again emphasize that in discussing patterns of migration (Table II) and the consumption of different migrant groups (Table VII) I have removed from the sample all countries which have had more than 1% of their population displaced by internal conflict.

(a) Model

Consider an economy with two perfectly competitive industries, urban and rural, which produce output according to the constant returns to scale production functions

$$(6) Q_i = A_i S_i^{\alpha_i} US_i^{1-\alpha_i}$$

where i denotes the sector (urban or rural), S_i and US_i the input of skilled and unskilled labour, A_i a productivity parameter, and α_i the skilled share of total factor payments. The urban sector is more skill intensive than the rural sector, i.e. $\alpha_U > \alpha_R$. Due to unspecified positive externalities or complementary local factors, industries are geographically concentrated, so workers working in urban industry must reside in urban areas and workers working in rural industry must reside in rural areas. There are, however, no barriers or costs to the movement of labour. Thus, in equilibrium skilled workers are paid w_S everywhere and unskilled workers w_{US} .

The well-known first order conditions for each firm's optimal use of inputs imply

$$(7) \frac{S_i}{US_i} = \frac{\alpha_i}{1-\alpha_i} \frac{w_{US}}{w_S}$$

With S and US representing the total supply of skilled and unskilled labour, let $\Pi_S^i = S_i / S$ denote the share of the skilled labour force working in industry i , or equivalently the probability a skilled worker is employed and resides in sector i , with $\Pi_{US}^i = US_i / US$ denoting the same for unskilled labour. Obviously, $\Pi_x^R = 1 - \Pi_x^U$ for $x = S, US$. Using this fact and (7) for both industries, one easily derives the relation:

$$(8) \frac{\Pi_S^U}{1 - \Pi_S^U} = \frac{\alpha_U}{1 - \alpha_U} \frac{1 - \alpha_R}{\alpha_R} \frac{\Pi_{US}^U}{1 - \Pi_{US}^U}$$

As $\alpha_U > \alpha_R$, in equilibrium

$$(9) \Pi_S^U > \Pi_{US}^U \quad \text{and} \quad \Pi_S^R < \Pi_{US}^R$$

Given the greater skill intensity of the urban sector, a skilled worker is more like to work and

reside in the urban sector than an unskilled worker, and vice-versa for the rural sector.

Workers' human capital is composed of observable and unobservable components. While E , the years an individual spent in formal education, is directly observable, the actual outcome of that education is not. With probability $P(E)$ a student graduates to become a skilled worker and with probability $1-P(E)$ graduates as an unskilled worker, with $P'(E) > 0$. Normalizing nominal values so that the wage of the unskilled equals one, we see that the expected ln income of someone with education E is given by $\ln(w_s) * P(E)$. Consequently, the Mincerian return to education (R_E) equals $\ln(w_s) * P'(E)$. A positive Mincerian returns requires that $w_s > 1$, i.e. that factor supplies and demand are such that in equilibrium skilled workers earn more than unskilled workers, which I shall assume throughout.

Turning to the urban-rural gap, let $P(E,i)$ denote the probability a worker is skilled given that they have educational attainment E and are observed working in sector i . This is given by:

$$(10) \quad P(E,i) = \frac{P(E)\Pi_s^i}{P(E)\Pi_s^i + (1-P(E))\Pi_{US}^i}$$

where the denominator is the probability workers of educational attainment E work in sector i , while the numerator equals the probability they work in sector i and are skilled. From (9) it follows that $P(E,U) > P(E,R)$, for any given level of educational attainment a worker observed in the urban sector is more likely to be skilled. This is a natural consequence of the urban sector's higher relative demand for skilled workers. Since by assumption skilled workers earn more on average than unskilled workers, these probabilities produce an urban-rural gap, as for a given educational attainment urban workers earn more than comparable rural workers. Measured in units of equivalent education, this is given by:

$$(11) \text{ UR Gap} = \frac{\ln(w_S)[P(\mu_E, U) - P(\mu_E, R)]}{R_E}$$

$$= \frac{1}{P'(\mu_E)} \left[\frac{P(\mu_E)\Pi_S^U}{P(\mu_E)\Pi_S^U + (1 - P(\mu_E))\Pi_{US}^U} - \frac{P(\mu_E)\Pi_S^R}{P(\mu_E)\Pi_S^R + (1 - P(\mu_E))\Pi_{US}^R} \right]$$

where I have substituted μ_E (the average educational attainment) for E in the equations, as the econometric estimates are determined, more or less, by the characteristics of the average worker.

Equation (11) highlights two points. First, it shows that, modulo the production function $P(E)$ mapping from education to skill, the urban-rural residence probabilities are “sufficient statistics” for the urban-rural gap. Thus, while a full panoply of elements (demand, factor supplies, and factor intensities) underlies the determination of these probabilities,³⁰ given information on these probabilities one can ignore all of the underlying general equilibrium structure. Thus, in presenting the model I have skimmed over or avoided such details because, conditional on the residence probabilities, they are completely irrelevant. Second, the model predicts that when the urban residence probabilities of the skilled and unskilled are equal, so that $\Pi_S^U = \Pi_{US}^U = \Pi^U$ (the overall urban residence probability), the urban-rural gap is zero. Thus, it is not higher or lower rates of urbanization per se that generates the urban-rural gap, as confirmed by Table VI earlier which found no relation with overall urbanization, but rather the differences in the urbanization rates of different types of workers. I explore this prediction further below.

To summarize, differing factor intensities produce differences in the ratio of skilled to unskilled workers across sectors. In equilibrium, it is more likely that a skilled worker will choose to work in the urban sector than an unskilled worker. While observable educational attainment determines the probability a worker is skilled, the correlation between educational attainment and skill is not perfect. The relative factor demands of the different sectors, however,

³⁰So, countries with a greater supply of educated labour will, by the usual Rybcinski effects, have a greater fraction of workers of all types in urban areas. Similarly, levels of development and proximity to large overseas markets will, through the income elasticity of demand and trade, respectively, affect the allocation of labour to the different sectors.

produce a sorting of workers, so that a worker of a given educational attainment working in the urban sector is more likely to be skilled. Consequently, urban workers end up earning more than rural workers for a given educational attainment, producing the observed "urban-rural gap."

(b) Empirical Predictions

While the model described above qualitatively matches the empirical patterns chronicled in earlier sections, it also produces strong additional, testable, predictions concerning the determinants of the urban-rural gap, i.e. the difference in urban and rural living standards for a given level of education. Taking a first order expansion of (11) around the points

$\Pi_S^U = \Pi_{US}^U = \bar{\Pi}$ and $\mu_E = \bar{\mu}$, one finds that:

$$(12) \text{ UR Gap} = \left[\frac{P(\bar{\mu})[1-P(\bar{\mu})]}{P'(\bar{\mu})\bar{\Pi}(1-\bar{\Pi})} \right] (\Pi_S^U - \bar{\Pi}) - \left[\frac{P(\bar{\mu})[1-P(\bar{\mu})]}{P'(\bar{\mu})\bar{\Pi}(1-\bar{\Pi})} \right] (\Pi_{US}^U - \bar{\Pi}) + 0^*(\mu_E - \bar{\mu})$$

This suggests the regression:

$$(13) \text{ UR Gap}(c) = \alpha + \beta_1(\Pi_S^U(c) - \bar{\Pi}) + \beta_2(\Pi_{US}^U(c) - \bar{\Pi}) + \beta_3(\mu_E(c) - \bar{\mu}) + \varepsilon(c)$$

where c denotes the country which is the unit of observation and ε is the error term motivated by the random error in the measure of the urban-rural gap brought about by product and household sampling. From the theory described above, we see that the model predicts:

$$(14) \hat{\alpha} = \hat{\beta}_3 = 0 \text{ and } \hat{\beta}_1 = -\hat{\beta}_2 \text{ (as well as } \hat{\beta}_1 > 0, \hat{\beta}_2 < 0 \text{)}.$$

For concreteness, in the regressions which follow I will specify $\bar{\Pi}$ and $\bar{\mu}$ to be the mean, across the observations, of the national urbanization rate and average educational attainment, but in a linear regression of this sort the test statistics on (14) are completely insensitive to the choice of the expansion point.³¹ Equation (14) constitutes a fairly demanding and restrictive set of predictions, in that it is rare to ask of a regression that the constant term be zero and the

³¹That is, as one varies $\bar{\Pi}$ and $\bar{\mu}$ the estimate of α moves, but the p-values on the hypothesis $\alpha = 0$ and the joint hypothesis $\alpha = \beta_3 = 0$ and $\beta_1 = -\beta_2$ are unchanged.

coefficients be of equal magnitude and opposite and determinate sign. Two regressors, by themselves, and with restrictions on their signs and relative values, should explain all of the average urban-rural gap.³²

In the DHS data, Π_S^U and Π_{US}^U are not directly observed. However, the urbanization rates of extreme education groups provide reasonable proxy measures. In the model described above, the urbanization rate of workers with educational attainment E is given by:

$$(15) \Pi_E^U = P(E) * \Pi_S^U + (1 - P(E)) * \Pi_{US}^U$$

One can think of estimating, survey by survey, the probability individuals of educational attainment E reside in urban areas, $\hat{\Pi}_E^U$. With $P'(E)$, as E increases (decreases) this probability is increasingly representative of Π_S^U (Π_{US}^U). If one posits that, for all intents and purposes, the probability of acquiring skill is 1 for individuals at the extreme upper end of the educational distribution (say those with 16 years of attainment) and 0 for individuals at the extreme lower end of the educational distribution (say those with 0 years of attainment),³³ one has

$$(16) \Pi_S^U = \hat{\Pi}_{16}^U \quad \text{and} \quad \Pi_{US}^U = \hat{\Pi}_0^U.$$

These estimated values can then be substituted into a regression of the form of (13).³⁴

Before turning to the results, it is necessary to address a potentially important econometric

³²If one believes that the urban rural gap has nothing to do with differences in residence probabilities or education rates, the alternative hypothesis is that $\beta_1 = \beta_2 = \beta_3 = 0$ and $\alpha =$ mean residual (net of education) urban-rural gap in living standards (approximately equal to 9 yrs of equivalent education).

³³For the reader crying out (mentally) that uneducated workers often have skills, in the colloquial sense, not possessed by those with rarified tertiary education, allow me to reemphasize: the term “skill” in this paper merely means an ability used relatively more intensively in the urban sector that is acquired through education and confers substantially higher average incomes.

³⁴As the consumption data are at the household level, I estimate the residence equation at the household level as well, running a simple logit of a household’s residence (urban or rural) on the average educational attainment of its adult members. As the samples are not balanced, the logit is weighted by the relative urban-rural population weights so as to produce a population-accurate prediction of the urban-rural population distribution. As described above, the estimated residence probabilities at the upper and lower extremes of the educational distribution (i.e. 16 and 0 years of education) are then used as proxies for the urban residence probability of the skilled and unskilled, respectively.

issue. In the regressions which follow I regress one estimate (household consumption) on another set of estimates (residence probabilities and average educational attainment). Although these are based on different aspects of the data, they share in common the sample distribution of household educational attainment. Sampling variation, consequently, may produce correlated errors in the estimates of underlying parameters. Thus, the observation that two sets of parameter estimates are correlated may reflect the correlation of the underlying parameters or simply the correlation of their estimation error. I address this problem by first using bootstrap sampling techniques to re-estimate all of the consumption, residence and average attainment equations 250 times,³⁵ calculating the correlation between the estimation error of the different estimates. I then use maximum likelihood techniques to calculate the correlation matrix of the estimated variables purged of the correlation brought about by estimation error, as described more fully in Appendix B. In tables below I present results with and without this correction.

(c) Results

Table VIII regresses the residual urban-rural gap on a constant, the urban residence probabilities of skilled and unskilled individuals, and mean educational attainment. I begin in column (1) of the upper panel by proxying the skilled and unskilled with individuals of 16 and 0 years of education, as planned earlier above. The results support the model. The constant term and coefficient on mean educational attainment are insignificantly different from zero, and the coefficients on the urban residence probabilities of the highly and poorly educated are pretty much equal and opposite in sign. The joint test of all the coefficient restrictions has a gargantuan p-value. Adjusting the estimation procedure to take into account the estimation induced covariance between the dependent and independent variables, in column (4) of the table,

³⁵Rather than random sampling the households, I stratify the sample at the urban-rural level and resample clusters of households, as this more accurately reproduces the typical survey sampling framework.

produces similar results.

Enquiring minds might want to know how sensitive the results are to the choice of proxies for the skilled and unskilled. Column (2) of the upper panel of Table VIII substitutes the residence probabilities of those with 12 and 4 years of education as the relevant regressors. The p-value on the joint test is smaller, but still completely insignificant. Column (3) of the table moves in asymmetrically, retaining the residence probability of those with 16 years as a proxy for the skilled, but substituting those with 4 years for the unskilled. The joint hypothesis is now rejected at the 5% level, principally because the coefficients on the opposing residence probabilities are no longer equal in magnitude. However, moving the opposite way, using the residence probability of those with 12 years for the skilled but retaining that of 0 years for the unskilled, produces a p-value of .288 (not shown). Since the probability of being skilled is rising in educational attainment, the best proxies for the urban residence probabilities of the skilled and unskilled are the urban residence probabilities of the extremes of the education distribution, i.e. those with 16 and 0 years of attainment, and these produce estimates that are most consistent with the null hypothesis. Nevertheless, if one wants to play around with the choice of proxies, one can produce less favourable results.

The bottom panel of Table VIII adds a variety of regressors to a baseline regression with a constant and the urban residence probabilities of those with 16 and 0 years of education proxying for those of the skilled and unskilled. Since, as seen by comparing the right and left hand sides of the upper panel, the estimation error covariance between the estimates is not driving any of the results, I focus on standard OLS procedures (estimates with the adjustment are similar). I enter, one at a time, the urban population share as estimated by the DHS and as estimated by the World Bank, real income per capita as estimated by the Penn World Tables and as estimated by the

United Nations, and a sub-Saharan dummy. In each case, the null hypothesis that the constant and the coefficient on the additional variable are zero, and the coefficients on the urban residence probabilities are of opposite sign and equal magnitude, is never even close to being rejected. In the final column of the table I enter all of the additional regressors, as well as the mean educational attainment of the population, and find that the null hypothesis that all of them, plus the constant, are 0 and the coefficients on the urban residence probabilities are of opposite sign and equal magnitude, has a p-value of .210.³⁶ As predicted by the model, the urban residence probabilities of the skilled and unskilled, by themselves, and constrained in sign and magnitude, are sufficient to explain all of the urban-rural gap. Nothing else matters.

(d) Discussion

A model that explains urban-rural gaps by appealing to geographic sorting on the basis of unobservable skill appears, at first glance, to be nothing short of tautological, producing results that are no more than its assumptions. However, an examination of the equations describing the urban-rural gap, as in (11) and (12) above, produces the desired gentlemanly distance between assumptions and results. To a first order approximation, the influence of the urban residence probability of the highly and poorly educated on the urban-rural gap should be opposite in sign and equal in magnitude, and together they should explain the whole of the urban-rural gap, leaving nothing for the constant term or average educational attainment. These surprising restrictions are supported by the data.

Missing from the presentation above is a prediction of migration flows. If skilled and

³⁶Since this last regression is flooded with regressors, the standard errors are large, which might lead one to erroneously conclude that the urban-rural gap is independent of everything. A test of the null hypothesis that $\beta_1 = \beta_2 = 0$, i.e. the residence probabilities don't matter, is rejected with a p-value of .004. In response to a query from a referee, I also note that the seemingly peculiar opposite sign of the DHS and World Bank urbanization rates and PWT and UN real gdp measures comes from the fact that they are all entered simultaneously in this final regression. Although these variables are highly correlated with each other, their coefficients come from their orthogonal variation, so there is no reason to expect their signs to be the same in a multiple regression.

unskilled workers are completely homogeneous and moving costs are literally zero, as described, then migration flows are somewhat indeterminate as, up to certain limits imposed by regional employment and endowments, urban and rural born workers can exchange positions without disturbing any equilibrium relation. Introducing some heterogeneity in workers' urban and rural productivity conditional on their skill status solves this problem. As I show in an online appendix, a model of this sort easily explains the relative educational characteristics of migrants. The intuition is obvious. As education is correlated with skill, sorting on the basis of unobservable skill produces observable patterns as well. Thus, the concentration of the demand for skilled workers in urban areas and unskilled workers in rural areas ensures that, on average, the typical rural to urban migrant will be better educated than the typical rural born permanent resident, while the typical urban to rural migrant will be less educated than the typical urban born permanent resident. The appendix also shows how by allowing that urban education is somewhat more efficacious in producing skill than rural education the model can be extended to simultaneously explain both differences in the urban residence probabilities of the urban and rural born and in their consumption levels when residing in the same region.

It is useful to restate the central stylized facts which motivate the model and consider alternative explanations of the data. In the DHS one observes rural born young adults migrating to urban areas where they are observed to enjoy much higher consumption levels at middle age than their non-migrant rural cousins, without any increase in the variance of outcomes, and one also observes urban born young adults migrating to rural areas where they are observed to enjoy much lower consumption levels at middle age than their non-migrant urban cousins, without any decrease in the variance of consumption. While the rural to urban facts can be explained by barriers, which only a lucky few are able to overcome, the apparently voluntary urban to rural

movement encourages the search for alternative explanations. One possible explanation, as put forward in this paper, is that, on the margin, the mean consumption outcomes of migrants would not have been different if they had remained in the region of their birth, i.e. that the different migrant groups represent opposite tails of the consumption distribution of their respective native born populations.

An alternative explanation is that migrants are getting or losing something that is unobserved in the DHS data. Perhaps, as suggested by Gollin, Parente & Rogerson (2004) there is some form of unobserved home production and consumption in rural areas. Perhaps there is simply the pleasure of fresh air and rural living. Whatever they are, the preference for these unmeasured gains must be inversely related to education, as educated people on average appear to have a higher proclivity to select the measured material gains (for given levels of education) of urban living. With this premise in place, however, the resulting model is by and large observationally equivalent to the one set out above. Cross-country differences in the urban-rural residence probabilities of highly and poorly educated individuals identify the degree to which there are unobserved benefits to rural living that generally are more favoured by the less educated, and these explain the urban-rural gap in measured living standards. It is noteworthy, however, that this alternative framework is in the spirit of the one set out above: the urban-rural gap is not in and of itself a distortion, it simply reflects efficient sorting conditional on unobserved characteristics of urban-rural life.

Efficient sorting does not necessarily imply, however, that the conditions that induce the sorting are themselves efficient. Thus, a conscientious referee has pointed out that the higher incomes of the better educated might come from non-competitive government jobs, i.e. that “skill” itself might be nothing other than a watchword for entrance into non-competitive

protected professions which employ more people in urban areas than in rural areas. In principle, this is true. Similarly, the unobserved characteristics that make urban or rural living more or less pleasant might stem from market failures, e.g. the failure to properly penalize pollution externalities in congested urban living. Such discussions easily become ideological, as each new fact can be explained away, depending upon one's proclivities, as the manifestation of (unconsidered) distortions or the efficient competitive allocation subject to (unconsidered) factors, preferences and technology. Every individual has a different threshold. In the case of this author, having for decades believed the paradigm that urban-rural differences in developing countries represent a gigantic, unexploited, arbitrage opportunity, the DHS evidence of the large scale voluntary movement of urban born individuals to places where they seemingly consume dramatically less suggests the need to reevaluate preconceived notions and seriously consider the plausibility of alternative, efficiency based, explanations.

VI. Conclusion

There is an inherent risk in appealing to unobservables as an explanatory factor, as the argument easily becomes dogmatic and non-falsifiable. There are circumstances, however, in which the implications of unobservables are so great one feels compelled to give them consideration. In this paper I show that the urban-rural gap accounts for the lion's share of the level and cross-country variation in within country inequality. If this gap reflects wedges, a gap between the value marginal product of otherwise identical people living in urban and rural areas, one must conclude that this urban-rural distortion is responsible for generating much of the inequality within countries. Caselli (2005) and Restuccia, Yang and Zhu (2008) have argued that the physical productivity differences between agricultural and non-agricultural industry are larger in poorer economies and, within the context of modern development accounting, that their

removal would greatly reduce income inequality between countries. If the urban-rural gap reflects the sorting of labour based upon unobserved skill, none of these arguments carry through.

In the model presented in this paper the urban-rural gap reflects the efficient allocation of labour in response to the regional demand for skill. Where it is larger, it simply reflects a greater difference in the skill intensity of urban and rural industry. Perhaps at higher capital-labour ratios, which are associated with higher levels of GDP per capita, the difference in the relative skill intensity of agricultural and non-agricultural industry falls. Perhaps some countries simply have endowments that lead to patterns of comparative advantage with greater urban-rural skill differences. In either case, there is no way in which one can think of the urban-rural gap itself as a proximate cause of inequality or poverty. There is no urban-rural distortion to be removed to generate equity and wealth, no gigantic market failure to be identified as the bane of poor countries. There is simply the overall endowment and distribution of resources and technology which, needless to say, can always be beneficially improved.

Alwyn Young

London School of Economics

Appendix A: Demographic and Health Survey Data

Table A1 below lists the 170 DHS surveys used in the paper, 167 of which have data on at least four measures of durable goods or housing consumption and are used to estimate consumption inequality. The remaining 3 have data on wages or migration and are added to the sample for those topics. 27 surveys have data on individual wage income and 146 have data on at least one of the two measures of migration.

The DHS survey codes corresponding to the living standard variables listed in Table I above are ("hv" variables come from the household file, all others from the women's file):

Radio (hv207), television (hv208), refrigerator (hv209), bicycle (hv210), motorcycle (hv211), car (hv212), telephone (hv221), electricity (hv206), tap drinking water (hv201), flush toilet (hv205), constructed floor (hv213), diarrhea (h11), fever (h22), cough (h31), alive (b5), attending school (hv121 or hv110 if unavailable), working (v714), gave birth past year (v209), ever married (v502).

All "don't know" or "missing" responses are dropped from the sample. Some variables are recoded into broad dichotomous 0/1 categories, or to correct survey anomalies and differences, as follows:

Constructed floor: hv213 \leq 13 (dirt/sand/dung) = 0, otherwise (cement/wood/tiles/etc) = 1.
Flush toilet: hv205 $<$ 21 (including septic tanks) = 1, otherwise (pit/latrine/bush/etc) = 0. Tap drinking water: hv201 $<$ 21 (tapped or piped) = 1, otherwise (well/stream/lake/etc) = 0.
Diarrhea, fever and cough in past 2 weeks: yes answers 1 or 2 coded as 1 (extra detail on past 24 hours not universal across surveys and not used), no coded as 0. Gave birth past year: one or more births coded as 1, none coded as 0. Marital status: currently and formerly coded as 1, never coded as 0.

Conditioning/demographic variables are constructed as follows:

Ln number of household members (number of hvidx household records); young children's sex (b4) and age in months (v008-b3); youth's sex (hv104) and age (hv105); married women's age (v012, or hv105 for the marital status equations).

Because of changes in the coverage of DHS survey questionnaires over time, samples are restricted to generate consistent samples, as follows:

Children's health variables: children aged 35 months or less (i.e. born within 35 months of the survey). Women's fertility and work variables: currently married women only.

Migration measures are based on v103, (childhood place of residence), v104 (years spent in place of residence), & v105 (type of place of previous residence), and the corresponding "mv" data from the men's files. For both v103 and v105 I code "abroad" as missing, recode various city and township locales as urban and countryside as rural, and compare with hv025 (current urban/rural location of the household) to determine the migrant status of individuals. V104 pairs with v105 and is used to separate out urban to urban from urban permanent residents (and same

for rural) based upon whether individuals claim to have always lived in the locale, as well determine the number of years since an individual migrated (as reported in Table II).

There are numerous coding errors in both the consumption and migration measures, which I have identified and eliminated by working through the survey questionnaires. For example, in Uganda 00 individuals were asked where they lived 5 years ago, and this is recoded as childhood residence in the recode files. It is also not uncommon for surveys to use different codes for responses, e.g. 3 for countryside and 4 for abroad or 4 for countryside and 5 for abroad, which the DHS programmers fail to spot and recode based upon a standardized framework. I have also recoded dates to the Gregorian calendar (an issue for Ethiopia and Nepal), eliminated duplicate household observations and corrected inconsistencies and errors in rural/urban location within clusters.

For the wage regressions reported in Sections III, I restrict myself to female and male individuals aged 25 to 65 reporting that they work for others ($v719$ or $mv719 = 2$, "m" denotes the male questionnaire). Annual earnings are constructed from $v736/mv736$ data, with the earnings of individuals reporting annual, monthly and weekly wages multiplied by 1, 12 and 50, respectively (individuals reporting an hourly or daily wage, numbering about 1/5 of those working for others and reporting wage data, are dropped from the sample). As I have painstakingly recoded all the educational data for the household files, but have not done the same for the male and female questionnaires, I get individual age and educational characteristics by merging the individual files (which contain the earnings data) with the household files using the individual id numbers, eliminating cases where there is a discrepancy between sources of more than 2 years in the reported age (roughly 7 percent of cases that meet the other wage sample eligibility criteria).

Employment, schooling and marital status pose special problems. On women's employment, variation in the question form has dramatic effects on average responses. The standard questionnaire first asks women if, apart from housework, they are currently working and then follows up with a question that explains that women may work in a variety of ways (for cash or in kind, selling things, in their businesses, on farms or in the family business) and asks the respondent if she is currently doing any of these. The combination of these two questions forms the basis for DHS code $v714$. An occasional third question, on whether the woman has done any work in the past 12 months, then produces $v731$. The problem is that many DHS surveys vary this pattern, omitting the first or second of the two part $v714$ question, omitting the preliminary $v714$ questions in their entirety (but including the $v731$ question), and even modifying the questions to focus on working for cash only. To keep the data as consistent as possible, I restrict my employment measure to $v714$ in those surveys with the standard two-part question.³⁷

On schooling, some questionnaires ask whether the household member attended school in the past year ($hv121$) and others whether the household member is currently in school or still in school ($hv110$). The form of this question does not seem to be important, as the differences within surveys where the two questions overlap and between surveys when the questions change

³⁷There is some slight variation in the wording of the question, with the phrase "last week" introduced in some early and most recent surveys, which I allow in order to avoid losing too many observations.

are small. Consequently, I take hv121 when it is available, and use hv110 as a reasonable substitute when it is not. The main problems that arise in the educational data are that (1) in some surveys individuals who, when questioned on educational attainment, say they have never been to school are automatically coded as not currently attending school, whereas in other surveys they are not; (2) the educational attendance question is generally restricted to individuals 6 to 24, but in some surveys the age range is further restricted, while those who were not asked the question are automatically coded as not attending. I solve these problems by coding all individuals whose educational attainment is listed as having never attended school as not currently attending and, in cases where problem (2) arises for 6 year olds only, coding all 6 year olds as missing. For the Indian surveys, problem (2) arises for individuals older than 14, 17 or 18 (depending on the survey), eliminating most of the 15-24 age group. Consequently, I eliminate India from the sample for this variable. In the case of the few surveys with missing data for 6 year olds, I deem that the age controls and the existence of data for the remainder of youths aged 6-14 allow me to keep them in the sample.

On marital status (never vs currently/formerly), this is reported in some surveys in the household module, and always in the women's question module. The latter, however, is sometimes restricted to ever-married women. I take a woman's marital status from the household module when it is available there, and from the women's module when not. Where the women's module is restricted to ever-married women, I go back to the household module and identify the additional eligibility criterion for the women's module (usually "slept last night") and code all women in the household file meeting the additional eligibility criterion who are also listed as "not eligible" for the women's questionnaire as "never married", merging these records with the marriage data from the women's question module. The marital status of women who do not meet the additional eligibility criterion is uncertain (they are excluded from the female survey even if they are married), so they are dropped from the marital status sample.

Finally, I turn to educational attainment. The DHS questionnaires ask respondents for their educational attainment, measured as grade level achieved, not the number of years attended. The DHS "recode" takes this raw data, converts it into a broad categorical variable (hv106 = none, primary, secondary, tertiary), a measure of years at that level (hv107), and total years of attainment (hv108). Unfortunately, the procedures used by programmers to generate these conversions over the years have varied, with, for example, the number of years of education falling in each hv106 category varying even within countries. Most fundamentally, there are extraordinary errors and inconsistencies in reaching the final years of attainment (hv108), with, to cite some examples, those responding "don't know", a code of 8 in many surveys, credited 8 years of education; reaching tertiary education (not counting years there) being credited anything from 10 to 19 years base (sometimes, within the same country); upper secondary systems that require 10 formal levels to reach being coded as 6 years; etc. Working with the DHS questionnaires, original "raw" non-recode data generously provided by the DHS programmers, and summaries of educational systems and their history found on websites hosted by UNESCO, jstor, education.stateuniversity.com, and the education ministries of different countries, I have recoded all the educational attainment data to represent years of formal attainment within each country's educational ladder, taking the level of entering 6 year olds as the starting point. In cases where systems change over time (e.g. an old system primary lasted 6 years and a new system primary lasts 8 years, so "completed primary" has different meanings), I use the timing of institutional

reform, an individual's birth cohort, and sample information on the distribution of years of attainment by age group (e.g. those with uncompleted primary up to a certain birth cohort indicate no more than 6 years) to impute an appropriate estimate of years of completed education to different birth cohorts.

Appendix B: Regressions and Correlations for Estimated Variables

Section V regresses estimated variables on each other. Independent measurement error in the dependent and independent variables leads to attenuation bias, while correlated measurement error produces spurious relations. In this section I explain how I address this problem.

Let the vector of M random variables X be distributed normally with mean μ and covariance matrix Ω . Each element m of μ is a separate function of k conditioning variables X_μ , $\mu_m = X_\mu \beta_{\mu m}$. X_μ may be merely a constant or a variety of independent variables, but in all cases each element of μ is a function of the same conditioning variables which are observed without error. Individual observations on the vector X are acquired with mean zero error:

$$(B1) \hat{X}_i = X_i + \varepsilon_i, \quad E(\varepsilon_i) = 0, \quad E(\varepsilon_i \varepsilon'_i) = \hat{\Omega}_i.$$

The covariance matrix of the estimation errors ($\hat{\Omega}_i$) may be given by the estimation procedure or by a bootstrap simulation. Given this model, each vector of observations \hat{X}_i follows a normal distribution with mean μ and covariance matrix $\Omega + \hat{\Omega}_i$. I estimate the parameters $\beta_{\mu m}$ and the elements σ_{ij} of Ω by maximizing the likelihood of the overall sample and compute their covariance matrix using the usual inverse of the empirical negative Hessian.

Let X_1 be the first element of X and let X_2 identify the remaining variables, with σ_{11} , Ω_{22} , and Ω_{12} denoting their variance and covariance matrices. The $M-1$ coefficients in the regression of X_1 on X_2 (with X_μ as additional controls) are given by $\beta_2 = \Omega_{22}^{-1} \Omega_{21}$ and the partial R^2 is given by $\beta_{12}' \Omega_{22}^{-1} \beta_{12} / \sigma_{11}$. The implicit k coefficients on X_μ in the same regression are recovered by $\beta_\mu = \beta_{\mu 1} - \beta_{\mu 2} \beta_2$, where (recall) $\beta_{\mu 1}$ and $\beta_{\mu 2}$ are the $k \times 1$ and $k \times (M-1)$ coefficients on X_μ used to determine the M vector of means μ in the likelihood. The covariance matrix (standard errors) of these coefficients are estimated by applying the delta method using the covariance matrix of the estimates $\beta_{\mu \mu}$ and σ_{ij} .

The procedure described above is simply partial regression with an adjustment, after the influence of the variables X_μ is partialled out, for the estimation error in the remaining variables. As such, when the error covariance matrices $\hat{\Omega}_i$ are matrices of 0s (i.e. there is no estimation error), it produces the usual OLS regression coefficients and partial R^2 s. Although degrees of freedom adjustments are not justified by asymptotic maximum likelihood theory, they are often

implemented in practice³⁸ and I include such an adjustment here as well. The degrees of freedom adjustment ensures that when the $\hat{\Omega}_i$ are matrices of 0s the estimated standard errors and test statistics equal their usual OLS counterparts.

I confirm the accuracy of the procedure by simulating it using the data used in Table VIII. Since I will simulate tests under the assumption that the null hypothesis is true, I begin by estimating the coefficients in the regression of the 65 country mean values on a constant term and deviations of the urban residence probabilities of 16 and 0 years of education from the mean urban residence probability subject to the constraints $\alpha = 0$ and $\beta_{16} = -\beta_0$.³⁹ Using the resulting constrained estimates of μ and Ω , and the given observation by observation estimation error $\hat{\Omega}_i$, I then produce 10,000 random replications of the data and estimate the *unconstrained* model (as was done in Table VIII).⁴⁰ Table B1 reports the results. As shown, the mean estimated value of the coefficients is virtually identical to the underlying values, while the mean estimated standard error of the same is very close to the actual variation. With regards to the F test of the null hypothesis, the nominal size of the test and the actual rejection probability are virtually identical at all levels. In sum, in a small sample of 65 observations the procedure appears to produce unbiased coefficients and standard errors and exact test statistics.

References

- Becker, Gary S., Tomas. J. Philipson, and Rodrigo R. Soares, "The Quantity and Quality of Life and the Evolution of World Inequality," American Economic Review, Vol. 95, No. 1 (March 2005), 277-291.
- Butler, J.S. and Robert Moffitt, "A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model," Econometrica, 50 (1982), 761-764.
- Caselli, Francesco, "Accounting for cross-country income differences," in Phillipe Aghion and Steven N. Durlauf (eds), Handbook of Economic Growth, Vol 1A, North Holland Elsevier, New York, NY, 2005, 679-741.

³⁸Stata, for example, introduces them in some maximum likelihood procedures.

³⁹Since the number of terms σ_{ij} rises with the square of M, I drop the mean educational attainment variable of Table VIII for the purpose of simulation.

⁴⁰To repeat, to examine the validity of the tests (under the assumption that the null is true), the null has to be true for the data generating process. Thus, I use the actual data to estimate parameters that are consistent with the null (i.e. subject to its constraints), use the parameter estimates to produce 10,000 replications of the data which by construction satisfy the null, and then run the unconstrained model and test the null hypothesis.

- Filmer, Deon and Lant H. Pritchett, "Estimating Wealth Effects Without Expenditure Data – or Tears: An Application to Educational Enrollment in States of India," Demography, Vol. 38, No. 1 (Feb. 2001), 115-132.
- Gollin, Douglas, David Lagakos, and Michael E. Waugh, "The Agricultural Productivity gap in Developing Countries," manuscript, March 2012.
- Gollin, Douglas, Stephen L. Parente and Richard Rogerson, "Farm work, home work and international productivity differences," Review of Economic Dynamics, 7 (2004), 827-850.
- Herrendorf, Berthold and Schoellman, Todd, "Why is Measured Productivity so Low in Agriculture?," manuscript, Arizona State University, October 2012.
- Jones, Charles and Peter Klenow, "Beyond GDP? Welfare across Countries and Time," manuscript, Stanford University, February 2011.
- Lagakos, David and Michael E. Waugh, "Selection, Agriculture, and Cross-Country Productivity Differences," manuscript, Nov. 2011.
- Lewis, W. Arthur, "Economic Development with Unlimited Supplies of Labor," The Manchester School of Economic and Social Studies, 22 (1954), 139-191.
- Psacharopoulos, George, "Returns to Investment in Education: A Global Update," World Development, 22 (1994), 1325-1343.
- Restuccia, Diego, Dennis Tao Yang and Xiaodong Zhu, "Agriculture and aggregate productivity: A quantitative cross-country analysis," Journal of Monetary Economics, 55 (2008), 234-250.
- Roy, A.D, "Some Thoughts on the Distribution of Earnings," Oxford Economic Papers, New Series, Vol. 3, No. 2 (June 1951), 135-146.
- Young, Alwyn, "The Gini Coefficient for a Mixture of Ln-Normal Populations," manuscript, London School of Economics, December 2011.
- Young, Alwyn, "The African Growth Miracle," Journal of Political Economy, Vol. 120, No. 4 (August 2012), 696-739.

TABLE I
DHS REAL CONSUMPTION MEASURES BY CATEGORY
(167 surveys across 2.1 million households in 65 countries)

| | N | Mean | | N | Mean |
|-----------------------------------|---------|------|--|---------|------|
| Ownership of Durables (household) | | | Household Time and Family Economics (individual) | | |
| Radio | 2062257 | .583 | Attending School (age 6-14) | 2405172 | .736 |
| Television | 2082338 | .448 | Attending School (age 15-24) | 1617008 | .366 |
| Refrigerator | 1978069 | .289 | Working (women age 15-24) | 301269 | .401 |
| Bicycle | 1958128 | .291 | Working (women age 25-49) | 978038 | .547 |
| Motorcycle | 1899527 | .117 | Gave Birth Past Year (age 15-24) | 348919 | .311 |
| Car | 1964232 | .077 | Gave Birth Past Year (age 25-49) | 1139948 | .139 |
| Telephone (landline) | 1633104 | .181 | Ever Married (women age 15-24) | 964597 | .415 |
| | | | Ever Married (women age 25-49) | 1464709 | .924 |
| Housing Conditions (household) | | | Children's Health (individual) | | |
| Electricity | 2039093 | .561 | Diarrhea | 726884 | .196 |
| Tap Drinking Water | 2058331 | .458 | Fever | 715084 | .308 |
| Flush Toilet | 1934458 | .363 | Cough | 725314 | .330 |
| Constructed Floor | 1905055 | .636 | Alive | 801877 | .935 |

Notes: All variables coded as 0/1. Ownership of Durables: at least one such item in the household; Housing Conditions: constructed floor means made of other than dirt, sand or dung. Household Time: individual variables, i.e. coded separately for each individual of that age in the household; recent fertility and market participation refer to currently married women only. Children's Health: individually coded for each child born within 35 months of the survey; diarrhea, cough and fever referring to the occurrence of these for the individual in question (if alive) in the preceding two weeks. N = number of households or individuals.

TABLE II
AVERAGE CHARACTERISTICS OF MIGRANTS AND NATIVE RESIDENTS (INDIVIDUALS)

| | Based on current locale & locale when \leq 12 yrs | | | | Based on when moved to current locale & locale before move | | | | |
|----------------------------|---|------|----------|-------|--|------|----------|-------|-------|
| Women aged 25 to 49 | | | | | | | | | |
| | Pop Shares | | Educ | Age | Pop Shares | | Educ | Age | Date |
| | Dest | Orig | (yrs) | (yrs) | Dest | Orig | (yrs) | (yrs) | (yrs) |
| (1) Lifelong Rural | .859 | .775 | 3.1 | 35.2 | .492 | .474 | 3.8 | 35.3 | NA |
| (2) Rural to Rural Migrant | | | | | .376 | .359 | 3.8 | 35.3 | 12.6 |
| (3) Urban to Rural Migrant | .141 | .230 | 4.9 | 34.4 | .132 | .202 | 5.5 | 34.7 | 10.2 |
| (4) Lifelong Urban | .657 | .770 | 7.2 | 34.3 | .398 | .418 | 7.1 | 34.6 | NA |
| (5) Urban to Urban Migrant | | | | | .363 | .380 | 7.8 | 35.1 | 10.7 |
| (6) Rural to Urban Migrant | .343 | .225 | 4.8 | 35.3 | .239 | .167 | 5.4 | 35.4 | 12.3 |
| # individuals | | | 389591 | | | | 623453 | | |
| # surveys in # countries | | | 62 in 34 | | | | 81 in 39 | | |
| Men aged 25 to 49 | | | | | | | | | |
| | Pop Shares | | Educ | Age | Pop Shares | | Educ | Age | Date |
| | Dest | Orig | (yrs) | (yrs) | Dest | Orig | (yrs) | (yrs) | (yrs) |
| (1) Lifelong Rural | .839 | .779 | 4.1 | 35.6 | .599 | .613 | 4.7 | 35.5 | NA |
| (2) Rural to Rural Migrant | | | | | .228 | .236 | 4.9 | 36.4 | 11.6 |
| (3) Urban to Rural Migrant | .161 | .277 | 6.1 | 35.3 | .173 | .263 | 6.7 | 35.7 | 9.6 |
| (4) Lifelong Urban | .622 | .723 | 8.6 | 34.5 | .379 | .368 | 8.3 | 34.4 | NA |
| (5) Urban to Urban Migrant | | | | | .400 | .369 | 9.3 | 35.7 | 10.4 |
| (6) Rural to Urban Migrant | .378 | .221 | 6.8 | 36.1 | .221 | .151 | 6.9 | 36.1 | 12.5 |
| # individuals | | | 44911 | | | | 118832 | | |
| # surveys in # countries | | | 33 in 25 | | | | 47 in 30 | | |

Notes: Since the number of individual observations differs greatly across surveys, as do the number of surveys per country, I take the country as the unit of observation. Thus, I calculate the measure for each survey, average the surveys for each country, and then calculate the average across the countries, reporting the final result in the table. Pop Shares – share of destination or origin population; Date = number of years since arrival. NA – not applicable.

TABLE III
LOGIT OF INDIVIDUAL MIGRANT STATUS ON MEAN MIGRANT STATUS OF CLUSTER NEIGHBOURS

| | Women aged 25-49 | | Men aged 25-49 | | |
|-------------------------------|------------------|-------------|----------------|-------------|------|
| | Urban | Rural | Urban | Rural | |
| MigMean | 3.24 (.029) | 4.37 (.034) | 1.73 (.072) | 2.36 (.099) | |
| N | 190611 | 212301 | 17795 | 28314 | |
| Predicted average probability | | | | | |
| | actual | .287 | .146 | .396 | .116 |
| MigMean = | 0 | .123 | .061 | .244 | .076 |
| | 1 | .774 | .818 | .626 | .429 |

Notes: Migrant status coded as 1/0 (yes/no) based upon childhood vs current place of residence. MigMean = average migrant status of survey cluster members of the same sex and age group outside of the individual's household. N = number of individuals. All logits include survey dummies. Standard errors are adjusted for clustering. Sample surveys are the same as those used in Table II, excluding countries with internal displacements and surveys with inconsistent urban/rural vs city/town/countryside coding, as discussed in the text above.

TABLE IV
COMPONENT ESTIMATES

| | $\sigma_U(u)$ | $\sigma_R(u)$ | $\sigma_U(E)$ | $\sigma_R(E)$ | D_U-D_R | E_U-E_R |
|---|---------------|---------------|---------------|---------------|-----------|-----------|
| Cross-Survey Distribution (167 Surveys) | | | | | | |
| mean coef. estimate | 5.47 | 5.73 | 4.09 | 3.10 | 9.04 | 3.13 |
| st. dev. of coef. est. | 1.16 | 1.46 | .669 | .795 | 4.23 | .871 |
| mean s.e. of coef. est. | .303 | .334 | .054 | .033 | .870 | .092 |
| Cross-Country Distribution (65 Countries) | | | | | | |
| mean coef. estimate | 5.57 | 5.76 | 3.98 | 2.96 | 9.38 | 3.00 |
| st. dev. of coef. est. | 1.21 | 1.59 | .713 | .748 | 4.33 | .911 |
| mean s.e. of coef. est. | .339 | .372 | .055 | .033 | .957 | .092 |

Notes: $\sigma_i(u)$ = within region standard deviation of residual, net of education, ln consumption; $\sigma_i(E)$ = within region standard deviation of educational attainment; D_U-D_R = difference between urban and rural consumption net of (at zero) educational attainment; E_U-E_R = difference between urban and rural mean educational attainment. All coefficients represent years of equivalent education, as described in the previous section, estimated using all available durables, housing, health and family economics products whose average regional consumption level lies between .01 and .99. Top panel, calculated off of the 167 survey estimates; bottom panel, calculated off of the 65 country means of the survey estimates.

TABLE V
WITHIN COUNTRY STANDARD DEVIATION OF LN CONSUMPTION
IN YEARS OF EQUIVALENT EDUCATION IN 65 COUNTRIES

| | | Mean Across Countries | Standard Deviation Across Countries |
|-----------------|---------------------------------------|--------------------------|--|
| Actual estimate | | 8.85 | 2.22 |
| | educational inequality | 7.25 | 2.15 |
| | residual inequality | 3.66 | 0.70 |
| Absent | urban rural differences | 6.69 | 1.38 |
| | educ. differences | 8.06 | 2.04 |
| | residual differences | 6.84 | 1.38 |
| | within region residual differences | 6.67 | 1.96 |
| | within region educational differences | 8.13 | 2.30 |

Note: calculated off of 65 observations of mean country values (i.e. the country means across the 167 survey observations).

TABLE VI
CORRELATIONS WITH URBANIZATION AND GDP
(regressions of 65 country mean values of ln dependent variable in each column)

| | $\sigma_U(u)$ | $\sigma_R(u)$ | D_U-D_R | $\sigma_U(E)$ | $\sigma_R(E)$ | E_U-E_R | σ_U | σ_R | UR gap | σ |
|-----------------------|------------------|-----------------|------------------|------------------|------------------|------------------|-----------------|------------------|------------------|------------------|
| $S_{U\text{ DHS}}$ | 0.327* (.156) | 0.279 (.196) | -0.089 (.382) | -0.153 (.143) | 0.263 (.206) | -0.310 (.292) | 0.170 (.120) | 0.252 (.179) | -0.067 (.308) | 0.391* (.172) |
| $S_{U\text{ WB}}$ | 0.282 (.155) | 0.316 (.192) | -0.153 (.375) | -0.108 (.141) | 0.311 (.201) | -0.291 (.287) | 0.152 (.119) | 0.293 (.175) | -0.126 (.302) | 0.360* (.170) |
| $\ln Y_{\text{ PWT}}$ | 0.038 (.038) | 0.068 (.045) | -0.028 (.087) | -0.034 (.033) | 0.099* (.047) | -0.045 (.069) | 0.016 (.029) | 0.073 (.041) | -0.027 (.070) | 0.056 (.040) |
| $\ln Y_{\text{ UN}}$ | 0.029 (.030) | 0.067 (.036) | -0.030 (.071) | -0.026 (.027) | 0.083* (.037) | -0.032 (.055) | 0.010 (.023) | 0.069* (.033) | -0.026 (.057) | 0.053 (.033) |

Notes: Each cell is a separate regression of 65 country means on the dependent variable listed on the left-hand side plus a sub-Saharan dummy. S_U = urban population share; $\ln Y$ = \ln gdp per capita; subscripts refer to sources. DHS = DHS household share; WB = World Bank databank urban population share; PWT = 7.0 (rgdpl); UN = constant US dollars in UN national accounts database; dependent variables as defined in text above; * = significant at the 5% level. PWT regressions exclude Zimbabwe. Between PWT 6.3 and PWT 7.0 the year 2005 \ln GDP per capita of Zimbabwe relative to the \ln mean of the entire 65 country sample falls by -2.3, i.e. relative gdp falls to .1 of its original value. This no doubt reflects the difficulties involved in measuring expenditure and prices during a hyperinflation, as PWT 7.0 uses 2005 as its benchmark year.

TABLE VII
 CONSUMPTION CHARACTERISTICS OF MIGRANTS AND NATIVE RESIDENTS (HOUSEHOLDS)
 (residual, net of education, consumption in yrs of equivalent education)

| | Based on current locale and locale when ≤ 12 yrs | | Based on when moved to current locale & locale before move | |
|--|--|------------|---|------------|
| | Relative ln C | SD of ln C | Relative ln C | SD of ln C |
| Household status determined by migrant status of women aged 25 to 49 | | | | |
| (1) Lifelong Rural | NA | 4.8 (.283) | NA | 4.8 (.330) |
| (2) Rural to Rural Migrant | | | 0.2 (.429) | 4.9 (.376) |
| (3) Urban to Rural Migrant | 2.4 (.550) | 4.9 (.461) | 2.4 (.568) | 5.0 (.464) |
| (4) Lifelong Urban | 8.8 (.743) | 4.6 (.272) | 7.8 (.759) | 4.8 (.338) |
| (5) Urban to Urban Migrant | | | 8.8 (.753) | 4.7 (.315) |
| (6) Rural to Urban Migrant | 7.5 (.722) | 4.8 (.316) | 7.2 (.735) | 4.8 (.345) |
| # households | | 351458 | | 593207 |
| # surveys in # countries | | 60 in 33 | | 79 in 38 |
| Household status determined by migrant status of men aged 25 to 49 | | | | |
| (1) Lifelong Rural | NA | 5.2 (.486) | NA | 5.3 (.570) |
| (2) Rural to Rural Migrant | | | 0.7 (.897) | 5.7 (.683) |
| (3) Urban to Rural Migrant | 2.1 (1.00) | 5.2 (.818) | 2.4 (.933) | 5.5 (.768) |
| (4) Lifelong Urban | 10.0 (1.21) | 5.0 (.560) | 9.1 (1.25) | 5.4 (.660) |
| (5) Urban to Urban Migrant | | | 10.0 (1.28) | 5.1 (.548) |
| (6) Rural to Urban Migrant | 8.7 (1.14) | 5.2 (.583) | 8.6 (1.20) | 5.1 (.633) |
| # households | | 45422 | | 123274 |
| # surveys in # countries | | 33 in 25 | | 47 in 30 |

Notes: In the notation of Section III: Relative ln C = $D_i - D_{base}$, where i is the group in question and base are lifelong rural households; SD of ln C = $\sigma_i(u)$. NA = not applicable, lifelong rural are the base category for mean consumption. The numbers reported are the mean country values and standard errors (i.e. mean of the country standard errors, not standard error of the country means).

TABLE VIII
REGRESSION OF URBAN-RURAL RESIDUAL CONSUMPTION GAPS ON
RESIDENCE PROBABILITIES AND MEAN EDUCATIONAL ATTAINMENT (65 COUNTRY VALUES)

| | Unadjusted | | | Adjusted for Error Covariance | | |
|--|-----------------------|-----------------------|-----------------------|-------------------------------|-----------------------|-----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Π_{16}^U, Π_0^U | Π_{12}^U, Π_4^U | Π_{16}^U, Π_4^U | Π_{16}^U, Π_0^U | Π_{12}^U, Π_4^U | Π_{16}^U, Π_4^U |
| α : constant | -696 (2.90) | 2.03 (2.06) | -654 (3.01) | .121 (2.97) | 2.63 (2.04) | .214 (3.12) |
| β_1 : $\Pi_S^U - \bar{\Pi}$ | 16.2 (5.12) | 19.4 (5.35) | 20.7 (6.17) | 14.5 (5.19) | 17.3 (5.31) | 18.4 (6.38) |
| β_2 : $\Pi_{US}^U - \bar{\Pi}$ | -12.1 (4.28) | -16.0 (4.88) | -10.1 (3.86) | -11.0 (4.18) | -14.3 (4.80) | -8.91 (3.79) |
| β_3 : $\mu_E - \bar{\mu}$ | .132 (.180) | .144 (.184) | .069 (.183) | .122 (.172) | .133 (.176) | .066 (.175) |
| R^2 | .173 | .182 | .159 | .151 | .165 | .135 |
| H_0 : $\alpha = \beta_3 = 0$ & $\beta_1 = -\beta_2$ (p-value) | .806 | .317 | .028 | .792 | .217 | .020 |
| α : constant | -1.31 (3.43) | -1.69 (3.51) | 5.45 (5.86) | -.746 (5.44) | -.397 (2.80) | 10.2 (8.27) |
| β_1 : $\Pi_{16}^U - \bar{\Pi}$ | 14.7 (5.01) | 14.7 (4.98) | 14.8 (4.92) | 15.2 (4.95) | 14.4 (4.91) | 16.2 (8.82) |
| β_2 : $\Pi_0^U - \bar{\Pi}$ | -13.9 (5.43) | -14.4 (5.48) | -9.95 (4.65) | -12.3 (4.86) | -11.4 (4.26) | -10.5 (11.1) |
| β_3 : $S_{U \text{ DHS}}$ | 2.23 (3.96) | | | | | -5.25 (17.2) |
| $S_{U \text{ WB}}$ | | 2.81 (3.96) | | | | 4.20 (9.27) |
| $\ln Y_{\text{PWT}}$ | | | -667 (.606) | | | -4.34 (1.77) |
| $\ln Y_{\text{UN}}$ | | | | .069 (.608) | | 3.33 (1.60) |
| Africa | | | | | 1.39 (.998) | .016 (1.42) |
| $\mu_E - \bar{\mu}$ | | | | | | .380 (.572) |
| H_0 : $\alpha = \beta_3 = 0$ & $\beta_1 = -\beta_2$ (p-value) | .859 | .814 | .647 | .929 | .500 | .210 |

Notes: Π_E^U = estimated urban residence probability of households with educational attainment E. Π_S^U, Π_{US}^U = urban residence probability of the skilled and unskilled, proxied in the top panel by those with the educational attainment listed in the column headings. μ_E = estimated mean household educational attainment. $\bar{\Pi}$ and $\bar{\mu}$ = average urbanization rate and mean household educational attainment of the sample. S_U and $\ln Y$ = urban population share and \ln gdp per capita, subscripts refer to sources (DHS = DHS urban household share, WB = World Bank databank urban population share, PWT = 7.0 (rgdpl), UN = constant US dollars in UN national accounts database). Africa = dummy for sub-Saharan Africa. Adjustment for error covariance = see Appendix B.

TABLE A1
DHS AND ASSOCIATED SURVEYS USED IN THE PAPER

| | | | | | |
|---------------|-----------------|--------------|----------------------------|-------------|-------------------------|
| Benin | 96*, 01, 06 | Rwanda | 92, 00, 05 | Bolivia | 94*, 98*, 03, 08 |
| Burkina Faso | 92, 98, 03 | Sao Tome | 08 | Brazil | 91, 96 |
| Cameroon | 91, 98, 04 | Senegal | 92, 05 | Colombia | 90, 95*, 00, 05, 10 |
| Cen. Af. Rep. | 94* | Sierra Leone | 08 | Dom. Rep. | 91, 96*, 99, 02, 07 |
| Chad | 96* | South Africa | 98* | Guatemala | 95*, 98* |
| Comoros | 96* | Swaziland | 06 | Guyana | 05, 09 |
| Congo | 05, 09 | Tanzania | 92, 96, 99, 03, 04, 07, 10 | Haiti | 94, 00, 05 |
| Congo Dem Rep | 07 | Togo | 98* | Honduras | 05 |
| Cote D'Ivoire | 94, 98, 05 | Uganda | 95*, 00, 06 | Nicaragua | 97*, 01 |
| Ethiopia | 00, 05 | Zambia | 92, 96*, 01, 07 | Paraguay | 90 |
| Gabon | 00 | Zimbabwe | 94*, 99, 06 | Peru | 92, 96*, 00, 04 |
| Ghana | 93, 98*, 03, 08 | | | | |
| Guinea | 99, 05 | Bangladesh | 93, 96, 99, 04, 07 | Albania | 08 |
| Kenya | 93, 98, 03, 08 | Cambodia | 00, 05, 10 | Armenia | 00, 05 |
| Lesotho | 04, 09 | India | 92, 98, 05 | Azerbaijan | 06 |
| Liberia | 07 | Indonesia | 91, 94, 97, 02, 07 | Egypt | 92, 95*, 00, 03, 05, 08 |
| Madagascar | 92, 97*, 03, 08 | Maldives | 09 | Jordan | 97*, 02, 07, 09 |
| Malawi | 92, 00, 04, 10 | Nepal | 96*, 01, 06 | Kazakhstan | 95, 99 |
| Mali | 95*, 01, 06 | Pakistan | 90, 06 | Kyrgyz Rep. | 97 |
| Mozambique | 97*, 03, 09 | Philippines | 93, 98*, 03, 08 | Moldova | 05 |
| Namibia | 92, 00, 06 | Timor Leste | 09 | Morocco | 92, 03 |
| Niger | 92, 98, 06 | Vietnam | 97, 02 | Turkey | 93, 98*, 03 |
| Nigeria | 90, 99*, 03, 08 | | | Ukraine | 07 |
| | | | | Uzbekistan | 96 |

Notes: Years denote date when survey began; data collection often continues into the following year. 9x numbers are 199x, 0x numbers are 200x, 10 is 2010. (*) Surveys with wage income data. Chad 96, Nigeria 90 and Paraguay 90 do not have the necessary product data for the estimation of inequality, but have data on wages or migration.

TABLE B1
 10,000 SIMULATED REPLICATIONS OF THE ESTIMATION PROCEDURE
 (regression of 65 country mean consumption gaps on estimated residence probabilities)

| | Coefficients | | | | Standard errors | | | | |
|--|------------------|----------------------|------------------|----------------------|-----------------|------|------|------|------|
| | underlying value | mean estimated value | actual variation | mean estimated value | | | | | |
| constant | 0 | -.009 | 2.90 | 2.89 | | | | | |
| $\Pi_{16}^U - \bar{\Pi}$ | 13.45 | 13.45 | 5.01 | 4.99 | | | | | |
| $\Pi_0^U - \bar{\Pi}$ | -13.45 | -13.47 | 4.23 | 4.19 | | | | | |
| Size of $H_0: \alpha = 0 \text{ \& } \beta_1 = -\beta_2$ | | | | | | | | | |
| Nominal | .010 | .050 | .100 | .250 | .500 | .750 | .900 | .950 | .990 |
| Actual | .009 | .049 | .102 | .254 | .502 | .751 | .900 | .949 | .991 |