**CGEB WORKING PAPER SERIES**

Center for Global Economy and Business
NYU Stern School of Business

# SELECTION, AGRICULTURE, AND CROSS-COUNTRY PRODUCTIVITY DIFFERENCES

David Lagakos
Arizona State Univerisy

Michael E. Waugh
New York University

**Working Paper No. 05**
**March 2012**



Center for Global Economy and Business
NYU Stern School of Business
44 West 4th Street, Suite 7-190
New York, NY 10012
http://www.stern.nyu.edu/cgeb

# Selection, Agriculture, and Cross-Country Productivity Differences

David Lagakos

Arizona State University

Michael E. Waugh

New York University

First Version: February 2009

This Version: November 2011

ABSTRACT

Cross-country labor productivity differences are larger in agriculture than in non-agriculture. We propose a new explanation for these patterns in which the self-selection of heterogeneous workers determines sector productivity. We formalize our theory in a general equilibrium Roy model with preferences featuring a subsistence food requirement. In the model, subsistence requirements induce workers that are relatively unproductive at agriculture work to nonetheless select into the agriculture sector in poor countries. When parameterized, the model predicts that agriculture productivity differences are twice as large as those in non-agriculture even when economies differ by an economy-wide efficiency term that affects both sectors uniformly.

Email: lagakos@asu.edu, mwaugh@stern.nyu.edu. For helpful comments we thank Tasso Adamopoulos, David Atkin, Francesco Caselli, Berthold Herrendorf, Ayse Imrohoroglu, Narayana Kocherlakota, Erzo Luttmer, Ellen McGrattan, Guillermo Ordoñez, Diego Restuccia, Andrés Rodríguez-Clare, Richard Rogerson, Jim Schmitz, Todd Schoellman, Jon Vogel and seminar participants at ASU, Boston University, Clemson, the Federal Reserve Banks of Minneapolis and San Francisco, Notre Dame, Ohio State, Iowa, Syracuse, Toronto, UCLA, USC, York, the 2009 Midwest Macro Meetings, the 2009 NBER Summer Institute Economic Growth Meeting, the SED (Istanbul), and the 2009 SITE Growth and Development session. This version supersedes an earlier version with the title "Specialization, Agriculture, and Cross-Country Productivity Differences".

## 1. Introduction

Cross-country labor productivity differences are much larger in agriculture than in the non-agriculture sector (Caselli (2005), Restuccia, Yang, and Zhu (2008)). Because developing countries have most of their workers in agriculture, their low productivity in agriculture accounts for nearly all of their low productivity in the aggregate. This implies that understanding why productivity differences in agriculture are so large compared to those of the non-agriculture sector is at the heart of understanding world income inequality.[1]

In this paper we propose a new explanation for these productivity patterns in which the self-selection of heterogeneous workers determines sector productivity. We start from the well-known idea that in poor countries, where economy-wide efficiency is low, most people must work in the agriculture sector in order to satisfy subsistence consumption needs. This is what Schultz (1953) famously called the "food problem." Our insight is that precisely because the majority of workers in poor countries are employed in agriculture, many of these workers must be relatively unproductive at agricultural work. In contrast in rich countries, where economy-wide efficiency is high, those few workers selecting into agriculture must be those who are relatively most productive at agriculture work. Thus, two economies that differ in economy-wide efficiency will have even larger measured differences in agriculture productivity. By the same mechanism, they will have even smaller measured non-agriculture productivity differences.

Our theory has two main ingredients. The first is non-homothetic preferences, and in particular a subsistence consumption requirement in the agricultural good. This leads to an income elasticity of demand for agricultural goods less than one. The second ingredient is heterogeneity in individual (worker) productivity in each sector, combined with the assumption that workers choose where to supply their labor. This is the Roy (1951) model of self-selection based on comparative advantage. We combine these features into a two-sector general equilibrium version of the Roy model. Countries differ only in an economy-wide efficiency term; preferences and the distribution of individual productivity are taken to be identical across countries.

Within this economic environment, we provide a general condition on the heterogeneity in individual productivity that leads to productivity differences that are larger in agriculture than non-agriculture when countries differ only by an economy-wide efficiency term. The key condition is simple and economically meaningful: that comparative advantage aligns with absolute advantage. As long as workers who have a comparative advantage in a given sector have an absolute advantage (on average) in that sector, then our model qualitatively replicates the larger cross-country productivity differences in agriculture and smaller differences in non-agriculture.

[1] Versions of this argument have been made by Caselli (2005), Restuccia, Yang, and Zhu (2008), Chanda and Dalgaard (2008), Vollrath (2009).

To measure the quantitative importance of selection in explaining the sector productivity patterns at hand, we make flexible parametric assumptions on the distribution of individual productivity. In particular, we assume that sector productivities are drawn from dependent Fréchet distributions, where the dependence is captured parsimoniously using a copula. These assumptions allow us to calibrate the distribution parameters using simple moments from the cross-sectional distribution of wages in the United States, namely the variance of log wages in each sector and the ratio of sector average wages.

Our main quantitative finding is that selection leads to roughly twice as much variation in agriculture productivity than in non-agriculture productivity across rich and poor countries. In the data, there is just over ten times as much productivity variation in agriculture than non-agriculture. This implies that selection accounts for around one-fifth of the greater cross-country productivity variation in agriculture. We reach this conclusion using our benchmark model, which features only labor as an input to production, and our main quantitative experiment, which varies economy-wide efficiency to match the difference in aggregate GDP per worker between the ninetieth and tenth percentile countries of the world income distribution, and then computes the model's implications for sector productivity differences.

We extend the model to include capital and land, and find that these forces increase the overall explanatory power of the model while leaving the importance of the selection channel largely unchanged. When calibrated, the extended model produces four times as much variation in productivity in agriculture as non-agriculture. The improved performance comes from the well known role that land plays as a fixed factor in agriculture (see e.g. the models of Restuccia, Yang, and Zhu (2008), Adamopoulos and Restuccia (2010) and Herrendorf and Teixeira (2011)). While the importance of selection is similar in magnitude as in the benchmark model, decomposing the results into the contribution from land versus selection shows that selection is as important or more than the effects from land alone.

We find that, in either version of the model, the quantitative predictions are consistent with other important features of the data not targeted directly. In particular, both versions predict a large wage gap between agriculture and non-agricultural workers, as in the data. This is in contrast to other papers in the literature, which reconcile this wage gap using some sort of exogenous barrier to workers moving out of agriculture (e.g. Caselli and Coleman (2001), Restuccia, Yang, and Zhu (2008), Adamopoulos and Restuccia (2010), Tombe (2011) and Herrendorf and Teixeira (2011)). Both versions of our model are also quantitatively consistent with the higher employment shares in agriculture in poor countries, and the higher relative prices of agriculture goods in poor countries.

To illustrate how our theory works in practice, we provide one concrete example of how agriculture workers in developing countries are on average less productive at agriculture work than their counterparts in rich countries. Specifically, we cite evidence that women are less productive at men on average in agricultural work, and use cross-country data to document that women form a much larger fraction of all agricultural workers in developing countries than they do in richer countries. Putting these together implies that poor countries have lower measured productivity in agriculture in part because they employ more workers with relatively low productivity in at agriculture work, just as our theory predicts.

Our paper is the first to propose and assess the role of selection in understanding why productivity differences in agriculture are so much larger than in other sectors. This mechanism is distinct from previous explanations in the literature, most of which focus on distortions that are specific to the agriculture sector. For example, Restuccia, Yang, and Zhu (2008) argue that the larger productivity differences in agriculture are due partly to barriers to the adoption of intermediate goods in agriculture; Adamopoulos and Restuccia (2010) focus on the role of policies that misallocate farm land in developing countries.

One key difference is that our paper can in part reconcile the sector productivity patterns even when distortions in poor countries do not disproportionately affect agriculture. Instead, they can arise from general factors, such as weak institutions, as emphasized by e.g. Hall and Jones (1999) and Acemoglu, Johnson, and Robinson (2001, 2002), plus the selection channel studied in the current paper. The policy implications of our paper differ as well. The emphasis on agriculture-specific distortions in the previous literature suggests that the focus in poor countries should be on removing distortions that are specific to agriculture. Under our view, this implication may be misguided.

Still, it is worth emphasizing that our explanation and previous ones in the literature are complements, in the sense that selection forces along with distortions of either a general or sector-specific nature lead to measured productivity differences in agriculture that are larger than they otherwise would be, and non-agriculture differences that are smaller than they otherwise would be. This observation is important because it is unlikely that one story alone can completely explain why there is so much more productivity variation in agriculture than in other sectors, given the enormous magnitude of the difference.

## 2. Motivating Evidence

In this section, we review the evidence that cross-country labor productivity differences are much larger in agriculture than in the non-agriculture sector. We then provide new calculations, and discuss existing evidence, suggesting that these sector labor productivity differences largely reflect sector TFP differences.

**Table 1: Sector Labor Productivity and Employment Shares**

| | Agriculture | Aggregate | Non-Agriculture | Ag/Non-Ag Ratio |
|---|---|---|---|---|
| 90-10 Labor Productivity Differences | 45 | 22 | 4 | 10.7 |
| Employment Share 90th Percentile Country | 3 | — | 97 | — |
| Employment Share 10th Percentile Country | 78 | — | 22 | — |

**Source:** Caselli (2005)

Table 1 reproduces the findings of Caselli (2005), who constructs Purchasing Power Parity (PPP)-adjusted measures of labor productivity in the agriculture and non-agriculture sectors of 79 countries. His calculations combine PPP-adjusted GDP per worker data from the Penn World Tables with PPP-adjusted agriculture value-added-per-worker data from the Food and Agriculture Organization (FAO) constructed by Rao (1993).

The first row of Table 1 reports that the ratio of aggregate output per worker in the 90th to 10th percentile of the world income distribution is a factor 22. In agriculture, this ratio is a factor 45, while in non-agriculture it is a factor of just 4. Thus, agriculture productivity differences across countries are much larger than those of non-agriculture.[2] The last column shows that the ratio of agriculture to non-agriculture productivity differences is 10.7. In other words, there is more than ten times as much variation in agriculture productivity across countries than there is in non-agricultural productivity.

The second and third rows report the percent of employment in agriculture in the 90th and 10th percentile countries. In the 90th percentile country, just 3 percent of labor is in agriculture, while the other 97 percent is in the non-agricultural sector. In the 10th percentile country, in contrast, 78 percent of workers are in agriculture, compared to 22 percent in non-agriculture. In short, a key distinction between rich and poor countries is that agriculture employs most people in the poorest countries and virtually nobody in the richest countries.

Simple accounting exercises show that the divide between agriculture and non-agriculture accounts for much of aggregate productivity differences. Caselli (2005) computes the hypothetical 90-10 ratio of aggregate output per worker by giving the agricultural productivity level of the 90th percentile country to all countries. He finds that the 90-10 ratio would be a factor of 1.6,

[2] In independent work, Restuccia, Yang, and Zhu (2008) arrive at a very similar conclusion.

down from the actual factor of 22! Similarly, by hypothetically giving an agricultural employment share of 3 percent, as in the 90th percentile country, to all countries, the 90-10 ratio would be just a factor 4.2.

One simple explanation of these sector labor productivity patterns is that developing countries use much less capital per worker in agriculture than in rich countries, and use only modestly less capital per worker in non-agriculture. The main challenge to testing this hypothesis is the limited data on capital stocks by sector across countries. Caselli (2005) addresses this limitation by making the plausible assumption that rates of return to capital are equated across sectors, and then using aggregate capital stock data to allocate capital to each sector. For a set of 65 countries for which comparisons can be made, he finds that capital explains 15 percent of cross-country productivity differences in agriculture, and 59 percent in non-agriculture. Thus, his calculations suggest capital differences are indeed important in both sectors, but there are still bigger residual productivity differences in agriculture even after taking capital into consideration.

To complement these findings, we conducted our own accounting exercises for a smaller set of countries using data on agricultural capital stocks constructed by Butzer, Mundlak, and Larson (2010). These data contain the values of machinery, equipment, livestock and tree stock used in agriculture production in a set of 28 countries from all income levels. As we detail in Appendix B, we combine these data with estimates of the aggregate capital stocks constructed by the PWT to create estimates of the non-agricultural capital stocks in each country. The resulting sector capital data allow us to conduct accounting exercises in the same manner as Caselli (2005).

We find that using these new data, capital accounts for 22 percent of cross-country productivity differences in agriculture, and 29 percent in non-agriculture. Thus, these exercises largely corroborate the findings of Caselli (2005). While both sets of calculations have their limitations, both suggest that capital-per-worker differences are important in both sectors, but unlikely to be the main cause of the larger differences in agriculture labor productivity across countries. In this sense, our findings are consistent with those of Chanda and Dalgaard (2008) and Vollrath (2009), who conclude that low agriculture (and aggregate) labor productivity in the developing world largely reflects their low TFP in agriculture.

## 3. Model of Agricultural and Non-Agricultural Productivity

In this section we formalize our model economy and characterize its equilibrium. The model predicts, under conditions that we describe, that exogenous differences in economy-wide efficiency across economies lead to even larger differences in agriculture productivity, and smaller differences in non-agriculture productivity.

### 3.1. Preferences and Endowments

There are measure one of workers, indexed by $i$, who differ in productivity, as explained below. Preferences are given by

$$U^i = \log(c_a^i - \bar{a}) + \nu \log(c_n^i), \tag{1}$$

where $c_a^i$ is agricultural good (food) consumption, $c_n^i$ is non-agricultural good consumption, $\bar{a}$ is a parameter representing a subsistence consumption requirement, and $\nu$ governs the relative taste for non-agriculture consumption. These "Stone-Geary" preferences ensure that Engel's Law holds, namely that the income elasticity of demand for food is less than one.

Each worker is endowed with one unit of time which she supplies inelastically to the labor market. Each worker is also endowed with a vector of "individual productivities," denoted $\{z_a^i, z_n^i\}$, which represent the efficiency of one unit of labor in sectors $a$ and $n$. Individual productivities are drawn from a distribution $G(z_a, z_n)$ with support on the positive reals. The budget constraint of worker $i$ is

$$p_a c_a^i + c_n^i \leq y^i, \tag{2}$$

where $y^i$ is labor income (described in more detail below), $p_a$ is the relative price of agriculture, and the non-agricultural good is taken to be the numeraire.

### 3.2. Production

There is a competitive market in both sectors, and each has its own aggregate production function. Both sector technologies are freely available and operated by competitive entrepreneurs. The technologies are given by

$$Y_a = AL_a \quad \text{and} \quad Y_n = AL_n, \tag{3}$$

where $A$ is exogenous and captures "economy-wide efficiency" of production, and $L_a$ and $L_n$ represent the total number of effective labor units employed in the two sectors. Economies differ only in $A$, and we assume that each economy is closed.[3]

Let $\Omega^a$ and $\Omega^n$ denote the sets of workers choosing to work in agriculture and non-agriculture. The sector aggregate labor inputs $L_a$ and $L_n$ are defined as

$$L_a \equiv \int_{i \in \Omega^a} z_a^i \, dGi \quad \text{and} \quad L_n \equiv \int_{i \in \Omega^n} z_n^i \, dGi$$

[3] In Section 5.4 we discuss how opening the economy to trade would change our results. See also Gollin, Lagakos, and Waugh (2011b) for more on the open-economy implications of selection in multi-sector models, and Tombe (2011) for a theory of the lack of agriculture imports by developing countries.

and represent the sum of all individual productivity employed in the sectors. The total number of workers in each sector are defined as

$$N_a \equiv \int_{i \in \Omega^a} dGi \quad \text{and} \quad N_n \equiv \int_{i \in \Omega^n} dGi.$$

### 3.3. Optimization and Equilibrium

An equilibrium of the economy consists of a relative agriculture price, $p_a$, wages per efficiency unit of labor in each sector, $w_a$ and $w_n$, and allocations for all workers, such that all workers optimize and both labor markets and output markets clear. Measured labor productivity in equilibrium is denoted by $Y_a/N_a$ in agriculture and $Y_n/N_n$ in non-agriculture, and represent the physical quantity of output produced per worker in each sector.

Workers take prices and wages as given when they optimize. The problem for a worker is first to choose which sector to supply their labor, and then to maximize her utility, (1), subject to her budget constraint, (2). Because of competition, the wages per efficiency unit of labor are

$$w_a = p_a A \quad \text{and} \quad w_n = A.$$

A simple cutoff rule in relative individual productivity, or *comparative advantage*, determines the optimal occupational choice for each worker. Working in non-agriculture is optimal for worker $i$ if and only if

$$\frac{z_n^i}{z_a^i} \geq p_a. \tag{4}$$

Thus, the workers that enter non-agriculture are those whose productivity is sufficiently high relative to their productivity in agriculture. Labor income under the optimal sector choice is defined as $y^i \equiv \max\{z_a^i w_a, z_n^i w_n\}$.

The remainder of the worker's problem is standard, and optimal demands are:

$$c_a^i = \frac{y^i + \bar{a} p_a \nu}{p_a(1+\nu)} \quad \text{and} \quad c_n^i = \frac{\nu(y^i - \bar{a} p_a)}{1+\nu}. \tag{5}$$

Due to the subsistence consumption requirement, workers consume relatively more agricultural goods when their income is lower. The lower is $\nu$, the higher is the ratio of agriculture to non-agriculture consumption.

### 3.4. Qualitative Features of Equilibrium

We now show that, in equilibrium, economy-wide efficiency determines the relative price of agriculture, which in turn determines the selection of workers and productivity in each sector.

### The Relative Price of Agriculture Goods is Higher in Poorer Economies

The first important result is that in equilibrium, the relative price of agriculture is higher in economies with lower economy-wide efficiency. We formalize this result as:

**Proposition 1** *Consider two economies, rich and poor, with efficiency terms $A^R$ and $A^P$ such that $A^R > A^P$. In equilibrium, the relative price of agriculture is higher in the poor economy: $p_a^P > p_a^R$.*

The intuition is that a higher price of agricultural goods is needed in the poor economy in order to induce workers to work in the agriculture sector. To see this, let $p_a^R$ be the equilibrium relative price in rich economy. If $p_a^R$ were the equilibrium price in the poor economy as well, then by (4), the sector labor-supply cutoffs would be the same in both countries, and so would the share of workers in agriculture. But because of the subsistence consumption requirement, the poor economy demands a much larger fraction of agricultural goods, and thus there would be excess demand for food in the poor economy. It follows that the relative price of agriculture could not be the same in the two economies, and in fact must be higher in the poor economy.

### Individual Productivity Distribution and Sectoral Productivity Differences

We now turn to the link between the distribution of individual productivity and sector aggregate productivity in equilibrium. Proposition 2 describes conditions on the individual productivity distribution that are sufficient for economy-wide efficiency differences to lead to larger differences in agriculture labor productivity and smaller differences in non-agriculture labor productivity.

**Proposition 2** *Consider two economies with efficiency terms $A^R$ and $A^P$ such that $A^R > A^P$. Let the individual productivity distribution be such that $E(z_a|z_a/z_n > x)$ and $E(z_n|z_n/z_a > x)$ are increasing in $x$. Then equilibrium sector labor productivities are such that*

$$\frac{Y_a^R/N_a^R}{Y_a^P/N_a^P} > \frac{A^R}{A^P} \quad \textit{and} \quad \frac{Y_n^R/N_n^R}{Y_n^P/N_n^P} < \frac{A^R}{A^P}.$$

Intuitively, Proposition 2 says that as long as workers who have a comparative advantage in a given sector have an absolute advantage (on average) in that sector, then productivity differences will be larger in agriculture than non-agriculture across the two economies. The reason is as follows. As $A$ rises, the relative price of agriculture falls (by Proposition 1), and only workers with a greater comparative advantage in agriculture (i.e. a higher $z_a/z_n$ ratio) choose to work in agriculture. Then, since workers with a greater comparative advantage also have a greater absolute advantage, it follows that agriculture sector productivity increases. The second part of Proposition 2 says that, for a similar reason, non-agricultural productivity differences are smaller than $A$ differences if workers with a greater comparative advantage in non-agriculture have a higher expected productivity in that sector.

Note that both heterogeneity in worker productivity and non-homothetic preferences are necessary for Proposition 2 to hold. When all workers are identical in productivity, then changes in $A$ induce changes in the share of workers in agriculture but (trivially) do not change the average individual productivity by sector. When preferences are homothetic, relative prices and hence the allocation of workers by sector are independent of $A$. Thus, in each sector, average individual productivity is identical across countries.

At least one of the conditions of Proposition 2 must hold (see Heckman and Honoré (1990)). Thus, at the very least, our theory qualitatively delivers productivity differences in one sector that differ from the aggregate in a way consistent with the data. Of course, it can also explain the patterns of both sectors. We now turn to an example where both conditions on the individual productivity distributions are satisfied, and in which simple analytical expressions help provide intuition for how the model works.[4]

### 3.5. Simple Analytical Model: Independent Fréchet Individual Productivities

In this section, we illustrate the mechanics of the theory using a simple analytical version of the model which assumes independent Fréchet distributions on individual productivity. This example helps demonstrate how the size of the mechanism's effects depend on (i) the variance in individual productivity, and (ii) differences in the sector employment shares across the economies being compared. Furthermore, this example is a special case of the individual productivity distribution used for quantitative analysis in Section 4.

**Assumption 1** *Let $z_a$ and $z_n$ be drawn independently from Fréchet distributions:*

$$G(z_a) = e^{-z_a^{-\theta}} \quad \textit{and} \quad G(z_n) = e^{-z_n^{-\theta}}.$$

[4]One can show that both conditions of Proposition 2 hold whenever individual productivities are independent across sectors and distributed log-concave in each sector. Prominent examples are Normal, Pareto and Uniform distributions. However, none has the analytic tractability productivity of independent Fréchet distributions that we focus on below.

The parameter $\theta$ controls the dispersion of individual productivity in each sector, with a smaller $\theta$ implying more productivity dispersion across individuals and a higher $\theta$ meaning less dispersion.[5] This distributional assumption conveniently relates equilibrium employment shares in agriculture, the relative price of agriculture, and parameter $\theta$. The equilibrium share of workers in agriculture is

$$\pi_a = \text{Prob}\left\{Az_n^i \leq p_a A z_a^i\right\} = \frac{1}{p_a^{-\theta}+1}. \tag{6}$$

By (6), one can see that as $p_a$ rises, the share of workers in agriculture rises as well. Furthermore, the responsiveness of the share of workers in agriculture to $p_a$ depends on the productivity-dispersion parameter $\theta$. Manipulating (6), and a similar equation for non-agriculture, yields a log-linear relationship in the ratio of the agriculture to non-agriculture worker shares $(\pi_n)$ and the relative price of agricultural goods:

$$\log\left(\pi_a/\pi_n\right) = \theta \log(p_a). \tag{7}$$

Intuitively, with a low $\theta$, meaning high productivity dispersion across workers, large changes in the relative price of agriculture are needed to induce workers to switch sectors. On the other hand, a higher $\theta$, meaning small productivity dispersion, implies that only small changes in the relative price of agriculture are needed to induce workers to switch sectors.

Both conditions on the productivity distribution in Proposition 2 hold in this example. That is, expected worker productivity in a sector is larger when its workers have a greater comparative advantage in that sector. To see this note that expected individual productivity in the two sectors are

$$E(z_a|z_a/z_n > 1/p_a) = \gamma\, \pi_a^{\frac{-1}{\theta}}, \text{ and } E(z_n|z_n/z_a > p_a) = \gamma \pi_n^{\frac{-1}{\theta}}, \tag{8}$$

where the constant $\gamma$ is the Gamma function evaluated at $(\theta-1)/\theta$. Equation (8) relates expected individual productivity to the share of workers in a sector and through equation (7) the relative price. A decrease in the relative price of agriculture decreases the share of workers in agriculture. This then leaves a more selected set of workers in agriculture with higher average agricultural productivity. Similarly, because the share of workers increases in non-agriculture, non-agriculture productivity decreases. The magnitude of these changes depends on the parameter $\theta$.

Differences in $A$ across economies will lead to relative price differences (Proposition 1). This then leads to differences in employment shares (equation (7)) and hence to larger productivity

[5] This distribution has been used by Eaton and Kortum (2002) and others to analytically solve multi-country Ricardian models of international trade. To our knowledge, we are the first to exploit the analytical properties of this distribution to study the Roy model.

differences in agriculture and smaller ones in non-agriculture across economies. We formalize this as:

**Corollary 1** *Consider two economies with efficiency terms $A^R$ and $A^P$, such that $A^R > A^P$, and let Assumption 1 hold. Then, the ratios of sector labor productivities are*

$$\frac{Y_a^R/N_a^R}{Y_a^P/N_a^P} = \left(\frac{\pi_a^P}{\pi_a^R}\right)^{\frac{1}{\theta}} \left(\frac{A^R}{A^P}\right) > \frac{A^R}{A^P} \quad \textit{and} \quad \frac{Y_n^R/N_n^R}{Y_n^P/N_n^P} = \left(\frac{\pi_n^P}{\pi_n^R}\right)^{\frac{1}{\theta}} \left(\frac{A^R}{A^P}\right) < \frac{A^R}{A^P}. \tag{9}$$

Dispersion in individual productivity controls the magnitude of the sector productivity difference from the aggregate. A lower $\theta$ leads agriculture productivity to be larger than the aggregate (since $\pi_a^R < \pi_a^P$ in equilibrium). As $\theta$ approaches infinity, heterogeneity in individual productivity disappears, selection effects are diminished, and the ratio of agriculture productivity converges downward toward the aggregate productivity ratio. A similar argument illustrates that the non-agriculture productivity difference is smaller than the difference in $A$, with the magnitude of the difference again shrinking to zero as individual productivity dispersion is reduced to zero.

With large cross-country differences in employment shares these effects can be potent. Consider for example the countries at the 10th and 90th percentile of the world income distribution, which have 78 and 3 percent of their workforce in agriculture. With a $\theta$ of 5, these differences in agriculture employment shares amplify the underlying $A$ differences by a factor of two. In contrast, for countries that do not differ dramatically in agriculture employment shares, these effects will be more modest.

## 4. Quantitative Analysis

We now present a richer model that we calibrate and use to assess the quantitative importance of the mechanism. This richer model differs from the simple analytical model of the previous section by allowing for correlation between individual productivity draws and different degrees of productivity dispersion in the two sectors.

Introducing these richer features is important for two reasons. First, it allows our theory to fail. In particular, there is nothing inherent in the richer model that assures both assumptions on the individual productivity distribution in Proposition 2 hold. Whether both conditions hold will be dictated by the data in the calibration. Second, it allows for greater flexibility in matching the data, and hence a more accurate assessment of the quantitative importance of the theory.

### 4.1. Dependent Fréchet Individual Productivity Distribution

We set the joint distribution of individual productivities to be

$$G(z_a, z_n) = C[F(z_a), H(z_n)],$$

$$\text{where} \qquad F(z_a) = e^{-z_a^{-\theta_a}} \text{ and } \quad H(z_n) = e^{-z_n^{-\theta_n}},$$

$$\text{and} \qquad C[u, v] = \frac{-1}{\rho} \log\left\{1 + \frac{(e^{-\rho u} - 1)(e^{-\rho v} - 1)}{e^{-\rho} - 1}\right\}.$$

The function $C[F(z_a), H(z_n)]$ is a *Frank copula*, which allows for dependence between draws from distributions $F(z_a)$ and $H(z_n)$.[6] The parameter $\rho \in (-\infty, \infty)\backslash\{0\}$ determines the extent of dependence, with a positive (negative) value of $\rho$ representing positive (negative) dependence between the draws.[7] The marginal distributions themselves are Fréchet, with dispersion parameters $\theta_a$ and $\theta_n$ and the means are normalized to be the same. The lower are $\theta_a$ and $\theta_n$, the higher is the variation in individual productivity in agriculture and non-agriculture.

This parameterization introduces two dimensions of richness relative to the analytical example of Section 3.5. First, individual productivity draws are no longer independent across sectors. This allows for characteristics that make a worker more productive in both types of activity. Second, dispersion in individual productivity is no longer the same in each sector. Since non-agriculture work is a stand-in for many different types of economic activities, one might expect that individual productivity dispersion is larger in non-agriculture than in agriculture. This parameterization allows for this possibility.

We choose this functional form for our quantitative analysis for several reasons. First, it allows for a transparent calibration of the distribution parameters, while also allowing for dependence and differing sector dispersion in individual productivity. As we show in the following section, the three parameters of the distribution ($\theta_a$, $\theta_n$ and $\rho$) are disciplined by three simple moments calculated from a single cross section of wages.[8]

[6] A copula is a function that allows for the creation of multivariate distributions out of arbitrary univariate distributions; see e.g. Nelsen (2006). The Frank copula generates dependence between draws that is *radially symmetric*, i.e. not systematically stronger when closer to the right or left tails of the distribution. Other copulae, such as the Clayton or the Gumbel copula, do not have this feature.

[7] When $\rho = 0$, $C[u, v]$ is defined as $u \cdot v$.

[8] Individual productivity distributions in the Roy model cannot be identified from cross-sectional wage data without making assumptions about the functional form of the distributions (see Heckman and Honoré (1990)). Because one observes only the maximum of each worker's draws, but not both draws themselves, if individual productivity distributions are allowed to take on an arbitrary form, there are many distributions that can generate a given set of observations on wages and sector choices by individuals.

Second, the choice of Fréchet distributions for individual productivity in each sector contains a sensible economic interpretation, which is as follows. The Fréchet distribution is an extreme-value distribution, representing the distribution of the maximum of independent draws from some underlying distribution.[9] Thus, the draw $z_n^i$ can be thought of as the maximum of worker $i$'s individual productivity draws in a large set of distinct non-agricultural tasks. A similar interpretation can be given to $z_a^i$.[10]

### 4.2. Calibration of Individual Productivities

To calibrate the individual-productivity distribution parameters, our strategy uses cross-sectional wage data from the United States. Formally, we jointly calibrate $\theta_a$, $\theta_n$ and $\rho$ to match three moments: the standard deviations of log wages in agriculture and non-agriculture (adjusted as we describe below) and the ratio of average wages in agriculture and non-agriculture.

While all three parameters are jointly determined, each has an intuitive relationship with one of the moments picked. The parameters $\theta_a$ and $\theta_n$ are disciplined by cross-sectional wage variation in agriculture and non-agriculture. Because a worker's wage in the model equals the value of her marginal product, variation in individual productivity maps into variation in wages across workers.

The dependence parameter $\rho$ is disciplined by the ratio of average wages in agriculture to average wages in non-agriculture, with a lower ratio implying a higher $\rho$. The intuition is as follows. For high values of $\rho$, workers tend to get either two high draws or two low ones. Because of the higher variance in non-agricultural productivity (implied by the calibration procedure, as we explain below), those with the high draws are more likely to have a comparative advantage in non-agriculture. This implies that most of the high-wage workers are in the non-agricultural sector, and that the ratio of average wages is low. For low values of $\rho$, in contrast, each sector employs worker with high sector-specific skills, and higher wage individuals are more equally distributed across sectors. Hence, the ratio of average wages is higher.

Our cross-sectional wage data comes from the U.S. Current Population Survey (CPS) for 2010, which is the most recent year available. Our sample includes all individuals who have non-missing data on income and hours worked, including both self-employed and salaried workers. We calculate each individual's wage as the sum of salary income, business income and

[9]By the extreme value theorem, the maximum of independent draws from any distribution converges in distribution (once properly normalized) to one of three extreme value distributions: the Fréchet, the Gumbel, or the Weibull.

[10]Yet another advantage of Fréchet distributions is that they produce wage distributions with fat right tails, as in the data, while other prominent distributions fail in this dimension. For example, we find that a version of our model with log normal individual-productivity distributions generates tails that are too thin compared to the data. Heckman and Sedlacek (1985) arrive at a similar conclusion. Details of our calculations are available on request.

farm income in the previous year divided by hours worked in the previous year. We restrict the sample to include only those earning at least the Federal minimum wage. We define agricultural workers to be those whose primary industry of employment is agriculture, forestry or fishing, and non-agricultural workers to be all other workers.[11]

We then adjust the variation in wages to keep only the "permanent" component of wages, rather than the "transitory" component. We do so because wage variation in the model arises only from productivity differences across workers, whereas wage variation in the data may include other factors unrelated to productivity. Following Guvenen and Kuruscu (2009), we subtract off estimates of the transitory component of wages using the estimates of Guvenen (2009), who calculates that the variance of the transitory component of log wages is $0.14$. We end up with adjusted standard deviations of log wages in agriculture and non-agriculture of $0.33$ and $0.46$, which we target in our calibration.

The ratio of average wages in agriculture to average wages in non-agriculture is the final moment we target in our calibration. Using the CPS data, we calculate this ratio to be 0.69.[12]

These moments imply parameter values of $\theta_a = 5.5$, $\theta_n = 2.8$ and $\rho = 2.2$. The estimates of $\theta_a$ and $\theta_n$ mean that there is more variation in individual productivity in non-agriculture work than in agricultural work, which seems reasonable given that non-agriculture work encompasses more types of economic activities. While $\rho$ itself is hard to interpret, the associated Spearman rank correlation coefficient is 0.24 and the linear correlation coefficient is 0.31. This suggests that there is a modest amount of positive correlation in individual productivities: if an worker is productive in one sector, she is likely to be productive in the other as well.

### 4.3. Calibration of Preference Parameters

For the preference parameters, we pick $\nu$ and $\bar{a}$ to jointly match two moments from U.S. data. The first moment we target is the fraction of workers in agriculture from U.S. data, which is just below two percent. The second is a long-run agriculture expenditure share of $0.5$ percent, which has been used by others in the literature, in particular Restuccia, Yang, and Zhu (2008).

The resulting parameter $\bar{a}$ is consistent with independent estimates of the size of the subsistence consumption requirement in developing countries. Rosenweig and Wolpin (1993) and Atkeson and Ogaki (1996), both of which use panel data from a sample of rural households in

[11] We find that our results are similar when using other plausible sample selection criteria, such as that of Heathcote, Perri, and Violante (2009), who further restrict the sample to include only those aged between 25 and 60 years old and working at least 35 hours per week. See Appendix C for more on the cross-sectional data we employ.

[12] If the cost of living is lower in rural areas, then the ratio of real average wages would be higher. Adjusting for this implies that we would infer a lower $\rho$ in our calibration, which would strengthen our results. We skip such an adjustment however, as data on prices in rural areas in the United States are not systematically collected.

**Table 2: 90-10 Productivity Differences, Data and Benchmark Model**

| | Agriculture | Aggregate | Non-Agriculture | Ag/Non-Ag Ratio |
|---|---|---|---|---|
| Data | 45 | 22 | 4 | 10.7 |
| Model | 32 | 22 | 14 | 2.3 |
| Without Selection | 20 | 20 | 20 | 1.0 |

**Note:** The aggregate productivity difference is the ratio of GDP per worker expressed at Gheary-Khamis international prices between the 90th and 10th percentile countries.

India, estimate a subsistence consumption need of around 33 percent of the average income of Indian villagers. When we compute the the subsistence consumption requirement in our model economy with $A$ calibrated to mimic India's per capita GDP relative to the U.S. we find that $\bar{a}$ is 30 percent of average income.

### 4.4. Quantitative Predictions for Sector Productivity Differences

To explore the quantitative implications of our model, we perform the following experiment. Beginning with a value of $A$ normalized to one for the benchmark economy (calibrated to the U.S.), we lower $A$ to match GDP per worker for a country in the 90th percentile of the income distribution, and then for a country in the 10th percentile. We then compare the model's predictions for sector labor productivity in the 10th and 90th percentile countries to those of the data.

Table 2 shows the the results of the experiment. By construction, the gap in aggregate labor productivity is a factor 22 in both the model and data. This gap is generated with a difference in $A$ of 20.[13] At the sector level, the model predicts a factor 32 difference in agriculture productivity, and a factor 14 difference in non-agriculture. In the data, the differences are a factor 45 in agriculture and a factor 4 in non-agriculture. Thus, the selection channel in the model generates quantitatively large differences between sector and aggregate productivity differences, but not quite as large as in the data.

To provide a more concrete metric for the overall quantitative importance of selection, the last column of Table 2 shows the ratio of the productivity differences in agriculture to those of

[13] Aggregate labor productivity is expressed as GDP per worker at Gheary-Khamis international prices. The difference between the $A$ difference and GDP per worker difference comes from the higher relative price of agriculture goods in the poorer countries, which our model generates endogenously as workers with progressively lower agriculture productivity are induced to enter agriculture.

**Table 3: Selection and Individual Productivity**

| Country | Agriculture, $\frac{E(z_a\|z_a/z_n>p_a)}{E(z_a)}$ | Non-Agriculture, $\frac{E(z_n\|z_n/z_a>p_a)}{E(z_n)}$ |
|---|---|---|
| 90th Percentile | 1.62 | 1.01 |
| 10th Percentile | 1.02 | 1.48 |
| Ratio | 1.58 | 0.68 |

non-agriculture in the model, which is 2.3. The implication is that if selection were the only phenomenon at work, agriculture productivity differences would be 2.3 times as large as productivity differences in non-agriculture. The equivalent figure in the data is 10.7. Thus, this experiment implies that selection accounts for roughly a factor of two of the cross-country variation in agriculture productivity compared to non-agriculture. For illustration, the bottom row presents the model's predictions without selection, i.e. when worker heterogeneity is shut down. In this case the ratio is 1.0, as the productivity gaps are the same in each sector as the $A$ differences themselves.

Another way of gauging the quantitative importance of selection is consider the effect on each sector separately. For agriculture, the model predicts that productivity differences that are roughly 60 percent larger than the underlying $A$ differences (32 versus 20). Similarly, in non-agriculture, selection amplifies down the $A$ differences by roughly 30 percent (14 versus 20). Note that the combined impact on the ratio of agriculture to non-agriculture productivity differences, which is 2.3, can be computed by taking the amplification factor in agriculture (1.6) and dividing by the amplification factor in non-agriculture (0.7).

Table 3 provides more insight about where the selection effects come from. For each country, the table reports the expected individual productivity of workers in each sector relative to the population mean (unconditional expected productivity). In the 90th percentile country, the average agriculture worker is 1.62 times as productive as the population mean. Recall that the 90th percentile country in the model has a small fraction of workers in agriculture. What the model predicts is that this small set of workers are in fact much more productive in agriculture than a worker taken at random from the population. In the 10th percentile country, in contrast, agriculture workers are just 1.02 times as productive as the population mean. Essentially, agriculture workers in the poor country are roughly comparable to the population mean. The ratio of average productivity of agriculture workers in the two countries is 1.58. Note that this corresponds exactly to the amplification factor in agriculture discussed above.

In non-agriculture, selection forces works in the opposite direction. In the 90th percentile country, non-agriculture workers are just 1.01 times as productive as the population mean. This is not surprising as 97 percent of workers are employed in the non-agriculture sector in this country. In the 10th percentile country, with 22 percent of workers employed in non-agriculture sector, non-agricultural workers are 1.48 times as productive as the population mean. Taking a ratio of the 90th to 10th percentile of the country income distribution gives 0.68, which corresponds to the non-agriculture amplification factor.

These observations imply that workers with a comparative advantage in agriculture (non-agriculture) also have an absolute advantage in agriculture (non-agriculture.) This means that both conditions on the individual-productivity distribution of Proposition 2 hold. We note that there was nothing in our calibration strategy that guaranteed this outcome. Indeed, there exist parameter combinations for which one of the conditions fails. The sensitivity analysis of Section 5.2 provides one such example, and illustrates the dimensions on which it is counterfactual to the data.

### 4.5. Assessment of Calibrated Model's Cross-Country Implications

The model has a variety of other predictions for the cross section of countries. Below we highlight some of its other main quantitative implications and compare them to cross-country data.

**Agriculture Wage Gaps.** One novel prediction of our model is that average wages are much lower in agriculture than non-agriculture even though there there are no barriers to workers moving between sectors (as in the models of Caselli and Coleman (2001), Restuccia, Yang, and Zhu (2008), Herrendorf and Teixeira (2011), Adamopoulos and Restuccia (2010), and Tombe (2011)). In our model, this agricultural wage gap is driven entirely by selection: in equilibrium, most of the high-wage individuals are those who possess a comparative advantage in non-agriculture production and self select into that sector (see Section 4.2). Figure 1 plots the ratio of average wages in agriculture to non-agriculture against GDP per worker using wage data from the ILO and the predictions from our model. In the data, virtually all countries exhibit a large agricultural wage gap, with ratios of average wages below one in countries at all levels of the income distribution and substantially less than one in the poorest countries. The model also predicts a ratio less than one in all countries, with a slight decline as GDP per worker declines.

The success of our model on this dimension is important because new evidence suggests several prominent explanations are unable to account for these wage gaps. Using new data from a large set of developing countries, Gollin, Lagakos, and Waugh (2011a) find that even after adjusting for human capital differences between sectors, as emphasized by Caselli and Coleman (2001), lower costs of living in agricultural areas, as documented by Ravallion, Chen, and Sangraula (2009), and hours worked differences, there are still large residual wage gaps in agriculture in developing countries.[14] They find that the average real wage in agriculture is roughly

[14]There, we refer to this agricultural wage gap synonymously as the "agricultural productivity gap."

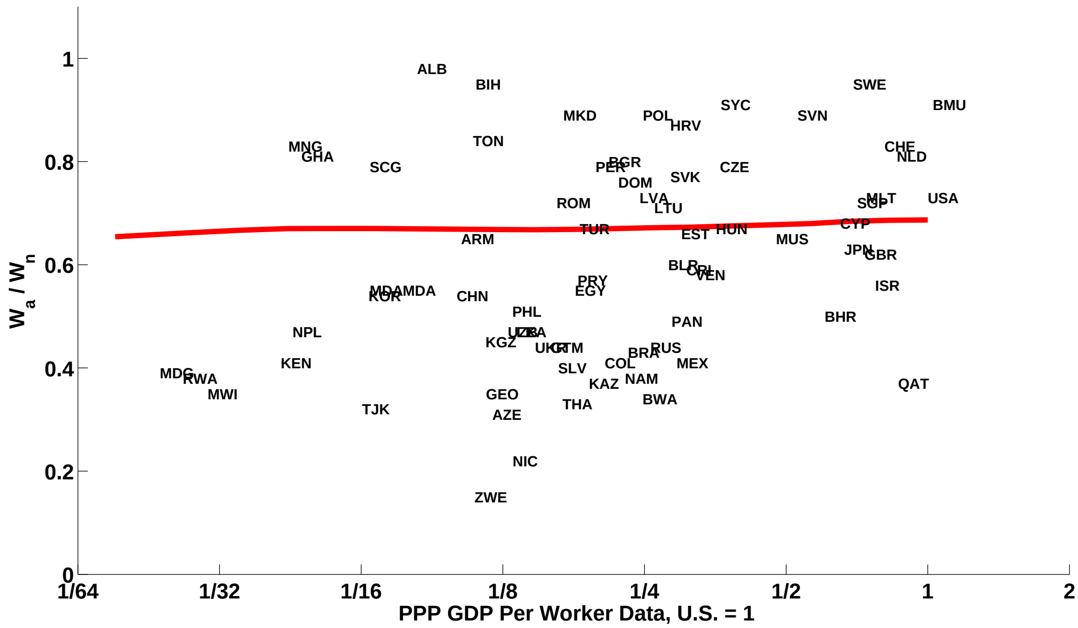


**Figure 1: Average Wage in Agriculture Relative to Non-agriculture, Data and Model**

one-half of the average wage in non-agriculture in the typical developing country, and argue that some other force besides these three must be behind these gaps.

The current paper suggests an alternative explanation, which is that selection forces induce higher wage workers to be disproportionately in the non-agriculture sector. Still, the model predicts that wages are roughly two-thirds as high in agriculture as non-agriculture, compared to the one-half number found by Gollin, Lagakos, and Waugh (2011a). This suggests that selection explains part but not all of the lower relative wages in agriculture in developing countries.

**Wage Inequality.** We now ask whether our model's predictions for income inequality and the income level are consistent with cross-country data. In the model, wage inequality increases slightly in income per capita. The reason is that as countries become richer, their workers move out of agriculture—a sector which has relatively low productivity variation across individuals—and into non-agriculture, which has higher productivity variation.

In the data, this relationship has been widely studied in the hunt for a Kuznets Curve, or hump shape in inequality (as measured by a Gini coefficient) as a function of GDP per capita. Barro (2000) summarizes the evidence by arguing that, while the cross-country data on income per capita and Gini coefficients do support a Kuznets curve, the bulk of the relationship between income level and income inequality remains unexplained by it. Thus, the model is in line with these data, and consistent with the Kuznets (1955) argument that as a country develops, movement out of agriculture activities will generate more inequality.

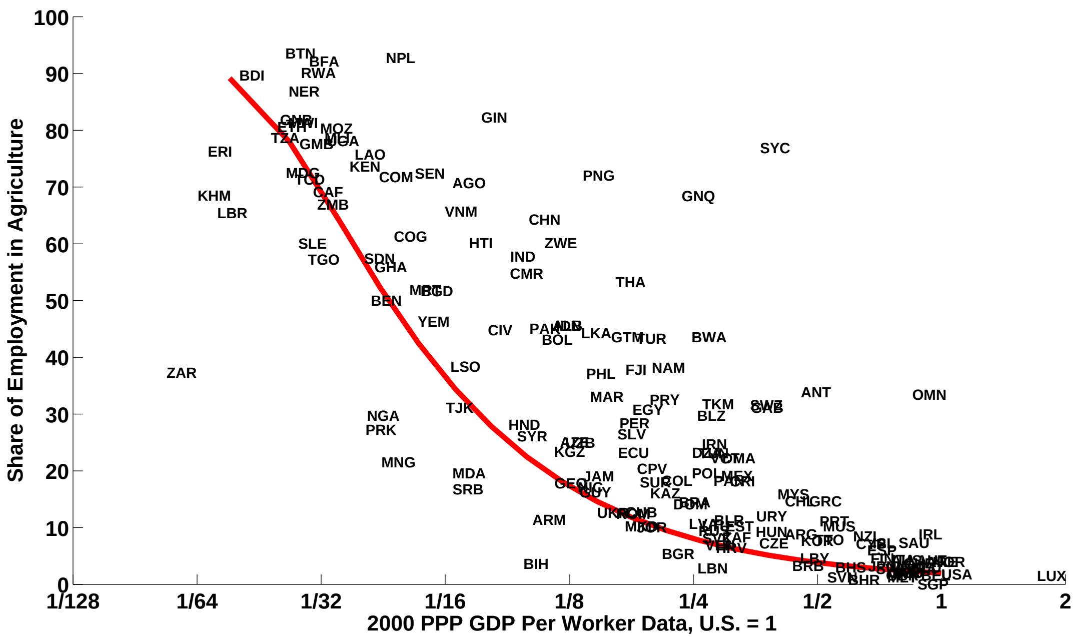


Figure 2: Share of Employment in Agriculture, Data and Model

**Share of Workers in Agriculture.** The model's non-homothetic preferences assure that the share of workers in agriculture declines in income per capita. Here we ask whether the model's quantitative implications for the share of workers in agriculture is consistent with the data, as is needed to accurately gauge the importance of the selection mechanism (see e.g. (9) in the analytical version of the model). Figure 2 plots data on the percent of employment in agriculture against GDP per worker data and the predictions from our calibrated model. In the data, the country in the 10th percentile of the income distribution has an employment share in agriculture of 78 percent, whereas the country in the 90th percentile of the income distribution has a share of three percent. The model predicts that a country in the bottom 10th percentile in GDP per worker should have around 63 percent of workers in agriculture, and that the percent declines with increases in GDP per worker in a way that is quantitatively consistent with the data.

**The Relative Price of Agriculture.** As Proposition 1 shows, the model predicts that relative agriculture prices are higher in poor countries than rich countries. Figure 3 plots the predictions of our quantitative model, as well as data on the relative price of agriculture and GDP per worker. Our data on relative agriculture prices are constructed using 2005 data from the International Comparison Programme (ICP); Appendix C provides the complete details. Figure 3 shows that relative agriculture prices systematically decline in GDP per worker, with a ratio of relative prices between countries in the 90th and 10th percentiles of GDP per worker of 2.5. The solid line in Figure 3 plots the model's prediction. In the model, relative agriculture prices also systematically decline with GDP per worker, and the ratio between the 90th and

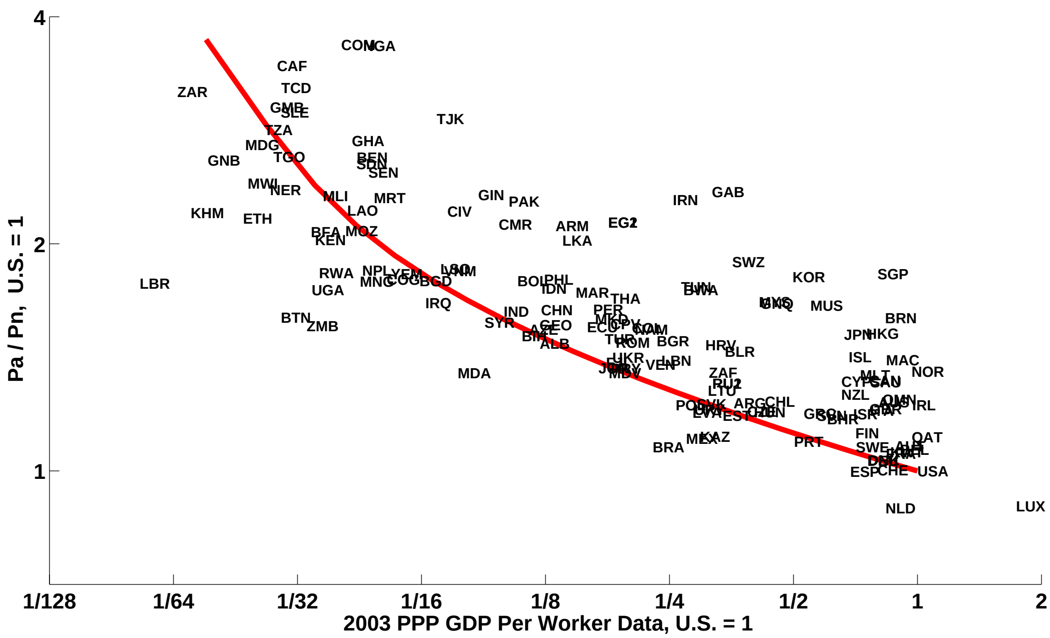


**Figure 3: Relative Price of Agricultural Goods, Data and Model**

10th percentiles is 2.3.[15]

One concern is that the data are based on the prices that consumers pay for goods, not the price that producers receive. This distinction would reflect distribution margins that are not in the model. If distribution margins vary systematically with the level of development (see for example Adamopoulos (2009)), then the relationship in Figure 3 may not reflect differences in relative agriculture-producer prices. To address this concern, we examined relative agriculture-price data using producer prices constructed by Restuccia, Yang, and Zhu (2008). We find that, by these measures, relative agriculture prices systematically decline in GDP per worker—as our model predicts—and, in fact, the relationship is even stronger than for consumer prices.

### 4.6. Evidence Using Proxies for Individual Productivities

In this section we take a different approach to assessing the plausibility of the calibrated model. In particular, we use two plausible proxies for agriculture and non-agriculture individual productivity that are observable independent of the sector the worker selects and provide evidence supporting key implications of our model.

[15]This fact is consistent with previous studies of variation in cross-country relative prices—e.g., Summers and Heston (1991), Jones (1994), Restuccia and Urrutia (2001), and Hsieh and Klenow (2007). In particular, Herrendorf and Valentinyi (2009), show that when partitioning ICP goods into agricultural and non-agricultural goods, the relative price of agriculture is higher in poor countries. They also show that partitioning goods into tradeable and non-tradeable goods implies a higher relative prices of tradeables in poor countries, and partitioning goods into consumption and investment goods implies a higher relative price of investment goods in poor countries.

The two proxies we use are height for agriculture and cognitive ability scores for non-agriculture. The rationale is that height reflects the "physical vigor" (Steckel (1995)) useful in physically demanding jobs such as agricultural work (see Pitt, Rosenzweig, and Hassan (2010), Pitt, Rosenzweig, and Hassan (1990), Steckel (1995), and Strauss and Thomas (1998) plus the references therein). Cognitive ability scores in turn plausibly capture the verbal, analytical or other non-physical capabilities often valued in non-agriculture activities (see e.g. Case and Paxson (2008) or Miguel and Hamory (2009)). While these proxies are certainly crude, they offer the advantage of being observable whether or not someone works in a particular sector, and have been (reasonably) widely measured in practice.

Using these proxies, we can compare the model's correlation between individual productivities and the correlation between the observed proxies. As discussed in Section 4.2, the model's correlation is positive, but modest in magnitude, with a linear correlation coefficient of 0.31. This correlation is very much consistent with the correlations between height and cognitive ability scores. Existing studies find correlations between height and cognitive ability in the range of 0.10 to 0.30 (see Case and Paxson (2008) and the references therein).

Another important implication of the model is that agriculture workers in rich countries are more productive in agriculture than the average worker (see e.g. Table 3). Using height as a proxy for agriculture productivity, we should find that agriculture workers in rich countries are taller than the average worker. To check this prediction, we draw on height data for U.S. adults collected in the 2009 National Health Interview Survey, conducted by the Center for Disease Control (CDC). We find that the average agriculture worker is 172.4 cm tall, while the average worker is 170.0 cm tall. The difference of 2.4 cm, or roughly one inch, is statistically significant at well below the 1 percent confidence level. Furthermore, it is economically significant: according to the CDC, this difference is equivalent in magnitude to the overall increase in average height in the U.S. from 1960 to 2009.

The developing country analog is that non-agriculture workers in poor countries are more productive in non-agricultural tasks than the average worker (again see Table 3). Using cognitive ability scores as a proxy for non-agriculture productivity, we should find that non-agriculture workers in poor countries have higher cognitive ability scores than average. While cognitive ability score data from developing countries are limited, the available evidence supports this implication. Using a unique data set from Kenya, Miguel and Hamory (2009) find that among rural Kenyan students, cognitive ability scores are a very strong predictor of who later migrates out of agricultural areas to take non-agricultural employment. Their estimates suggest that students that score one standard deviation higher on cognitive ability tests are roughly 17 percent more likely to migrate out of agriculture areas after finishing school. In addition, other studies have found that those with greater schooling attainment are far more likely to migrate to non-agricultural areas (e.g. Lanzona (1998) and Beegle, De Weerdt, and Dercon (2011)). As schooling attainment is correlated with cognitive abilities, this evidence also supports the model's predictions that non-agriculture workers in developing countries have higher cognitive ability than average.

We conclude that, when using height and cognitive ability scores as proxies for agricultural and non-agricultural individual productivity, the available evidence is in fact consistent with the model's predictions. In particular, the correlation between the proxies does appear to be positive but modest: agriculture workers in the U.S. do appear to be selected on height, and non-agricultural workers in developing countries do appear to be selected on cognitive ability. Of course given the crudeness of these proxies and limited availability of data, we take this evidence as supportive rather than definitive.

## 5. Robustness Exercises

### 5.1. Model Predictions for Countries at Intermediate Income Levels

In this section we compute the predictions of the benchmark model for intermediate income levels. We conclude that the role of selection is less important for understanding productivity differences between rich and intermediate income countries than between rich and poor countries. The reason is that shares of employment in agriculture are much more similar in rich and intermediate income countries, and hence differences in the average productivity of agriculture workers are much less pronounced then they are between rich and poor countries.

Table 4 illustrates the model's prediction for the 90th-50th ratio. As in the 90-10 experiment, $A$ differences are chosen to match the aggregate GDP per worker difference of a factor 3.1. The model predicts a factor 3.8 gap in agriculture and a factor 3.0 gap in non-agriculture. In the data, these gaps are a factor 11.1 in agriculture and 1.9 in non-agriculture. The last column shows that for these countries there is 5.8 times as variation in agriculture productivity as non-agriculture productivity. The model predicts just 1.3 times as much variation, or far smaller than in the data.

Why does the model fare so poorly in this case? The reason is that the employment shares in agriculture between the 90th and 50th percentile economies are not as different as they are for the 90th and 10th percentile countries. The share of workers in agriculture in the 50th-percentile country is 9 percent, compared to 3 percent in the 90th-percentile country. Thus, agriculture workers are highly selected based on agriculture productivity in both countries, and hence average worker productivity is only slightly lower in the 50th-percentile country. In contrast, in the 10th-percentile country, 78 percent are in agriculture, so the average worker has substantially lower productivity than the average agricultural worker in the 90th-percentile country. Equation (9) from our analytical example illustrates this point: selection plays a larger role when employment shares differ greatly, as they do between rich and poor countries, but not between rich and middle-income countries.

**Table 4: 90-50 Productivity Differences, Data and Benchmark Model**

| | Agriculture | Aggregate | Non-Agriculture | Ag/Non-Ag Ratio |
|---|---|---|---|---|
| Data | 11.1 | 3.1 | 1.9 | 5.8 |
| Model | 3.8 | 3.1 | 3.0 | 1.3 |
| Without Selection | 3.0 | 3.0 | 3.0 | 1.0 |

### 5.2. Sensitivity Analysis: Size of Correlation Between Individual Productivities

In this section we study the sensitivity of our results to the size of the correlation between individual productivity draws. This is an important issue because the correlation parameter helps determines the magnitude of the selection effects and when the conditions in Proposition 2 hold or not. Heckman and Honoré (1990) formalize this last point by showing that in the Roy model, the correlation in individual productivities determines how average productivity of workers in each sector relates to the unconditional averages, and in turn how comparative advantage aligns with absolute advantage, i.e. the conditions in Proposition 2.[16] Thus we explore how varying the correlation affects our results.

To explore these issues, we recompute the results of our main experiment (of Table 2) under a range of correlation coefficients running from 0.00 (independence) to 0.99 (near perfect correlation) by varying the dependence parameter $\rho$. In each case, we re-calibrate $\theta_a$ and $\theta_n$ to be consistent with the (adjusted) standard deviation of log wages in each sector and re-calibrate the preference parameters as described in Section 4.3. We do not attempt to match the ratio of average wages (since by varying $\rho$, we are no longer able to) but instead report the model's prediction for the sector wage ratio for each correlation value.

Table 5 shows the results of varying the model's correlation parameter in individual productivity, with the calibrated model in the center column (and marked with a star). The first row reports the Spearman rank correlation coefficient in each experiment. The second row reports the ratio of average wages in agriculture to non-agriculture. The third and fourth rows report the productivity differences between the 90th and 10th percentile countries in the two sectors. The final row presents the ratio of sector productivity differences.

One prominent feature in Table 5 is that higher values of correlation in individual productivity lead to smaller quantitative effects of selection. For example, starting with the calibrated model

[16] Heckman and Honoré (1990) refer to the case when comparative advantage aligns with absolute advantage as the "standard case," and the case when agents with a comparative advantage have an absolute disadvantage as the "non-standard case."

**Table 5: Sensitivity of Sector Productivity to Correlation**

| Correlation in individual productivity | 0.00 | 0.10 | 0.20 | 0.24* | 0.30 | 0.40 | 0.99 |
|---|---|---|---|---|---|---|---|
| Ratio of average wage $\bar{w}_a/\bar{w}_n$ | 0.74 | 0.73 | 0.71 | 0.69* | 0.66 | 0.61 | 0.48 |
| Agriculture Productivity Difference | 38.1 | 36.0 | 33.5 | 32.1* | 30.3 | 27.3 | 19.5 |
| Non-Agriculture Productivity Difference | 10.9 | 11.8 | 13.0 | 13.8* | 14.4 | 15.4 | 17.6 |
| Ag/ Non-Ag Ratio | 3.5 | 3.1 | 2.6 | 2.3* | 2.1 | 1.8 | 1.1 |

and increasing the correlation to 0.3 and 0.4 leads to agriculture gaps of 30.3 and 27.3 down from 32.1 in the calibrated model. Non-agriculture gaps rise to 14.4 and 15.4 up from 13.8. Thus, the model performs modestly worse in this range, with the combined affect of selection falling to a ratio 2.1 and 1.8 respectively. One challenge to correlation parameters in this range is the ratio of average sector wages are counterfactually low at 0.66 and 0.61, respectively.

In contrast, lower values of the correlation parameter lead to larger quantitative effects. Lowering the correlation to 0.2 and 0.1 leads to larger agriculture gaps of 33.5 and 36.0, and smaller non-agriculture gaps of 13.0 and 11.8. The combined effects rise to ratios of 2.6 and 3.1. The ratio of average wages also rises above the level found in the data, to 0.71 and 0.73.

The first and last data columns present some extreme examples of correlation, namely no correlation and near-perfect (0.99) correlation. In the zero-correlation case, the model performs better than in the benchmark case, with agriculture and non-agriculture gaps of 38.1 and 10.9, and an overall ratio of 3.5. The wage ratio is counterfactually high at 0.74. In the case of near-perfect correlation, the agriculture differences are just a factor 19.5, which is smaller than the underlying $A$ difference of factor 20. The non-agriculture difference is still smaller, at 17.6. The reason that the agriculture sector "flips" here is that in this case workers with a comparative advantage in agriculture have an absolute disadvantage there, i.e. one of the conditions in Proposition 2 does not hold. Thus, selection works in the opposite way as in the standard case, and agriculture productivity differences are smaller than $A$ differences. Of course a major limitation of having such a high correlation is that the average wage in agriculture relative to non-agriculture is strongly counterfactual, at 0.48.

### 5.3. Alternative Quantitative Experiment: Calibrate to Non-Agriculture Productivity Gaps

In this section we perform an alternative experiment that calibrates the model to target the observed non-agriculture productivity difference while maintaining agriculture employment shares consistent with the data. The motivation is two-fold. First, it allows us to highlight a key difference between our model and the existing literature, which is that our model requires underlying productivity differences larger than those of non-agriculture in order to match measured non-agriculture productivity differences. Second, it illustrates that the quantitative importance of selection does not depend on the assumption that the underlying productivity differences are sector neutral; similar results arise when the underlying exogenous productivity differences are larger in agriculture. Instead, what matters for the selection mechanism to work is that employment shares in agriculture are very different across rich and poor countries, as in the data.

To execute this experiment we will introduce one additional parameter, $A_a$, which allows agriculture efficiency to differ from non-agriculture efficiency. Formally, our agriculture production function is now $Y_a = A_a A L_a$. As a result, it is optimal to work in non-agriculture if and only if $\frac{z_n^i}{z_a^i} \geq p_a A_a$. This extra degree of freedom allows us to simultaneously target a non-agriculture productivity difference of four and match agriculture employment shares in the poor country. This last moment is important because an accurate measurement of the importance of selection should be based on differences in employment shares across model economies that are the same size as they are in the data for rich and poor countries.[17]

While we take this additional parameter $A_a$ as exogenous, it has several possible motivations. For one, it could represent agriculture-specific differences in land per worker or capital per worker, which we currently abstract from, but explore in Section 6. It could also represent the type of agriculture-specific distortion emphasized by the existing literature. For example it could be distortions to the use of intermediate inputs in agriculture as studied by Restuccia, Yang, and Zhu (2008), or restrictions on farm size, as emphasized in Adamopoulos and Restuccia (2010).

Beginning from the benchmark model calibrated as in the main experiment, we normalize $A_a$ to be one in the United States. We then lower $A$ and $A_a$ to match a productivity difference of 4 in non-agriculture, as in the experiments of Restuccia, Yang, and Zhu (2008) and Adamopoulos and Restuccia (2010), and an agriculture employment share of 78 percent, as in the 10th percentile country. We then compute the model's predictions for aggregate GDP per worker and agriculture productivity in the 90th and 10th percentile countries.

Table 6 presents the results of the alternate experiment. Note that the gap in non-agriculture productivity is a factor of 4 in the model (as in the data) by construction. It is important to understand that this requires an underlying $A$ difference of 8 (and not 4). This is because the

[17] The analytical example of Section 3.5 illustrates this point. In equation (8) and (9), a key determinant of selection's role is the share of workers in a sector. Thus, if our quantitative model does not deliver employment shares similar to the data, then we cannot provide an unbiased measurement of the role of selection.

**Table 6: Alternative Experiment, 90-10 Ratio, Data and Benchmark Model**

| | Agriculture | Aggregate | Non-Agriculture | Ag/Non-Ag Ratio |
|---|---|---|---|---|
| Data | 45 | 22 | 4 | 10.7 |
| Model | 42 | 16 | 4 | 9.9 |
| Without Selection | 26 | 18 | 8 | 3.4 |

**Note:** $A$ and $A_a$ are picked to match non-agriculture difference of 4 and a 78 percent agriculture employment share in the poor country.

non-agriculture productivity differences delivered by the calibrated model are smaller than the $A$ differences themselves. In contrast, the experiments of Restuccia, Yang, and Zhu (2008) and Adamopoulos and Restuccia (2010) require $A$ differences of the exact same size as the non-agriculture productivity differences. The gap in aggregate productivity is a factor 16, modestly less than in the data, and agriculture productivity is now a factor of 42, largely on par with the 45 in the data. Overall, the model delivers productivity differences that are 9.9 times as large in agriculture as non-agriculture, just slightly less than the data.

To measure the importance of selection, we re-solve the model without the selection channel, i.e. with worker heterogeneity shut down. In this case the model predicts a slightly larger aggregate difference of 18, a smaller agriculture difference of 26, and a non-agriculture difference of 8. Agriculture productivity is now just 3.4 times as variable across countries as non-agriculture. Thus, the model with selection generates 2.9 times larger variation in sectoral productivity than the model without selection (9.9/3.4). While this is a different exercise than the experiment in Section 4.4, the quantitative importance of selection is roughly comparable to the estimate of 2.3 computed in the main experiment.

### 5.4. Open-Economy Considerations

Up to this point we have treated each model economy as closed. This raises an important question: how will the model's predictions change if we allow for international trade? We argue that as long as a model with international trade generates labor allocations consistent with cross-country data, the model's quantitative predictions for sector productivity differences across countries will remain the same. This argument is clearly seen in the special case of our model in equation (9): if an open-economy model supports the same allocation of workers in agriculture and non-agriculture as the closed-economy model, then the open-economy model's predictions for productivity differences are the same. The only distinction between the models is how the relative price of agriculture is determined in equilibrium.

However, our model does have important implications for the impact from international trade. In Gollin, Lagakos, and Waugh (2011b) we build on the framework in the current paper within the Eaton and Kortum (2002) Ricardian model of trade. A key result is that the welfare gains from a trade liberalization are smaller relative to the standard Eaton and Kortum (2002) framework because of how labor productivity in each sector responds as workers reallocate following the liberalization. Less productive workers are drawn into the non-agriculture sector reducing a country's comparative advantage in that sector and reducing the scope and hence gains from trade. Thus, our model has important predictions for international trade in addition to its ability to explain the productivity patterns of Table 1.

## 6. Extended Model with Capital and Land

We now extend the model to include capital and land. Up to this point we abstracted from capital and land mainly for transparency. One concern with this abstraction is that capital and land may interact with selection in ways that diminish the importance of selection. A second concern is that, by ignoring capital and land, the calibration procedure may overestimate the amount of wage variation that is attributable to productivity variation across individuals, which would again lead to an overestimate of the importance of selection. As we show below, neither of these concerns turn out to be warranted.

### 6.1. Production with Capital and Land

In this extension, each worker has access to technologies to produce either the agriculture good or the non-agriculture good. The technologies are:

$$\begin{aligned} y_a^i &= Ak^{\phi_k}\ell^{\phi_l}(z_a^i)^{1-\phi}, \quad \text{with } \phi = \phi_k + \phi_l, \\ y_n^i &= Ak^{\alpha}(z_n^i)^{1-\alpha}, \end{aligned}$$

where $k$ represents capital, and $\ell$ represents land. Note that we abstract from land as a factor of production in the non-agriculture sector and allow for capital and labor's shares to potentially differ across sectors, as consistent with recent estimates (Valentinyi and Herrendorf (2008)).

To solve this model we work backwards by first characterizing the solution to the profit maximization problem given an occupational choice, and then characterizing the occupational choice. Given the decision to work in agriculture, the profit maximization problem is

$$\max_{k,\ell} \left\{ p_a Ak^{\phi_k}\ell^{\phi_l}(z_a^i)^{1-\phi} - rk - p_\ell \ell \right\}.$$

Workers rent capital and land to maximize profits. The price $r$ is the cost of renting one unit of capital, and $p_\ell$ is the price of renting one unit of land. Workers are the residual claimants on earnings after payments to capital and land are made; we denote individual $i$'s earnings as $w^i(z_a^i)$.[18]

Given the decision to work in non-agriculture, the profit maximization problem is

$$\max_{k} \left\{ Ak^{\alpha}(z_n^i)^{1-\alpha} - rk \right\}.$$

Here workers rent only capital to maximize profits, and again serve as residual claimants on earnings after payments to capital are made; we denote individual $i$'s earnings as $w^i(z_n^i)$.

Occupational choice comes down to a comparison of potential wage earnings in both sectors. These wages are

$$w^i(z_a^i) = z_a^i(1-\phi)(p_a A)^{\frac{1}{1-\phi}} \left(\frac{\phi_k}{r}\right)^{\frac{\phi_k}{1-\phi}}, \text{ and} \tag{10}$$

$$w^i(z_n^i) = z_n^i(1-\alpha)(A)^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\alpha}}. \tag{11}$$

Combining equations (10) and (11) yields a simple cutoff rule in relative individual productivity characterizing the optimal occupational choice for each worker. Working in non-agriculture is optimal for worker $i$ if and only if

$$\frac{z_n^i}{z_a^i} \geq \chi p_a^{\frac{1}{1-\phi}} A^{\left(\frac{1}{1-\phi}-\frac{1}{1-\alpha}\right)} r^{\left(\frac{\alpha}{1-\alpha}-\frac{\phi_k}{1-\phi}\right)} p_\ell^{\frac{-\phi_l}{1-\phi}}, \tag{12}$$

where $\chi$ is a collection of constants. While similar to the cutoff rule in equation (4) of the benchmark model, this cutoff rule differers in two respects. First, the price of capital and the price of land now factor into the decision where to work. Second, economy-wide efficiency directly enters into the equation, with its impact determined by the difference in labor shares between the two sectors (i.e. $1-\phi$ vs. $1-\alpha$).

### 6.2. Optimization and Equilibrium

Optimal consumption decisions are the same as in the benchmark model. A worker's income now consists of her labor earnings plus an equal share of the aggregate payments to capital and land.

[18]We model agriculture and non-agriculture workers as self-employed. Gollin (2008) provides evidence that self-employment in both agriculture and non-agriculture sectors is a key feature of the data in poor countries.

An equilibrium of the economy consists of an agriculture price, $p_a$, a price of capital, $r$, a price of land, $p_\ell$, wages per efficiency unit of labor in each sector, $w_a$ and $w_n$, and allocations for each worker, such that workers optimize and all markets clear.

### 6.3. Calibration of Extended Model

We begin with the calibration of the individual productivity distribution. One important feature of this extended model is that capital and land do not affect the calibration results of the dispersion parameters relative to the benchmark model. The reason is that in both models log wage variation only reflects variation in log individual productivity, i.e. $\text{var}(\log w^i(z_a^i)) = \text{var}(\log z_a^i)$. To see this in the extended model, equations (10) and (11) show that payments to labor are proportional to individual productivity, and the degree of proportionality is common across workers of different productivity. Hence, calibrating the model to the wage variance targets described in Section 4.2 results in the same $\theta_a$ and $\theta_n$ values as in the benchmark model.

Furthermore, given that the calibration results in the same variances of individual productivities as in the benchmark model, the correlation parameter $\rho$ must take on the same value as in the benchmark calibration. In both models, the calibration procedure must set the cutoff value in relative productivities in such a way to get two percent of workers in agriculture. Given that this cutoff value is the same, the set of workers in agriculture is the same in both models, and thus the ratio of average wages is the same. Hence, the value of $\rho$ that matches the average wage targeted in the data is the same here as in the benchmark model.

Incorporating capital and land adds several new parameters to calibrate: capital and land shares in agriculture production, $\phi_k$ and $\phi_l$, capital's share in non-agriculture production, $\alpha$, and aggregate capital and land stocks, which we denote $K$ and $\mathcal{L}$. We use the evidence of Valentinyi and Herrendorf (2008) on capital and land shares by sector in the U.S. to calibrate $\phi_k$, $\phi_l$ and $\alpha$. They find values for capital and land's share in agriculture production to be $0.36$ and $0.18$ which we assign $\phi_k$ and $\phi_l$ to take. While these values are for the U.S., they are also consistent with observed share-cropping arrangements in poor countries, where workers typically earn around one-half of all output; see Gollin, Lagakos, and Waugh (2011a) for a more detailed discussion. For non-agriculture, Valentinyi and Herrendorf (2008) find capital's share to be $0.33$, which we assign $\alpha$ to take.

To calibrate the aggregate capital stock, $K$, we pick this value so the capital-output ratio in the rich economy is 2.5, which is consistent with evidence from the U.S. To calibrate the aggregate land endowment we follow Adamopoulos and Restuccia (2010) and pick units such that average land per worker equals 169.3 hectares as they find in the U.S. data. Finally, we calibrate the preference parameters, $\bar{a}$ and $\nu$, as described in Section 4.3.

**Table 7: 90-10 Productivity Differences, Data and Extended Model**

| | Agriculture | Aggregate | Non-Agriculture | Ag/Non-Ag Ratio |
|---|---|---|---|---|
| Data | 45 | 22 | 4 | 10.7 |
| Model | 41 | 22 | 10 | 3.9 |
| Without Selection | 36 | 21 | 17 | 2.1 |
| Without Selection or Land | 17 | 16 | 15 | 1.1 |

### 6.4. Results for Extended Model

To explore the quantitative implications of the extended model, we normalize $A$ to equal one, and choose $K$ to match GDP per worker relative to the U.S. for a country in the 90th percentile of the income distribution and a capital-output ratio of 2.5. We then lower $A$ and $K$ to match the aggregate productivity difference of 22 between the 90th and 10th percentile countries, and a capital-output ratio of 1.0 in the 10th percentile county, as consistent with the data.

Table 7 presents the results. The extended model generates a factor 41 difference in productivity in agriculture and a factor 10 difference in non-agriculture. This amounts to 3.9 times as much variation in agriculture productivity relative to non-agriculture productivity across countries. This is higher than the 2.3 ratio found in the baseline experiment. Of course, the extended model has several other factors contributing to the larger differences in agriculture. In particular, land per worker is lower in the poor country, a feature that is present in other models with land as a fixed factor, such as Restuccia, Yang, and Zhu (2008), Adamopoulos and Restuccia (2010), and Herrendorf and Teixeira (2011). To measure the role of selection versus land, below we discuss two decompositions that allow us to assess the importance of each factor.[19]

To compute one measure of the importance of selection, we re-compute the model's predictions without the selection channel (i.e. with worker heterogeneity shut down) leaving all else the same. The third line in Table 7 reports the results. In this case the model predicts a lower agriculture difference of a factor 36, an aggregate difference of 21, and a higher non-agriculture difference of 17. The ratio of agriculture to non-agriculture productivity differences is now 2.1. Thus, the model with selection generates 1.9 times larger variation in sectoral productivity than

[19] We find that the implications of this calibration for other cross-country observables not directly targeted, such as the share of labor in agriculture and relative price of agriculture goods, are reasonable. One additional check of the model is in the average size of a farm in the 10th percentile country. Our model predicts an average size of 5.4 hectares, which is quite close to the value of 5.0 hectares reported by Adamopoulos and Restuccia (2010).

the model without selection (3.9/2.1). This shows that selection leads agriculture productivity differences to be roughly twice as large as non-agriculture differences, or just below the 2.3 of the main experiment.

As a frame of reference, consider the importance of land relative to selection. The fourth line in Table 7 reports the results when we remove land from the model. Agriculture productivity differences now fall to a factor 17, aggregate differences fall to a factor 16, and non-agriculture differences fall to a factor 15. The ratio of agriculture to non-agriculture differences falls to 1.1. Using the same logic as above, land-per-worker differences contribute a factor 1.9 (2.1/1.1) to understanding the ratio of agriculture to non-agriculture productivity differences, which is the same size as the selection channel.

One limitation of these experiments is that selection is not the only force that responds when the selection channel is shut down. In particular, the capital allocation changes leaving the ratio of capital per worker lower in agriculture and higher in non-agriculture in the poor country. This leads, all else equal, to larger labor productivity differences in agriculture and smaller ones in non-agriculture in the model without selection. Thus, this experiment understates the overall effects of selection.

An alternative way to measure the importance of selection is to consider the following decomposition of equilibrium output per worker in each sector. One can show that labor productivity in equilibrium can be written as

$$\frac{Y_a}{N_a} = (A)^{\frac{1}{1-\phi}} \left(\frac{K_a}{Y_a}\right)^{\frac{\phi_k}{1-\phi}} \left(\frac{\mathcal{L}}{Y_a}\right)^{\frac{\phi_l}{1-\phi}} \left(\frac{1}{N_a}\int_{i\in\Omega^a} z_a^i \, dGi\right), \text{ and} \tag{13}$$

$$\frac{Y_n}{N_n} = A^{\frac{1}{1-\alpha}} \left(\frac{K_n}{Y_n}\right)^{\frac{\alpha}{1-\alpha}} \left(\frac{1}{N_n}\int_{i\in\Omega^n} z_n^i \, dGi\right). \tag{14}$$

where the last bracketed term in equations (13) and (14) represent the contribution from selection. Expressing output in this way has the benefit of giving "credit" for variations in $K$ and $\mathcal{L}$ generated by selection and differences in $A$. For example, agents with higher individual productivity optimally use more capital and land per unit of labor — but capital-output ratios and land-output ratios only reflect aggregate scarcity of $K$ and $\mathcal{L}$. Klenow and Rodríguez-Clare (1997) and Hall and Jones (1999) make a similar argument for working with capital-output ratios rather than capital-labor ratios in the neoclassical growth model.

Taking a simple ratio of these bracketed terms in equations (13) and (14) across the rich and poor country decomposes the importance of each factor in accounting for the sector productivity differences in the model. The selection term contributes a factor of 1.6 to agriculture differences and 0.52 to non-agriculture differences. This implies that selection forces lead productivity differences in agriculture to be 3.0 times as large as those in non-agriculture (1.6/0.52), which is somewhat larger than in the main experiment.[20]

Together, these two decompositions establish bounds on the importance of selection in the extended model. The first decomposition suggests that selection leads to agriculture productivity differences that are 1.9 times as large as those of non-agriculture. The second decomposition suggests that selection leads to agriculture differences that are 3.0 times as large. Taken together, we conclude that the quantitative importance of selection is comparable in the extended model and benchmark model.

## 7. Evidence: The Prevalence of Women in Agriculture Across Countries

According to our theory, part of the large cross-country productivity differences in agriculture stem from poor countries having relatively more workers in agriculture who are unproductive at agriculture work (see e.g. Table 3). In this section, we provide one concrete example of this phenomenon. In particular, we cite evidence that women are less productive at agricultural work than men on average, and we show that in cross-country data, women form a larger fraction of agriculture workers in developing countries than in richer countries.

A large body of literature has found that women tend to earn lower wages than men in agricultural work (see e.g. Rosenzweig and Evenson (1977), Rosenzweig (1978), Psacharopoulos and Tzannatos (1992), and Horton (1996).)[21] One widely proposed hypothesis for this gender wage gap in agriculture is that women are less productive at agricultural work than men on average (e.g. Goldin and Sokoloff (1982, 1984), Foster and Rosenzweig (1996), Pitt, Rosenzweig, and Hassan (2010), and Alesina, Giuliano, and Nunn (2011).)

Several types of evidence support the hypothesis that women are less productive than men at agriculture work on average. As one piece of direct support, Pitt, Rosenzweig, and Hassan (2010) cite evidence from the U.S. and Bangladesh that men are physically stronger than women as measured by their grip strength. In Bangladesh, for example, 40 percent of men in a random

[20]This decomposition suggests a limited role for land. Because land is fixed and agriculture output is lower in the poor country than in the rich country, land-to-output ratios are actually slightly higher in the poor country than in the rich country. This suggests that land plays no role in explaining agriculture productivity differences other than through the selection channel. Put differently, the fixed quantity of land is not a limiting factor for agriculture in poor countries given the low average productivity of their agriculture workers.

[21]Rosenzweig and Evenson (1977) and Rosenzweig (1978) document that, in India, women earn roughly 0.75 as much as men in agriculture work. Psacharopoulos and Tzannatos (1992) document gender wage gaps in agriculture of 0.92 in Colombia, 0.70 in Costa Rica, 0.76 in Guatemala, and 0.69 in Peru, and Horton (1996) documents gaps of 0.89 in Thailand and 0.85 in the Philippines.

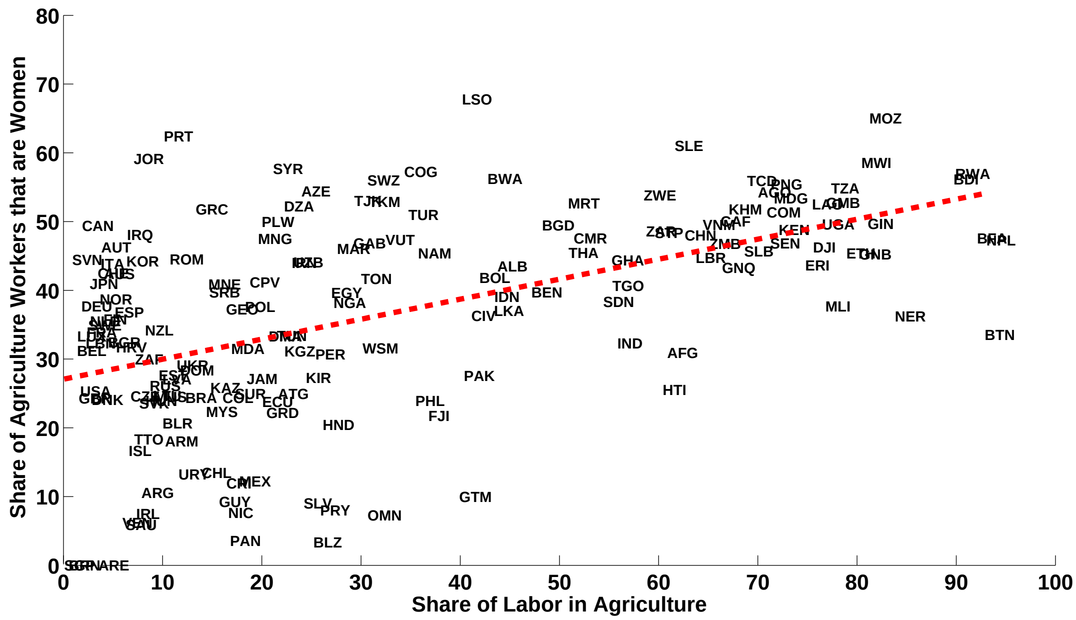


**Figure 4: Share of Agriculture Workers that are Women**

sample of adults had a stronger grip than the strongest woman. This matters for productivity since much of agriculture work, such as plowing, is strength-intensive. Further support comes from the sexual division of labor in agriculture. Foster and Rosenzweig (1996) show that in the agriculture sectors of many developing countries, most men are hired to do plowing, while most women are hired to do weeding.[22] Goldin and Sokoloff (1982, 1984) argue that a major reason women earned less than men in agriculture in the early U.S. was that women were generally less productive at plowing than men.[23]

Given evidence that women are relatively less productive in agriculture than men, we next show that women comprise a relatively larger fraction of the agriculture workforce in developing countries. In order to measure the prevalence of women in agriculture across countries, we draw on two independent sources of data. First, we use FAO data on the composition of agriculture workers by sex in 162 countries. The estimates come from a mix of labor force surveys and censuses of population. Second, we use data from **?**'s (**?**) Integrated Public Use Microdata Series (IPUMS) to compute the composition of agriculture workers by sex for 51 countries. These data come exclusively from nationally representative censuses of population, which in general have very large sample sizes. Using each data set we compute the fraction of each

[22] Alesina, Giuliano, and Nunn (2011) argue that, because women and children are less productive at plowing then men, societies that adopted plow agriculture earlier had lower demands for female and child labor, and hence have lower fertility rates today.

[23] Goldin and Sokoloff (1984) document a larger gender wage gap in agriculture in the North than the South, and attribute it to the North's predominance of hay and wheat farming, where plowing was required, compared to the South's focus on tobacco and cotton, for which a smaller stature was useful in harvesting.

country's agriculture workers that are women.

Figure 4 shows our calculations using the FAO data. We find that countries with higher shares of workers in agriculture tend to have a higher fraction of agriculture workers that are women.[24] For the countries with 70 percent or more of workers in agriculture, roughly 50 percent of agriculture workers are women. In contrast, in countries with 10 percent of workers in agriculture or less, on average 30 percent of agriculture workers are women. A linear regression of the share of agriculture workers that are women on the share of all workers in agriculture yields a slope coefficient of 0.29 with a P-value of 0.01. The IPUMS data (not pictured) paints a similar picture: a similar linear regression using the IPUMS data yields a slope coefficient of 0.33 with a P-value of 0.01.[25]

Putting these pieces together — (i) women are the less productive and agriculture work and (ii) women are more prevalent in agriculture in developing countries — provides a concrete example of how agriculture productivity differences across countries depend on the average productivity of workers in the agriculture sector as predicted by our theory.

## 8. Conclusion

We argue that cross-country productivity differences are larger in agriculture than in non-agriculture in part because subsistence food requirements lead workers that are relatively unproductive in agriculture work to nonetheless select into the agriculture sector in poor countries. In rich countries, in contrast, those few workers self-selecting into agriculture are those who are relatively most productive at farm work. As a result, measured labor productivity gaps are larger in agriculture than in the aggregate. Selection forces work in exactly the opposite way in non-agriculture, and productivity differences are smaller than those of the aggregate.

We formalize our theory in a general equilibrium Roy model, and calibrate it using data on the distribution of wages by sector. When calibrated, the model predicts that agriculture productivity differences are twice as large as those in non-agriculture, even when economies differ

[24]Time series evidence from the development experiences of the U.S. and Britain paint a picture consistent with our cross-country evidence. Goldin and Sokoloff (1982, 1984) show that as the U.S. grew in the 19th century, women shifted out of agriculture and into manufacturing much more rapidly than men. In Britain, the evidence of Allen (1994) shows that in 1700, 46 percent of adult agriculture workers were women, and by 1850 this fraction had fallen to just 29 percent. Authors' calculations using Allen (1994), Table 5.3.

[25]One alternative theory for why women are more prevalent in agriculture in developing countries is that higher fertility rates in the developing world make work on the family farm — where childcare can be provided easily — particularly attractive for women. One piece of evidence against this alternative theory is that the share of women *without children* in agriculture also increases sharply in the agricultural share of employment. A linear regression using our IPUMS data of the share of agriculture workers that are female without children under 5 on the agriculture of share of employment yields a slope coefficient of 0.21 with a P-value of 0.01.

only by an economy-wide efficiency term that affects both sectors uniformly. This result suggest that the larger cross-country productivity differences in agriculture may not exclusively be the result of distortions specific to agriculture in poor countries. Instead, they could be due to the optimal decisions of workers faced with subsistence consumption needs and low economy-wide efficiency. This low efficiency could in turn be due to weak institutions, poor protection of property rights, or poor social infrastructure, as emphasized by a growing macroeconomics literature (e.g. Hall and Jones (1999); Acemoglu, Johnson, and Robinson (2001, 2002)).

# A. Model Appendix

## 1.1. Proof of Proposition 1

Let $p_a^P$, $Y_a^P$ and $Y_n^P$ be the equilibrium relative price and quantities in an economy with economy-wide efficiency $A^P$. Denote by $p_a^R$, $Y_a^R$ and $Y_n^R$ the equilibrium of an economy with efficiency $A^R$.

Suppose that $p_a^R = p_a^P$, and that $p_a^R$ clears the output market in the rich economy. Then by (4), each worker $i$ would choose to work in the same sector in the two economies. Thus output in each sector would be scaled up by a factor equal to the ratio of the efficiency terms: $Y_a^R/Y_a^P = Y_n^R/Y_n^P = A^R/A^P$. But by (5), we know that workers must demand a higher fraction of non-agriculture goods in economy $A^R$ than $A^P$. But this implies that $Y_n^R/Y_n^P > Y_a^R/Y_a^P$, which is a contradiction. Thus $p_a^R \neq p_a^P$.

The only way to be consistent with the worker solutions, (5), is for more workers to supply labor in the non-agriculture sector in economy $A^R$ than economy $A^P$. By (4), this occurs if and only if $p_a^R < p_a^P$. ■

## 1.2. Proof of Proposition 2

Assume that $E(z_a|z_a/z_n > x)$ is increasing in $x$. By (4) we know that for any worker $i$ with individual productivities $z_a^i$ and $z_n^i$, if $i$ chooses to work in agriculture in country $P$ then $z_a^i/z_n^i > 1/p_a^P$, and if $i$ chooses to work in agriculture in country $R$ then $z_a^i/z_n^i > 1/p_a^R$. By Proposition 1 we know that $p_a^P > p_a^R$. Hence, by our assumption, $E(z_a|z_a/z_n > 1/p_a^P) < E(z_a|z_a/z_n > 1/p_a^R)$. Thus

$$\frac{Y_a^R/N_a^R}{Y_a^P/N_a^P} = \frac{A^R}{A^P} \cdot \frac{E(z_a|z_a/z_n > 1/p_a^R)}{E(z_a|z_a/z_n > 1/p_a^P)} > \frac{A^R}{A^P}.$$

A similar result holds when $E(z_n|z_n/z_a > x)$ is increasing in $x$. ■

## 1.3. Deriving Analytical Results for Independent Fréchet Individual Productivities

The probability we want to derive is Prob$\{z_n \leq p_a z_a\}$. To do so, note that this probability is represented by

$$\pi_a = \int_0^\infty \exp\{-(p_a z_a)^{-\theta}\} g(z_a) dza,$$

where the first term in the integral is the cumulative distribution function for productivity in non-agriculture evaluated at random variable $p_a z_a$, and the second term $g(z_a)$ is the individual productivity distribution function in agriculture. The anti-derivative for this integral is given by

$$\frac{1}{p_a^{-\theta}+1} \times \exp\{-(p_a^{-\theta}+1)z_a^{\theta}\}.$$

Evaluating the integral yields

$$\pi_a = \frac{1}{p_a^{-\theta}+1},$$

and similar arguments yields

$$\pi_n = \frac{p_a^{-\theta}}{p_a^{-\theta}+1}.$$

To compute the conditional average individual productivity in each sector, we make the following argument. First notice that the conditional productivity distribution for workers in non-agriculture is

$$\text{Prob}\{z_n < z | z_n > p_a z_a\} = \frac{\text{Prob}\{z_n < z, z_n > p_a z_a\}}{\text{Prob}\{z_n > p_a z_a\}}.$$

Then computing the probabilities in the numerator and the denominator we have

$$\frac{\text{Prob}\{z_n < z, z_n > p_a z_a\}}{\text{Prob}\{z_n > p_a z_a\}} = \exp\{-(p_a^{\theta}+1)z_n^{-\theta}\}.$$

Notice that the conditional productivity distribution of workers in non-agriculture is itself Fréchet distributed with centering parameter $(p_a^{\theta}+1)$. Using this insight we can now compute the average individual productivity of non-agriculture workers conditional on working in non-agriculture to be

$$E(z_n | p_a z_a < z_n) = (p_a^{\theta}+1)^{\frac{1}{\theta}}\gamma.$$

where the constant $\gamma$ is the gamma function evaluated at $\frac{\theta-1}{\theta}$. Similar arguments imply that average individual productivity of agriculture workers conditional on working in agriculture is

$$E(z_a | p_a z_a > z_n) = (p_a^{-\theta}+1)^{\frac{1}{\theta}}\gamma.$$

## B. Capital and Agriculture and Non-Agriculture Productivity Differences

To study the role of sector differences in capital per worker across countries, we use data on agricultural capital stocks constructed by Butzer, Mundlak, and Larson (2010). The capital stocks they construct represent estimates of the total value of machinery, structures, treestock and livestock used in agricultural production. They have estimates for a set of 30 countries from all levels of the world income distribution. One strength of this study is the effort to which the authors go to construct measures that are internationally comparable, which is no easy task given the data challenges inevitable in calculations of this nature. The main limitation is, as the authors point out, that there are still reasons to be skeptical of the international comparability of the data.

For our accounting calculations, we make use of their agricultural capital stock estimates from 1985, the year corresponding with the sector productivity data analyzed by Caselli (2005), and we express the capital stocks in international prices using the investment price deflators from the PWT. We construct the non-agricultural capital stocks by subtracting the agriculture capital from the total capital stocks used by Caselli (2005). We end up with estimates of both output and capital per worker, by sector, for 28 countries.

**Table 8: Role of Capital in Accounting for Sector Productivity Differences**

| Source | Sector | $success_1$ | $success_2$ |
|---|---|---|---|
| Our calculations | Agriculture | 0.22 | 0.12 |
| (n=28) | Non-agriculture | 0.29 | 0.50 |
| Caselli (2005) | Agriculture | 0.15 | 0.09 |
| (n=65) | Non-agriculture | 0.59 | 0.63 |

**Note:** Authors' calculations using data from Butzer, Mundlak, and Larson (2010) and Caselli (2005).

Table 8 reports our findings for the role of capital per worker differences in accounting for sector productivity differences. Here we employ Caselli's (2005) preferred metrics for the "success" of capital per worker differences. The first, $success_1$, is defined as the ratio of log variance in output per worker in a world with only capital per worker differences, divided by the actual log variance. The second, $success_2$, is defined as the 90-10 ratio of output per worker in a world with just capital per worker differences compared with the actual 90-10 ratio. The idea behind both of these metrics is that the lower they are, the larger is the role for TFP differences in explaining output per worker differences. For comparison, we also reproduce the results of Caselli (2005) (Table 5).

Our calculations suggest that TFP differences are the key component of output per worker differences and they seem to play an even larger role in explaining agriculture productivity differences across countries than in non-agriculture. As one can see in Table 8, by either metric, capital per worker differences far from fully account for sector productivity differences in either sector. For $success_1$, we find a ratio of 0.22 in agriculture and 0.29 in non-agriculture. For $success_2$, we find an even lower 0.12 in agriculture and 0.50 in non-agriculture. These calculations paint a very similar picture to those of Caselli (2005), even though we employ different methodology and a different set of countries.

# C. Data Appendix

## 3.1. Data Sources

- **GDP Per Worker** – From the Penn World Table version 6.2., variable "rgdpch".
- **Employment Share in Agriculture** — From the (online) FAO Statistical Yearbook 2004.
- **Agriculture Share in GDP** — These data come from Table G.1 in the FAO Statistical Yearbook online edition.
- **Relative Agriculture Prices** — Derived from author's calculations with original data from the World Bank's 2005 International Comparison Program online database. The sector "agriculture" is defined to be food and non-alcoholic beverages, alcoholic beverages and tobacco, codes (1101 and 1102). "Non-agriculture" is defined as all individual consumption, code (11), gross fixed investment, code (15), minus food, non-alcoholic beverages, alcoholic beverages and tobacco.
- **U.S. Cross-sectional Wage Data** — Our data come from the 2010 U.S. Current Population Survey (CPS), which is the most recent available. Our sample includes all individuals who have non-missing data on income and hours worked. We calculate each individual's wage as the sum of salary income, business income and farm income in the previous year divided by hours worked in the previous year. We restrict the sample to include only those earning at least the Federal minimum wage. We define agricultural workers to be those whose primary industry of employment is agriculture, forestry or fishing, and non-agricultural workers to be all other workers. All calculations are weighted using each individual's inverse probability of being sampled.
- **U.S. Height Data** — These data are taken from the 2009 National Health Interview Survey, a nationally representative survey of Americans conducted by the Center for Disease Control and Prevention (CDC). The data are freely available from the CDC website (http://www.cdc.gov/datastatistics/).

## References


ACEMOGLU, D., S. JOHNSON, AND J. ROBINSON (2001): "The colonial origins of comparative development: An empirical investigation," *American Economic Review*, 91(5), 1369–1401.

ACEMOGLU, D., S. JOHNSON, AND J. ROBINSON (2002): "Reversal of Fortune: Geography and Institutions in the Making of the ModernWorld Income Distribution," *Quarterly Journal of Economics*, 117(4), 1231–94.

ADAMOPOULOS, T. (2009): "Do Richer Countries Have Higher Distribution Margins?," Unpublished Manuscript, York University.

ADAMOPOULOS, T., AND D. RESTUCCIA (2010): "The Size Distribution of Farms and International Productivity Differences," Unpublished Manuscript, University of Toronto.

ALESINA, A., P. GIULIANO, AND N. NUNN (2011): "Fertility and Plough," NBER Working Paper No. 16718.

ALLEN, R. (1994): "Agriculture During the Industrial Revolution," in *The Economic History of Britain Since 1700*, ed. by R. Floud, and D. McCloskey, vol. 1, chap. 5, pp. 96–122. Cambridge University Press.

ATKESON, A., AND M. OGAKI (1996): "Wealth-Vary Intertemporal Elasticities of Substitution: Evidence from Paneland Aggregate Data," *Journal of Monetary Economics*, 38, 507–534.

BARRO, R. J. (2000): "Inequality and Growth in a Panel of Countries," *Journal of Economic Growth*, 5(1), 5–32.

BEEGLE, K., J. DE WEERDT, AND S. DERCON (2011): "Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey," *Review of Economics and Statistics*, 93(3), 1010–1033.

BUTZER, R., Y. MUNDLAK, AND D. F. LARSON (2010): "Measures of Fixed Capital in Agriculture," World Bank Policy Research Working Paper Number 5472.

CASE, A., AND C. PAXSON (2008): "Height, Ability, and Labor Market Outcomes," *Journal of Political Economy*, 116(3), 499–532.

CASELLI, F. (2005): "Accounting for Cross-Country Income Differences," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf.

CASELLI, F., AND W. J. COLEMAN (2001): "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, 109(3), 584–616.

Chanda, A., and C.-J. Dalgaard (2008): "Dual Economies and International Total Factor Productivity Differences:Channelling the Impact from Institutions, Trade and Geography," *Economica*, 75, 629–661.

Eaton, J., and S. Kortum (2002): "Technology, Geography and Trade," *Econometrica*, 70(5), 1741–1779.

Foster, A. D., and M. R. Rosenzweig (1996): "Comparative Advantage, Information and the Allocation of Workers to Tasks:Evidence from Agricultural Labor Markets," *Review of Economic Studies*, 63, 347–374.

Goldin, C., and K. Sokoloff (1982): "Women, Children, and Industrialization in the Early Republic: Evidence from the Manufacturing Censuses," *Journal of Economic History*, 42(4), 741–774.

——— (1984): "The Relative Productivity Hypothesis of Industrialization: The American Case, 1820 to 1850," *Quarterly Journal of Economics*, 99(3), 461–487.

Gollin, D. (2008): "Nobody's business but my own: Self-employment and small enterprise in economic development," *Journal of Monetary Economics*, 55(2), 219–233.

Gollin, D., D. Lagakos, and M. E. Waugh (2011a): "The Agricultural Productivity Gap in Developing Countries," Unpublished Manuscript, Arizona State University.

——— (2011b): "Selection, Agriculture, and the Gains from Trade," Unpublished manuscript, New York University.

Guvenen, F. (2009): "An empirical investigation of labor income processes," *Review of Economic Dynamics*, 12(1), 58–79.

Guvenen, F., and B. Kuruscu (2009): "A Quantitative Analysis of the Evolution of the U.S. Wage Distribution:1970-2000," *NBER Macro Annual*, Forthcoming.

Hall, R. E., and C. I. Jones (1999): "Why Do Some Countries Produce So Much More Output per Worker than Others?," *Quarterly Journal of Economics*, 114(1), 83–116.

Heathcote, J., F. Perri, and G. Violante (2009): "Unequal we Stand: An Empirical Analysis of Economic Inequality in the UnitedStates, 1967-2006," Prepared for *Review of Economic Dynamics* Special Issue.

Heckman, J. J., and B. Honoré (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121–1149.

HECKMAN, J. J., AND G. SEDLACEK (1985): "Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market," *Journal of Political Economy*, 93(6), 1077–1125.

HERRENDORF, B., AND A. TEIXEIRA (2011): "Barriers to Entry and Development," *International Economic Review*, 52(2), 573–602.

HERRENDORF, B., AND A. VALENTINYI (2009): "Which Sectors Make the Poor Countries So Unproductive?," Unpublished Manuscript, Arizona State University.

HORTON, S. (ed.) (1996): *Women and Industrialization in Asia*. Routledge.

HSIEH, C.-T., AND P. J. KLENOW (2007): "Relative Prices and Relative Prosperity," *American Economic Review*, 97(3), 562–585.

JONES, C. I. (1994): "Economic growth and the relative price of capital," *Journal of Monetary Economics*, 34(3), 359–382.

KLENOW, P. J., AND A. RODRÍGUEZ-CLARE (1997): "The Neoclassical Revival in Growth Economics: Has it Gone Too Far?," in *NBER Macroeconomics Annual 1997*, ed. by B. S. Bernanke, and J. Rotemberg. MIT Press, Cambridge, MA.

KUZNETS, S. (1955): "Economic Growth and Income Inequality," *American Economic Review*, 45, 1–28.

LANZONA, L. A. (1998): "Migration, Self-Selection and Earnings in Phillipine Rural Communities," *Journal of Development Economics*, 56, 27–50.

MIGUEL, E., AND J. HAMORY (2009): "Individual Ability and Selection into Migration in Kenya," UN Human Development Research Paper 2009/45.

NELSEN, R. (2006): *An introduction to copulas*. Springer Verlag.

PITT, M. M., M. R. ROSENZWEIG, AND N. HASSAN (1990): "Productivity, Health, and Inequality in the Intrahousehold Distribution of Food in Low-Income Countries," *American Economic Review*, 80(5), 1139–56.

——— (2010): "Human Capital Investment and the Gender Division of Labor in a Brawn-Based Economy," Yale Economic Growth Center Discussion Paper No. 989.

PSACHAROPOULOS, G., AND Z. TZANNATOS (eds.) (1992): *Case Studies on Women's Employment and Pay in Latin America*. The World Bank.

RAO, P. (1993): "Intercountry comparisons of agricultural output and productivity," *Etude FAO: Developpement Economique et Social (FAO)*.

Ravallion, M., S. Chen, and P. Sangraula (2009): "Dollar a Day Revisited," *World Bank Economic Review*, 23(2), 163–184.

Restuccia, D., and C. Urrutia (2001): "Relative prices and investment rates," *Journal of Monetary Economics*, 47(1), 93–121.

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

Rosenweig, M. R., and K. I. Wolpin (1993): "Credit Market Constraints, Consumption Smoothing, and the Accumulation ofDurable Production Assets in Low-Income Countries: Investments in Bullocksin India," *Journal of Political Economy*, 101, 223–244.

Rosenzweig, M. R. (1978): "Rural Wages, Labor Supply, and Land Reform: A Theoretical and Empirical Analysis," *American Economic Review*, 68(5), 847–861.

Rosenzweig, M. R., and R. Evenson (1977): "Fertility, Schooling, and the Economic Contribution of Children in RuralIndia: An Econometric Analysis," *Econometrica*, 45(5), 1065–1079.

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

Schultz, T. (1953): *The economic organization of agriculture*. McGraw-Hill.

Steckel, R. H. (1995): "Stature and the Standard of Living," *Journal of Economic Literature*, 33(4), 1903–1940.

Strauss, J., and D. Thomas (1998): "Health, Nutrition, and Economic Development," *Journal of Economic Literature*, 36, 766–817.

Summers, R., and A. Heston (1991): "The Penn World Table (Mark 5): An Expanded Set of International Comparions,1950-1988," *The Quarterly Journal of Economics*, 106.

Tombe, T. (2011): "The Missing Food Problem: How Low Agricultural Imports Contribute to International Income and Productivity Differences," Unpublished Manuscript, University of Toronto.

Valentinyi, A., and B. Herrendorf (2008): "Measuring factor income shares at the sectoral level," *Review of Economic Dynamics*, 11, 820–835.

Vollrath, D. (2009): "How Important are Dual Economy Effects for Aggregate Productivity?," *Journal of Development Economics*, pp. 325–334.