Specialization, Economic Development and Aggregate Productivity Differences

David Lagakos
Arizona State University

Michael E. Waugh
New York University

September 2009

PRELIMINARY AND INCOMPLETE,
Comments Welcome

ABSTRACT
Cross-country labor productivity differences are large in agriculture and much smaller in non-agriculture. We argue that these relative productivity differences arise when subsistence consumption needs prevent workers in poor countries from specializing in the sector in which they are most productive. We formalize our theory by embedding the Roy (1951) model of ability into a two-sector general-equilibrium growth model in which the agents’ preferences feature a subsistence food requirement. A parameterized version of the model predicts that output per worker gaps will be substantially larger across countries in agriculture than non-agriculture even though countries differ only by a sector-neutral efficiency term.

Email: lagakos@asu.edu, mwaugh@stern.nyu.edu. For helpful comments we thank Francesco Caselli, Berthold Herrendorf, Narayana Kocherlakota, Erzo Luttmer, Ellen McGrattan, Guillermo Ordoñez, Richard Rogerson, Jim Schmitz and seminar participants at Arizona State, the Federal Reserve Bank of Minneapolis, the 2009 Midwest Macro Meetings, the NBER Summer Institute Economic Growth group, the SED (Istanbul), and the SITE Growth and Development session. All errors are our own.
1 Introduction

Cross-country labor productivity differences are large in agriculture and much smaller in non-agriculture relative to aggregate differences (Caselli, 2005; Restuccia, Yang and Zhu, 2008). Development accounting exercises have shown that these sector productivity differences are key in accounting for aggregate productivity differences. If agricultural labor productivity were hypothetically raised to the U.S. level in every country, or if the share of labor in agriculture were hypothetically lowered to the U.S. level, then international variation in aggregate productivity would be virtually eliminated (Caselli, 2005). These results suggest that understanding productivity differences in agriculture and non-agriculture are at the heart of understanding world income inequality.

In this paper we provide a theory of relative labor productivity differences in the agriculture and non-agriculture sectors. We argue that these relative productivity differences arise when sector-neutral efficiency differences combine with subsistence food consumption needs to generate variation in the extent to which workers specialize in the sector where they are most productive.

The basic idea is that countries with low sector-neutral efficiency must deploy a large fraction of their workforce into the agriculture sector to satisfy subsistence food needs. As a result many of those working in agriculture are those whose comparative advantage is not in agricultural work, but rather in non-agricultural tasks such as writing newspaper articles, doing economic research, or teaching yoga classes. In countries with high sector-neutral efficiency a smaller fraction of workers are in agriculture, and those remaining in agriculture are those who are relatively most productive at farm work. As a result, physical productivity differences are larger in agriculture than in the aggregate, and smaller in non-agriculture than the aggregate.

We formalize our theory by embedding the Roy (1951) model of ability into a simple two-sector general-equilibrium growth model. Our theory has two main ingredients. First, workers are heterogenous in their ability to produce output in the two sectors. Second, preferences have a subsistence food requirement. Countries differ only in a sector-neutral efficiency term; preferences and the distribution of ability are identical across countries. Qualitatively, the model can generate productivity differences in agriculture are larger than aggregate differences, and non-agriculture productivity differences are smaller than in the aggregate. The novel feature of this result is it follows from optimal behavior only, as opposed to exogenous country-specific sectoral productivity differences, or barriers to agricultural production, as emphasized by other studies (e.g. Restuccia, Yang, Zhu, 2008).

Our main question of interest is whether the theory can generate quantitatively large differences in agriculture productivity compared to those in non-agriculture. To answer this question we
calibrate the model using parametric assumptions on the distribution of worker ability and observations on the distribution of wages in agriculture and non-agriculture in current U.S. data. Intuitively, the dispersion in ability for each sector is pinned down by the variances of wages for workers within the two sectors, while the correlation of the ability draws is disciplined by the ratio of average sector wages.

Our main exercise is to allow sector-neutral efficiency to vary across countries and study the implied differences in agriculture and non-agriculture productivity. When we feed sector-neutral efficiency differences into our model generating a factor of 22 difference in aggregate income, which corresponds to the differences between the 90th and 10th percentile of country income distribution, the model predicts a factor of 39 gap in agriculture labor productivity, and a factor of 9 gap in non-agriculture labor productivity. In the data there is a factor of 45 gap in agriculture and a factor of 4 gap in non-agriculture. Thus, our model explains nearly seventy five percent of the productivity differences in agriculture and non-agriculture between the 90th and 10th percentile of countries. We find that our model explains less of the sector productivity differences between the 90th percentile and countries with intermediate income levels, because differences in shares of labor in agriculture (and hence specialization differences) are smaller between these countries. We conclude that a completely frictionless economy with countries differing only by sector-neutral efficiency can generate meaningfully large differences in agriculture productivity and small differences in non-agriculture productivity between the richest and poorest countries.

We also show that the model performs quantitatively well in matching relevant development facts for the cross-section of counties. Specifically, the model matches the relationship between income per capita and the share of labor and GDP in agriculture, and performs moderately well in matching the relationship between income per capita and relative agricultural prices. We show that relative agricultural prices are higher in poor countries, with countries around the 10th percentile of the income distribution having prices around 2.5 times higher than countries in the 90th percentile. The model predicts this ratio should be around 4. We also compare our model’s predictions relative to the U.S. structural transformation and show how it replicates the non-linear relationship between relative agriculture prices and the share of employment in agriculture, both of which have steadily fallen over the last century.

The large quantitative effects delivered by our model result in large part from the large variance in ability across agents in our parameterized model. While this ability variance is disciplined by the wage distribution in the United States, some economists have argued that some fraction of wage dispersion is unrelated to ability differences. Postel-Vinay and Robin (2002), for example, argue that around half of wage variation is due instead to labor market imperfections. To address this concern we consider a more conservative calibration of our model with half the ability variance as the baseline calibration. In this more conservative calibration we
still find that our model explains around 50 percent of the productivity differences in agriculture and non-agriculture between the 90th and 10th percentile of countries, and makes other cross-country predictions that are in line with the data.

We conclude by providing direct evidence that our mechanism was at work in the development experiences of the United States and Britain. Two dimensions in which sector abilities are observable are sex and age: historians and development economists have argued that women and children have a comparative disadvantage in agricultural work relative to adult men. Our theory thus predicts that, during a structural transformation, women and children leave farm work and enter the industrial sector at a faster rate than men. We cite evidence that this is fact what happened in Britain and the United States in the 18th and 19th centuries.

Our theory has new implications for the way economists think about agricultural productivity in the developing world. Concretely, our model suggests that low aggregate productivity is not caused by large fractions of workers working in the relatively unproductive agriculture sector. Instead, low measured productivity in agriculture and large agricultural labor shares are consequences of low sector-neutral productivity. The distinction is important because it helps determine the extent to which future research efforts on aggregate productivity differences should focus on the determinants of productivity in agriculture per se, as opposed to more general potential determinants. The policy implications between the two views are different as well. While accounting exercises suggest that fixing agriculture is crucial to raising overall productivity, our theory predicts that improvements in technology, institutions, or social infrastructure (Hall and Jones, 1999; Acemoglu, Robinson and Johnson, 2002) are the key to improving living standards.

2 Agriculture’s Role in Accounting for Aggregate Productivity

In this section we highlight the important role of agricultural in understanding aggregate productivity. Specifically, we reproduce the findings of Caselli (2005) to illustrate how differences in labor productivity and shares of workers in agriculture account for much of the variation in aggregate output per worker across countries.

Panel A of Table 1 shows that labor productivity differences in agriculture are larger than aggregate differences, and that non-agriculture productivity differences are much smaller. The ratio of agricultural output per worker in the 90th and 10th percentiles of the income distribution is 45, compared to just 4 in non-agriculture. As a frame of reference, the ratio for the aggregates is 22. Panel B summarizes the well-known fact that poor countries have a much larger fraction of their work force in agriculture. A country whose per-capita income is in the 90th percentile
Table 1: Agriculture and Labor-Productivity Accounting

Panel A: Labor Productivity Differences

<table>
<thead>
<tr>
<th>Sector</th>
<th>Ratio of 90th-10th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate</td>
<td>22</td>
</tr>
<tr>
<td>Agriculture</td>
<td>45</td>
</tr>
<tr>
<td>Non-Agriculture</td>
<td>4</td>
</tr>
</tbody>
</table>

Panel B: Percent of Labor in Agriculture

<table>
<thead>
<tr>
<th>Country Income Percentile</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>90th</td>
<td>2.8</td>
</tr>
<tr>
<td>10th</td>
<td>78.3</td>
</tr>
</tbody>
</table>

Source: Caselli (2005)

has just 2.8% of its workers in agriculture, while the 10th percentile country has 78.3% of its workers in agriculture.

These two facts together suggest that labor productivity differences in the aggregate are almost completely accounted for by differences in agricultural productivity and shares of labor in agriculture. Caselli formalizes this argument by computing the hypothetical variance of cross-country aggregate output per worker assuming that agricultural productivity in all countries were equal to the U.S. level. His answer is just a factor of 1.6, down from the actual factor of 22! In other words, international labor productivity differences would be virtually eliminated. A similar experiment computes the hypothetical variance of aggregate output per worker assuming all countries had the U.S. share of workers in agriculture. This experiment yields a factor of 4.2 differences between the 90th and 10th percentile, which again is vastly lower than the 22 seen in the data.

One potential explanation for labor productivity differences in agriculture is physical capital per worker differences across countries. Caselli argues that labor productivity differences almost entirely represent total-factor productivity (TFP) differences. As he puts it, “the factor-only model explains virtually nothing of the observed per-capita income variance in agriculture: it’s entirely a story of TFP differences, even more so than for aggregate GDP.” (Caselli, 2005, page 49.) In independent work, Chanda and Dalgaard (2008) perform a similar set a counterfactual exercises using capital stock data from agriculture and non-agriculture, and conclude that around 85% of international TFP differences can be accounted for by TFP differences in agriculture relative to non-agriculture.
The results of this section suggest that cross-country differences in agriculture productivity and the share of workers in agriculture are central factors in accounting for aggregate productivity differences. It remains an unanswered question why the agriculture sector exhibits so much more variation in productivity across countries than the non-agriculture sector.

## 3 Model of Relative Agriculture Productivity

In this section we develop a model of productivity differences in agriculture relative to non-agriculture. We show that in the model sector-neutral efficiency differences across countries can generate relatively larger productivity differences in agriculture and relatively smaller differences in the non-agriculture sector than the efficiency differences themselves.

### 3.1 Households

There are measure one of agents, indexed by $i$, who differ by ability, as will be explained below. Preferences are given by

$$U^i = \log(c^i_a - \bar{a}) + \nu \log(c^i_n),$$

(1)

where $c^i_a$ is food consumption, $c^i_n$ is non-food consumption, $\bar{a}$ is a parameter representing a subsistence food requirement, and $\nu$ governs the relative taste for non-food consumption.

Each agent is endowed with one unit of time which she supplies inelastically to the labor market. Each agent is also endowed with a vector of abilities $\{z^i_a, z^i_n\}$ which represent the efficiency of one unit of labor in sectors $a$ and $n$. The population density of abilities is drawn from a distribution $G(z_a, z_n)$ with support on the positive reals, positive variance for each ability, and imperfect correlation between the two draws. Agents earn wage income $w^i$, which is described in more detail below. The budget constraint is

$$p_a c^i_a + c^i_n \leq w^i$$

(2)

where $p_a$ is the relative price of food, and the non-agricultural good is taken as the numeraire.

### 3.2 Production

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

\[ Y_a = A\tilde{L}_a \quad \text{and} \quad Y_n = A\tilde{L}_n \]  

in agriculture and non-agriculture, where \( A \) captures sector-neutral efficiency, and \( \tilde{L}_a \) and \( \tilde{L}_n \) represent the total number of effective labor units employed in the two sectors. Let \( \Omega^a \) and \( \Omega^n \) denote the sets of agents electing to work in agriculture and non-agriculture. The sector aggregate labor inputs \( \tilde{L}_a \) and \( \tilde{L}_n \) are defined as

\[ \tilde{L}_a \equiv \int_{i \in \Omega^a} z^i_a \, dG_i \quad \text{and} \quad \tilde{L}_n \equiv \int_{i \in \Omega^n} z^i_n \, dG_i \]

and represent the sum of all ability working in the respective sectors. Notice that our labor input differs from those of standard macro models in that ours sums up worker productivities, rather than workers themselves. The total number of workers in each sector is defined as

\[ L_a \equiv \int_{i \in \Omega^a} dG_i \quad \text{and} \quad L_n \equiv \int_{i \in \Omega^n} dG_i. \]

### 3.3 Optimization and Equilibrium

Agents take as given prices and a wage schedule which maps abilities into sector-specific wage offers. The problem for an agent is first to pick which sector to work in, and then to maximize (1) subject to (2). Because of competition in production markets, the schedule of wages offered to a worker with abilities \( z^i_a \) and \( z^i_n \) is equal to:

\[ w^i_a = p_a A z^i_a \quad \text{and} \quad w^i_n = A z^i_n \]

in the agricultural and non-agricultural sectors. A simple cutoff rule in relative ability determines the optimal occupational choice for each agent. Working in non-agriculture is optimal for agent \( i \) if and only if

\[ \frac{z^i_n}{z^i_a} \geq p_a. \]

Thus, the agents that enter non-agriculture are those whose ability there is sufficiently high relative to their ability in agriculture. Let the resulting wage under the optimal sector choice be defined as \( w^i \equiv \max\{w^i_a, w^i_n\} \).

The remainder of the agent’s problem is standard, and optimal demands are:

\[ c^i_a = \frac{w^i + \bar{a} p_a \nu}{p_a (1 + \nu)} \quad \text{and} \quad c^i_n = \frac{\nu (w^i - \bar{a} p_a)}{1 + \nu}. \]
Due to the subsistence consumption constraints, agents consume relatively more food when their wage is lower. The lower is $\nu$, the higher the ratio of food to non-food consumption when the agent’s wage is low.

An equilibrium of the economy consists of a relative food price, $p_a$, and allocations for all agents such that labor and output markets clear. Labor productivity in equilibrium is given by $Y_a/L_a$ in agriculture, and $Y_n/L_n$, and represent the physical quantity of output produced per worker in each sector.

### 3.4 Relative Price of Agriculture Higher in Poorer Economies

In this section we show that, in equilibrium, the relative price of agriculture declines in the efficiency level, $A$.

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

To see the intuition for why $p_a^P$ has to be higher than $p_a^R$, imagine in contradiction that they were the same. For expositional purposes, assume markets clear in the rich country. Then, by (5), the sector labor supply cutoffs would be the same in both countries, and hence so would the share of workers electing to supply labor in the agriculture sector. But because of the subsistence food requirement, the poorer economy demands a much larger fraction of food. Hence output markets would not clear in the poor economy. In order to induce enough workers to supply labor in agriculture in the poor economy, it must be true that $p_a^P$ is greater than $p_a^R$.

Figure 1 illustrates optimal sector choice in equilibrium. Each point on the figure represents one conceivable draw of $(z_a, z_n)$, corresponding to a pair of sector-specific abilities. The dotted lines stemming from the origin describe the set of ability pairs for which agents are indifferent between the two sectors, i.e. when $z_n^i/z_a^i$ equals $p_a^P$ and $p_a^R$ respectively. Points above the lines represent agents for which working in sector $n$ is optimal, and points below the lines meaning that working in $a$ is optimal. As in Proposition 1, because $p_a^P > p_a^R$, more agents work in non-agriculture in the richer economy. The shaded regions describe the set of agents that spend more than half their income on food.\(^1\) The poor economy has a larger fraction of such agents because of the subsistence food requirement.

\(^1\)The choice of one half income spent on food is arbitrary, and just meant to convey the higher food share in the poor country.
3.5 Productivity Differences in Agriculture and Non-agriculture

Our hypothesis is that sector neutral productivity differences lead to larger differences in output per worker in agriculture, and smaller differences in output per worker in non-agriculture. In this section we illustrate how this can occur, using a simple example where households differ only in their ability in one of the two sectors.

Consider the case when the ability distribution, $G(z_a, z_n)$, has $z_n = 1$ with probability one, and a density of $z_a$ is given by $g(z_a)$. Now, by (5), we see that households enter agriculture if and only if $z_a^i > \frac{1}{p_a}$. Figure 2 illustrates the ability distribution in two economies, one poor (low $A$) and one rich (high $A$), and which agents end up working in each sector. We know that the relative price of food is higher in the poor country, and hence $1/p_a^R < 1/p_a^P$. Thus more agents enter agriculture in the poor country, and the average ability of agriculture workers is lower. This means $Y_a^R/L_a^R < Y_a^P/L_a^P$, or in other words agriculture productivity differences are greater than the underlying efficiency differences.

While this simple example is useful for illustrative purposes, it has at least two major limitations. First, non-agriculture productivity differences will be proportional to differences in $A$. Thus, this model has very limited ability explain why non-agriculture output per worker differences are so much smaller across countries than aggregate differences. Second, it gives no sense of whether the mechanism produces economically large differences in agriculture and non-agriculture productivity. To address these two issues we parameterize a version of the model with heterogeneity in ability in both sectors.
4 Quantitative Analysis of the Model

In this section we parameterize the model and then assess its quantitative importance for understanding agriculture and non-agriculture labor productivity differences across countries. We also assess the model’s predictions for cross-country and historical U.S. data on shares of labor and GDP in agriculture, relative prices of agricultural goods, plus the current cross-section of wages in the U.S.

To parameterize the model we must select a distribution of ability, $G(z_a, z_n)$, plus values for the two taste parameters $\bar{a}$ and $\theta$; the efficiency term $A$ can be normalized to 1 for the United States. We discuss each in turn.

4.1 Parameterization of Ability Distribution

We set the joint distribution of abilities to be:

$$G(z_a, z_n) = (F(z_a)^{-\rho} + H(z_n)^{-\rho} - 1)^{\frac{1}{1-\rho}}$$

with $F(z_a) = e^{-z_a^\theta a}$ and $H(z_n) = e^{-z_n^\theta n}$.
The functions $F(z_a)$ and $H(z_n)$ are the cdfs of Fréchet random variables.\(^2\) The lower are $\theta_a$ and $\theta_n$, the higher is the variance of ability in agriculture and non-agriculture. Dependence of the ability draws is induced using the function $(u^{-\rho} + v^{-\rho} - 1)^{\frac{1}{\rho}}$, which is known as a Clayton copula. The parameter $\rho \in (-1,0) \cup (0,\infty)$ determines the extent of dependence, with a higher $\rho$ representing more dependence between draws.\(^3\)

Two natural questions are: why we chose a parametric form for the ability distribution, as opposed to something nonparametric, and why we chose this particular parametric form. The answer to the first question is that the Roy model, in spite of its apparent simplicity, cannot be identified from cross-sectional wage data without making assumptions on the functional form of the ability distribution (Heckman and Honore, 1990.) Because one only observes the maximum of each agents’ draws, but not both draws themselves, if ability distributions are allowed to take on an arbitrary form, there are many distributions which can generate a given set of observations on wages and sector choices by individuals.

There are three main justifications for this particular parametric form of the ability distribution. First, it parsimoniously allows for (potentially different) ability dispersion in each sector as well as dependence between ability draws. As will be shown in the following section, the three parameters of the distribution ($\theta_a$, $\theta_n$ and $\rho$) can each be disciplined in a transparent way using a single cross section of wages for agricultural and non-agricultural workers.

Second, the choice of Fréchet distributions for ability 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. Thus the draw $z_{i,n}$, for example, can be thought of as the maximum of household $i$’s ability draws in a large set of distinct non-agricultural tasks. A similar interpretation can be given to $z_{i,a}$.\(^4\)

Finally, Fréchet distributions for ability yield wage distributions – for the economy as a whole and by sector – which closely resemble their empirical counterparts, as we demonstrate in the following section. In particular, the model delivers wage distribution tails that mimic the data extremely closely, a dimension along which other distributions fail. Heckman and Honore (1990), for example, argue that a Roy model with lognormal ability distributions generates tails

---

\(^2\)This distribution has been used to parameterize Ricardian models of international trade originating with Eaton and Kortum (2002). It is also very convenient. In trade models, the Fréchet distribution yields a log-linear gravity equation relating trade flows to structural parameters. Similarly in our framework with independence and a common $\theta$ parameter, the Roy model yields a log-linear equation relating employment shares, the relative price of agriculture goods, and structural parameters.

\(^3\)Copulas can be used to create multivariate distributions out of arbitrary univariate distributions. See e.g. Cherubini, Luciano and Vecchiato (2004).

\(^4\)By the extreme value theorem, the maximum of a sample of i.i.d. draws from any distribution converges in distribution to one of three extreme value distributions: the Fréchet, the Gumbel, or the Weibull. See e.g. Kotz and Nadarajah (2000).
which are too thin compared to the data. Since our results are sensitive to the size of the tails of the ability distribution, it is essential that tails in the model are in line with the data.

To calibrate the ability distribution parameters, our basic strategy is to use cross sectional wage data from the United States. Formally, we jointly determine $\theta_a$, $\theta_n$ and $\rho$ to match three moments, the standard deviations of log wages in agriculture and non-agriculture plus the ratio of average wages in agriculture and non-agriculture. Intuitively, the our calibration strategy is as follows. The $\theta_a$ and $\theta_n$ terms determine the variation in ability across individuals, with higher $\theta$’s resulting in lower variation in abilities. Because wages are set equal to the value of marginal products, variation in ability maps into variation in wages across agents. Thus, observed wage variation in agriculture and non-agriculture wages are key in disciplining the parameters $\theta_a$ and $\theta_n$.

Next, $\rho$ is pinned down by the average wage in agriculture relative to non-agriculture, with a higher relative agriculture wage implying a higher $\rho$. Briefly, this is because when $\rho$ is higher, if an agent has a low draw in one sector it is likely to have a low draw in the other sector as well. Thus, the higher is $\rho$, the more of agents with two low draws there are in non-agriculture, and hence the lower are wages in non-agriculture compared to agriculture. This can be seen in Figure 3(a). Figure 3(a) plots a simulation with each point in the plane depicting the two wages that a household can receive. If a household is to the left of the 45° line than it is optimal to work in agriculture and vice versa. Figure 3(b) plots a simulation with no correlation. The key thing to notice is that workers in agriculture earn higher wages (compared to non-agriculture) in the model with dependence relative the model with no dependence. What this implies is that the dependence parameter $\rho$ affects the ratio of the average wages in the two sectors. Hence we pick $\rho$ to target this moment.

![Figure 3: Wages Offers From Model with Dependence and No Dependence](image)
We get our cross-sectional data from the U.S. Current Population Survey (CPS) for 2007. Following the study of U.S. wage inequality by Heathcote, Perri and Violante (2009), we take all individuals between 25 and 60 who have non-missing data on income and hours worked. We restrict the sample further to include only workers averaging at least 35 hours per week of work, and only those earning at least the Federal minimum wage. These restrictions provide more conservative estimates of cross-sectional wage variance, which will lead to more conservative variances of ability in the parameterized model (and hence less predictive power for our mechanism). We calculate that the standard deviation of log wages for agriculture workers is 0.45, while in non-agriculture the standard deviation of log wages is higher at 0.56. The average wage in agriculture is 0.77 times the average wage in non-agriculture.

These figures imply (along with the preference parameters calibrated below) choices of $\theta_a = 3.13$, $\theta_n = 2.22$ and $\rho = 0.15$. Since $\rho$ itself is hard to interpret (it runs from 0 to $\infty$), we computed the Kendall rank correlation coefficient to be 0.10. This suggests that there is only moderate amount of correlation across individuals in their abilities across the two activities. The estimates of $\theta_a$ and $\theta_n$ mean that there is more variance in ability in non-agriculture work than in agricultural work, again reasonable given that non-agriculture work encompasses so many more types of tasks.

### 4.2 Parameterization of Preferences

For the preference parameters, we pick $\nu$ to match average share of labor in agriculture across countries in the top 10th percentile of the income distribution. The resulting parameter implies a a long-run food expenditure share of 1.9%. Caselli and Coleman (2001) and Duarte and Restuccia (2009) pick a value of 1%, while Restuccia, Yang and Zhu (2008) pick a value of 0.5%. Admittedly, our model’s results are sensitive to the choice of this value, with lower values allowing us to explain more of the variation in agriculture and non-agriculture labor productivity. For this reason we stick with a more conservative value relative to others in the literature.

We set $\bar{a}$ to match a subsistence consumption need of 34% of average income in a model country with 7.5% of the U.S.’s per capita GDP. This is consistent with the independent estimates of subsistence food consumption requirements of Rosenzweig and Wolpin (1993), and Atkeson and Ogaki (1996), both of which use panel data from a sample of rural households from India (which had 7.5% of the U.S. per capita GDP in 1984).\(^5\)

---

\(^5\)Rosenzweig and Wolpin estimate a subsistence requirement of 1,469 rupees per agent per year. Townsend (1994) reports that average agent size in the sample is 6.7 and that average income per person in the Indian sample is 635 rupees.
4.3 Quantitative Predictions for Sector Productivity Differences

With the model parameterized, we now ask what it predicts quantitatively for agriculture productivity differences and non-agriculture differences in the cross section of countries. Specifically, we solve the model over a range of $A$ values covering the world income distribution, and compute its predictions for relative output per worker in the aggregate and the two sectors.

Table 2: Labor Productivity Differences

<table>
<thead>
<tr>
<th>Sector</th>
<th>Ratio of 90th-10th Percentile</th>
<th>Data</th>
<th>Model</th>
<th>Percent Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>45</td>
<td>39</td>
<td></td>
<td>74</td>
</tr>
<tr>
<td>Aggregate</td>
<td>22</td>
<td>22</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Non-Agriculture</td>
<td>4</td>
<td>9</td>
<td></td>
<td>72</td>
</tr>
</tbody>
</table>

Data Source: Caselli (2005)

Table 2 shows the model’s predictions for the ratio of the 90th to 10th percentile of countries in the model and data. The differences in aggregate output per worker (expressed as GDP per worker at Gheary-Khamis international prices) is a factor of 22 in the model and data by construction. The model predicts agriculture output per worker differences should be a factor of 39, and in non-agriculture it predicts a factor of 19 difference. In the data these ratios are (as described in Section 2) a factor of 45 and 4 respectively. The third column of the table shows that this corresponds to the model explaining 74% of agricultural differences and 72% of non-agricultural differences, relative to aggregate differences. The results in Table 2 show that differences in patterns of specialization are quantitatively important to understanding relative sector productivity differences between the richest and poorest countries.

Table 3 illustrates the model’s predictions for developing countries with relatively higher average income. Specifically, it shows the model’s prediction for the 90th-50th ratio and 90th-25th ratio. In the latter case aggregate productivity in the model and data differ by a factor of 9.4, again by construction. In the 90th-25th case, the model predicts a factor of 116.4 in agriculture and 7.4 in non-agriculture, compared to 31.1 and 2.7 in the data. The model predicts 32% and 30% of the agricultural and non-agricultural productivity differences, relative to the aggregate, as in the data, which is still large, but substantially lower than the 90th-10th percentile ratio.

In the 90th-50th case, the model fares much worse. The aggregate differences are chosen to be a factor of 3.1 as in the data. The model predicts differences in agriculture and non-agriculture of 3.8 and 3.04, compared to 11.1 and 1.9 in the data. This amounts to explaining just 8% and 5% of the sector differences, respectively.
### Table 3: Labor Productivity Differences – Intermediate Income Levels

<table>
<thead>
<tr>
<th>Sector</th>
<th>Ratio of 90th.-25th Percentile</th>
<th>Percent Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>Model</td>
</tr>
<tr>
<td>Agriculture</td>
<td>31.1</td>
<td>16.4</td>
</tr>
<tr>
<td>Aggregate</td>
<td>9.4</td>
<td>9.4</td>
</tr>
<tr>
<td>Non-Agriculture</td>
<td>2.7</td>
<td>7.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ratio of 90th-50th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
</tr>
<tr>
<td>Agriculture</td>
</tr>
<tr>
<td>Aggregate</td>
</tr>
<tr>
<td>Non-Agriculture</td>
</tr>
</tbody>
</table>

Data Source: Caselli (2005)

Why does the model fare so much less successfully when explaining differences between rich countries and those at intermediate income levels? The answer has to do with differences in the shares of labor in agriculture. Consider the case of the 90-50 differential. The percent of workers in agriculture in the 50th percentile country is just 9%, compared to 2% in the 90th percentile country. Thus, the average productivity of workers in agriculture is only slightly lower in the 50th percentile country. In contrast, in the 10th percentile country, 74% are in agriculture, and thus the average worker has substantially lower productivity than the average agricultural worker in the 90th percentile country. Hence, the model’s explanatory power is larger for differences between the the richest countries and the poorest countries than for the richest and middle income countries.

### 4.4 Assessing The Model’s Other Quantitative Predictions

The model generates large productivity differences in agriculture relative to non-agriculture across countries, at least between the richest and poorest countries. We now ask whether it is successful in matching other relevant data. In particular, we compute the model’s predictions for the share of labor in agriculture, the share of GDP spent in agriculture, the relative price of agriculture, U.S. historical relative prices, and the wages of agriculture workers relative to non-agriculture workers in the cross-section of countries, and assess whether they are quantitatively consistent with the data.
Figure 4: Employment and Agriculture Shares, Data and Model
4.4.1 Predictions for the Cross Section of Countries

Figure 4(a) shows the model’s predictions for the cross-section of countries for the share of labor in agriculture, along with the actual data. The horizontal axis displays purchasing-power parity GDP per worker for 2000 (on a log scale), and the y-axis displays the percent of workers in agriculture. As in the data, our model predicts that the poorest countries should have shares in the range of 70% to 90% of all workers, down to less than 10% for the richest. The model also captures the convex nature of this curve, which is driven in the model by the concavity of preferences along with the subsistence constraint in food. We conclude that this feature of the data is successfully captured by our model.

Next we turn, in Figure 4(b), to the model’s predictions for the share of GDP in agriculture. While similar to the labor shares shown above, note that in the data the GDP shares in agriculture are systematically lower than the labor shares in agriculture. In Kenya, for example, agriculture employs 74% of the workers but produces just 28% of GDP. While our model does a reasonable job of capturing GDP shares for agriculture in the countries with around 1/8 the U.S. income level or higher, it substantially over-predicts the GDP share in agriculture in countries with lower income.

One reason for the model’s inconsistencies with the data in this dimension may be that agricultural GDP itself is mis-measured in the poorest countries. If households spend much of their time in home production of agricultural goods, then measured GDP of agriculture will understate true agricultural output (Gollin, Parente, Rogerson, 2004).

An important prediction of our model is that the relative price of food is higher in poor countries than rich countries. We now ask whether this prediction is borne out in the data, and whether the model is quantitatively consistent with the cross-country relationship between relative food prices and income per capita. Figure 5 plots this relationship. The vertical axis contains the relative price of agricultural goods (expressed in log base 2) with the U.S. value normalized to one and the horizontal axis plots real gdp per worker relative to the U.S. also in log base 2 scale. Our data on relative food prices are constructed using 2005 data available from the International Comparison Programme (see Appendix B for details).

As can be seen in the Figure, relative food prices are indeed higher in lower income countries, as predicted by Proposition 1. The ratio of relative food prices in the 10th percentile of the country income distribution to 90th percentile is about 2.5. In the model it is around 4. We conclude that while model performs well except for the poorest countries, where it over-predicts the relationship found in this data.

One limitation of this comparison is that prices in our model are essentially producer prices, since our model ignores distribution margins. The data, on the other hand, are consumer prices,
which reflect this margin. This might be important, as much of the final price of agricultural products could reflect transportation costs from rural to urban areas or retail services after delivery. Adamopoulos (2008) examines the difference between retail and producer prices for agricultural goods across countries and finds that rich countries have higher distribution margins for food. This suggests that relative producer prices of food might be even higher in poor countries, compared to richer countries, than Figure 5 suggests.

4.4.2 Predictions for U.S. Time Series

Because of the possible measurement issues with the cross-country data, we also compared our models predictions relative to the U.S. time series of relative agriculture prices and labor shares in agriculture.\(^6\) Figure 6 plots the relative price of agriculture on the vertical axis with the U.S. value in 1990 normalized to one. On the horizontal axis, the share of employment in agriculture is reported as well. The model’s prediction is plotted in the solid line. Note that the model correctly picks up the decline in relative prices and the employment share in agriculture. More importantly our model replicates this fact in the non-linear way that the data exhibits.

---
\(^6\)The data appendix describes the source of the data used.
4.4.3 Predictions for U.S. Cross-Sectional Distribution of Wages

A final implication of our model worth testing is the entire distribution of wages within an economy. In our calibration we made parametric assumptions regarding the distribution of abilities and calibrated the parameters to target several moments of the wage distribution. One may be concerned that our functional form choice has unreasonable implications for relative to the entire distribution of wages. Figure 7(a) plots the empirical cumulative distribution function from the U.S. wage data and from data generated by our model. They track each other very closely. Figure 7(b) and 7(c) plot the same relationship for only those workers in agriculture and non-agriculture. The model performs reasonably well, particularly in its upper tails, which closely mimic those in the data. The only substantive deviation is that our calibration suggests there are slightly more low wage workers in than seen in the data. Overall, the Figures show that our parametric assumptions on the ability distribution yield realistic wage distributions.

4.5 More Conservative Calibration

While wage variation in the model arises only because of variation in efficiency of labor across individuals, some economists argue that wage variation in the data is due in part to factors
Figure 7: Wage Distributions, Model and Data
unrelated to ability, such as market imperfections resulting from search frictions. The largest estimate of the importance of these imperfections (that we could find) is that of Postel-Vinay and Robin (2002), who estimate that around one-half of the variance in log wages is due to market imperfections. In an effort to be as conservative as possible in determining the extent of ability differences across individuals, we follow the Postel-Vinay and Robin estimate, implying that the standard deviation of the log wages due to ability differences is 0.32 in agriculture and 0.40 in non-agriculture. We follow the our previous approach by calibrating $\rho$ to best fit the average wage in agriculture relative to non-agriculture.

Calibrating the ability distribution parameters to these moments results in $\theta_a = 3.1$, $\theta_n = 4.4$, and $\rho$ very close to zero (recall that it is not defined exactly at zero). This choice of $\rho$ leads to a slight overestimate the relative average wage suggesting a need for negative dependence in order to match the relative average wage. Since negative dependence increases the model’s explanatory power, we leave $\rho$ close to zero, and note that the predictions of this exercise will be even lower than they would have been with $\rho$ chosen to match the relative wages exactly.

With this conservative calibration, we now ask what it predicts quantitatively for agriculture productivity differences and non-agriculture differences in the cross section of countries. Table 4 presents the results for countries in the 90th and 10th percentile of the income distribution.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Ratio of 90th-10th Percentile</th>
<th>Percent Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>Model</td>
</tr>
<tr>
<td>Agriculture</td>
<td>45</td>
<td>32</td>
</tr>
<tr>
<td>Aggregate</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Non-Agriculture</td>
<td>4</td>
<td>13</td>
</tr>
</tbody>
</table>

Data Source: Caselli (2005)

The model predicts agriculture output per worker differences should be a factor of 32, and in non-agriculture it predicts a factor of 13 difference. In the data these ratios are a factor of 45 and 4 respectively. The third column of the table shows that this corresponds to the model explaining 44% of agricultural differences and 57% of non-agricultural differences, relative to aggregate differences. The key result from this exercise is that our mechanism still explains around half of the variation in sectoral productivity gaps even when we feed into a model which generates only one half the observed variation in wages.

Similar to the baseline calibration, the model is not as successful at explaining gaps for intermediate income levels simply because differences in aggregate labor allocations are not as large.
The other predictions of the model are quantitatively similar as well. The model correctly captures the employment share and agriculture share in GDP. The model’s predictions for relative prices improve with the model generating and elasticity of relative prices with respect to income level of $-0.23$. In the data this elasticity is $-0.21$. The model under-predicts relative to U.S. time series data on share of employment and relative food prices.

### 4.6 Predictions of Simple Alternative Theory

In this section, we study the implications of shutting down the key innovation of our paper—heterogenous workers with occupational choice—and study a simple model with homogenous labor and allow for sector specific total factor productivity differences. We ask whether a simple alternative to our story—namely homogenous workers with sector-specific TFP differences—is consistent with cross-country evidence. We find that it is not. The reason for its failure is because large sectoral productivity differences imply implausibly large differences in relative prices and the model does not generate a log-linear relationship between relative labor shares and relative prices as exhibited in the data.

To illustrate this point, consider the simplification of our model with homogenous labor. Now the production functions are simply

\[ Y_a = A_a A L_a \quad \text{and} \quad Y_n = A_n A L_n \]  

(7)

with \( L_a \) and \( L_n \) representing the number of workers in each sector, not the number of effective labor units as in our model. Furthermore, now production functions have a total factor productivity term that is common \( A \) and a sector specific productivity term \( A_a \) and \( A_n \). Preferences will remain the same as discussed in the text.

The key difficulty with this model is that it dramatically over predicts the amount of variation in relative prices across countries. To see this, note that in this model prices are pinned down entirely by the production side with the relative price being \( p_a = \frac{A_n}{A_a} \). Now comparing the relative price between a country in the 10th percentile and 90th percentile of income per worker, the relative price can be expressed as

\[ \frac{p_a^{10}}{p_a^{90}} = \left( \frac{A_n^{10}}{A_a^{10}} \right) \times \left( \frac{A_a^{90}}{A_n^{90}} \right) \]  

(8)

Plugging in numbers from Table 1 implies that the relative price between the 90th and 10th percentile should be a factor of 11 = .25 \times 45. In the data the relative price between a country in the 90th and 10th percentile is a factor of 2.5 — significantly less than that predicted by the model.
To further see the problems discussed above, we performed the following quantitative experiment. We solved this model over a range of $A, A_o,$ and $A_n$ generating differences in aggregate GDP per worker, agriculture GDP per worker, and non-agriculture GDP per worker similar to that in the data. The preference parameter $\nu$ is calibrated the same as in the text and the subsistence consumption parameter was calibrated to achieve a reasonable fit of labor and GDP shares in agriculture to the data. Figure 8 plots the relative food price against GDP per capita in the data, our model, and the simple model with homogeneous labor. As shown earlier, our model predicts a relative food price around 4 times higher in poor countries than rich countries, compared with a difference of around 2.5 in the data. The simple model predicts that relative prices should be 16 times higher in the poorest countries than the richest, which is implausibly large compared with the data.

5 Historical Evidence: Males versus Females

In this section we provide some direct evidence in support of our theory. While many dimensions of ability heterogeneity are not observable, along two particular observable dimensions, namely age and sex, there is concrete evidence of ability differences in agricultural and non-agricultural tasks. Historians and development economists have argued that women and children generally have a comparative disadvantage at farm work than adult men. As one piece
of evidence, Goldin and Sokoloff (1982, 1984) show that wages were much lower for women in farm work in the United States, earning roughly one third to one half as much as men in farm work in the nineteenth century, with smaller wage differentials in manufacturing work. Foster and Rosenzweig (1996) provide complementary evidence from a sample of farmers from the Philippines, among which they estimate males to have an absolute productivity advantage in several types of agricultural tasks.\footnote{They estimate a one-factor model with two tasks: ploughing and weeding. They find that men are more productive at both, with a larger productivity difference in ploughing.}

Our theory thus predicts that women and children should have been the first to move off the farms and into non-agricultural work. This is in fact what happened. In Britain, according to Allen (1994), the fraction of farmers that were women or children declined substantially during Britain’s industrial revolution. Table 5 shows Allen’s calculations for the composition of farm workers in England and Wales between 1700 and 1851. In 1700, a full 62.0% of farm workers in England were women and children, with the balance adult men. By 1800 this percent fell to 55.3%, and by 1851 it was down to 36.3%. For just women, the same figures are 32.5% in 1700, 30.3% in 1800, and 26.8% in 1851. Men, on the other hand, went from representing just over one third of farm workers to just under two-thirds.

<table>
<thead>
<tr>
<th></th>
<th>1700</th>
<th>1800</th>
<th>1851</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>38.3%</td>
<td>44.7%</td>
<td>63.7%</td>
</tr>
<tr>
<td>Women</td>
<td>32.5%</td>
<td>30.3%</td>
<td>26.8%</td>
</tr>
<tr>
<td>Women and Children</td>
<td>62.0%</td>
<td>55.3%</td>
<td>36.3%</td>
</tr>
</tbody>
</table>

Data Source: Allen (1994)

Goldin and Sokoloff (1982, 1984) provide evidence that this pattern held in the United States as well, using evidence from the manufacturing sector, which was a major component of the non-agricultural economy in the nineteenth century. In 1820, in the Northeast United States, roughly 55% of manufacturing workers were women and children. By 1890, this figure was down to 21%. The interpretation given by Goldin and Sokoloff is that as manufacturing work became available, women took manufacturing jobs at a faster rate than men, who stayed in agriculture work relatively longer.

Furthermore, Goldin and Sokoloff argue that the primary reason women moved into manufacturing relatively faster than men is that women had a comparative disadvantage at agriculture work, just as our theory predicts. To support their argument, Goldin and Sokoloff estimate that in 1820, women earned roughly 30% as much as men in the Middle Atlantic region, and roughly 37%
as much as men in New England. By 1850, they estimate relative wages of 51% in the Middle Atlantic and 46% in New England. While numerous factors were at play in this period, the authors argue that their finding of rising female wages “is consistent with the observations of many contemporaries of the early nineteenth century who reported that the relative productivity (and wages) of women and children compared to adult men was low in the agriculture and traditional sectors of the pre-industrial northeastern economy (1982, page 759).”

As additional support for the comparative advantage theory, Goldin and Sokoloff provide evidence that women faced greater comparative disadvantage in farming in the North than in the South, and entered manufacturing to a much greater extent in the North than in the South. The difference in the comparative disadvantage of women stemmed from the types of farm work common in the two regions. In the North, where strength-intensive wheat farming was prevalent, women earned around one third as much as men in the 1820. In the South, where Cotton and Tobacco farming were most common, women earned around one half as much as men, as dexterity played a more important role in farming these crops. Just as the theory predicts, as the U.S. structural change progressed in the second half of the 19th century, Northern women entered into factory work to a much larger extent than those in the South.

6 Extensions to Trade and Mobility Restrictions

We first discuss how the model behaves once we add the possibility of trade, and second, we discuss an extension which explores the implications of barriers to labor mobility.

6.1 Adding Trade to the Model

In this section we ask how allow the model’s predictions would change once we allow for trade. We draw two conclusions. First, allowing for frictionless trade would introduce strongly counterfactual assumptions about shares of labor in agriculture and relative prices across countries. Thus, to be useful, any extension to allowing trade should have to include trade frictions and account for differences in relative prices across countries. Yet, if a model with trade is consistent with relative agriculture prices, we conjecture that such a model would have a modest effect on the quantitative nature of the model’s predictions.

First consider a version of the model where each country has frictionless access to trade in world markets. Then the following is true.
Proposition 2 Imagine that the rich and poor economies can trade frictionlessly on world markets at a relative food price $p^W$. Then the following must hold:

\[
\frac{Y_a^P}{L_a^P} = \frac{Y_a^R}{L_a^R} \quad \text{and} \quad \frac{L_a^R}{L_a^R + L_n^R} = \frac{L_a^P}{L_a^P + L_n^P}.
\]

Proposition 2 says that under frictionless trade, two things are true. First, the extent of specialization would be the same in both countries, and hence labor productivity differences between the rich and poor countries would be the same in agriculture and non-agriculture. Second, the shares of labor in agriculture would be equated across countries. Both are true because, under a common relative price, the sector labor supply cutoff (5) is identical in both countries, and hence the composition of workers in each sector are identical as well.

But the prediction of labor shares being equal across countries is strongly counterfactual. As is well known, a substantially higher fraction of labor in poor countries is in agriculture than rich countries. In Section 2, for example, we cite evidence that the United States has just 2.8% of its labor in agriculture compared to over 78.3% of the labor in a country at the 10th percentile of the world per-capita income distribution. Thus, we conclude that adding frictionless trade will generate counterfactual predictions.

Nevertheless, we conjecture that adding frictional trade would have a modest quantitative effect on the model’s predictions in these three dimensions. The reason is that any reasonable model of trade would have to be in line with relative agriculture prices across countries, yet the baseline model is consistent with the data in terms of prices. Thus the model’s predictions for productivity differences would be changed little because weather the model is an open economy or not relative prices determine labor allocations, and labor allocations determine predictions about productivity.

6.2 Effects of Policies That Restrict Labor Mobility

Our model implies that government policies that restrict migration within a country are likely to reduce productivity. One component of these policies that has received relatively less attention is the government’s role in selecting which agents may move and which may not. Our theory says that letting the market allocate workers is more efficient than letting the government allocate them.

China is an important example of a growing country that is underdoing structural change and has restricted the flow of workers out of agricultural areas, typically in central China, into urban areas, typically on the eastern coast. One component of this hukou system is that the local and
federal government officials decide which workers are permitted to migrate and which must remain in rural areas (See e.g. Au and Henderson (2006) and Lau, Qian, and Roldand (2000).)\(^8\)

In future work we plan to use the model estimate the potential welfare cost of policies that restrict which individuals may leave agriculture and which may not. A rough idea is as follows. Imagine an economy like China in 1985 with 70% of its workforce in agriculture, and imagine it moves to a new steady state in which GDP per worker is (say) doubled. This corresponds roughly to Chinese growth from 1985 to the present. We can compute the welfare costs under two scenarios which restrict migration out of agriculture. The first randomly picks some fraction \(\omega\) of agricultural agents that may move from agriculture to non-agriculture in such a way that markets clear, and the second lets the market reallocate agent. The experiment is in the same vein as that of Alder (2009), who assesses the efficiency losses from non-assortative matching of managers to projects.

7 Conclusion

We argue that cross-country productivity differences in agriculture are larger than in non-agriculture because of differences in the extent to which workers specialize in sectors in which they are talented. In poor countries, virtually everyone works in agriculture, even though many of those workers have a comparative advantage that is not in farm work, but rather in non-agricultural tasks such as acting, teaching, or writing newspaper articles. In rich countries, in contrast, those remaining in agriculture are those who are relatively most productive at farm work. As a result, labor productivity differences are relatively larger in agriculture than the aggregate, and smaller in non-agriculture, even though countries differ only in general, sector-neutral, efficiency.

Our theory has new implications for the way economists think about agricultural productivity in the developing world. In contrast to other papers that emphasize barriers to efficient production in farming, we argue that low productivity in agriculture could represent the optimal response to low general efficiency in the face of subsistence food requirements. In this case it is optimal to employ many workers in agriculture who are less able in farm labor than other tasks. Concretely, our paper suggests that the source of low agriculture productivity might not be entirely found in the agriculture sector itself. It could, for example, 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, Robinson and Johnson, 2002).

\(^8\)In addition to policies restricting which workers leave agriculture, there is evidence of social norms which restrict worker movements across sectors. Hayashi and Prescott (2008) cite evidence of social customs in the period before World War II which kept first born sons of farmers in agriculture.
A Model Appendix

A.1 Proof of Proposition 1

Let \( p^1_a, Y^1_a \) and \( Y^1_n \) be the equilibrium relative price and quantities in an economy with general efficiency \( A^1 \). Let \( A^2 > A^1 \), and denote by \( p^2_a, Y^2_a \) and \( Y^2_n \) the equilibrium of an economy with efficiency \( A^2 \).

Suppose that \( p^2_a = p^1_a \). Then by (5), each agent \( i \) chooses to work in the same sector in \( A^2 \) as in economy \( A^1 \). Thus output in each sector would be scaled up by a factor equal to the ratio of the efficiency terms: \( Y^2_a/Y^1_a = Y^2_n/Y^1_n = A^2/A^1 \). But by (6), we know that agents must demand a higher fraction of non-agriculture goods in economy \( A^2 \) than \( A^1 \). Thus \( Y^2_n/Y^2_a > Y^1_n/Y^1_a \). But this implies that \( Y^2_n/Y^1_n > Y^2_a/Y^1_a \), which is a contradiction. Thus \( p^2_a \neq p^1_a \).

The only way to be consistent with the agent solutions', (6), is for more agents to supply labor in the non-agriculture sector in economy \( A^2 \) than economy \( A^1 \). By (5), this occurs if and only if \( p^2_a < p^1_a \). ■
B Data Appendix

- **GDP Per Worker** — This data is from the Penn World Table version 6.2. series “rgdpch”.

- **Labor Share in Agriculture** — This data comes from Table A.3 in the FAO Statistical Yearbook 2004 online edition.

- **Agriculture Share in GDP** — This data comes from Table G.1 in the FAO Statistical Yearbook online edition.

- **Relative Agriculture Prices** — This data is 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.

- **Relative Wages** — Wage data is from LABORSTA and the series wage data by economic activity is used. Agriculture corresponds directly with ISIC Revision 2 and 3 categories “Agriculture, hunting and forestry” and “Fishing”.

- **U.S. Historical Relative Prices** — U.S. historical relative prices are from Historical Statistics of the United States Millennial Edition Online, Table Cc125-137 - Wholesale price indexes for historical comparisons, by commodity group: 1860 - 1990. Agriculture price is defined to be farm products and non-food is all commodities other than food products. As an alternative price series, we also explored using a series from Table Cc1-2, Consumer price indexes BLS based, in the denominator instead which yielded similar results. This alternative series is the analog to that used in Caselli Colmen (2005). To match up with observations on employment in Farming, observations corresponding with 1880, 1890, and 1900 are taken to be decade averages.

- **U.S. Historical Farm Population** — U.S. historical farm population are from Historical Statistics of the United States Millennial Edition Online, Table Da14-27 - Farms number, population, land, and value of property: 1850 - 1997. This is taken to be a proxy for the share of employment in the United States in agriculture.
References


