Open access • Journal Article • DOI:10.1257/AER.103.2.948

Selection, Agriculture, and Cross-Country Productivity Differences — Source link

David Lagakos, Michael E. Waugh

Institutions: New York University

Published on: 01 Apr 2013 - The American Economic Review (American Economic Association)

Topics: Productivity, Subsistence agriculture, Agricultural productivity, Subsistence economy and Agriculture

Selection, Agriculture and Cross-Country Productivity Differences
David Lagakos and Michael E. Waugh
Online Appendix

A. Proofs of Propositions and Corollary

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 (3), 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 the demand functions, 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, the demand functions, is for more workers to supply labor in the non-agriculture sector in economy A^R than economy A^P. By (3), 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 (3) 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. Proof of Corollary 1

It suffices to prove that the $E(z_a|z_a/z_n > 1/p_a)$ is decreasing in p_a and $E(z_n|z_n/z_a > p_a)$ is increasing in p_a. To obtain closed-form expressions for the conditional expected productivities in question, one must derive Prob$\{z_n \leq p_a z_a\}$. To do so, note that this probability is represented by

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

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

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

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

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

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 γ 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. The Role of Capital in Explaining Sector 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 1 reports our findings for the role of capital per worker differences in accounting for sector productivity differences. Here we employ Caselli (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).

Table 1: 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).

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 1, 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. Estimation of the Non-Transitory Component of Wages

In this section we discuss how we estimate the variance of the non-transitory component of wages by sector to which we calibrate the model. The rationale for calibrating the model to match variation in the non-transitory component of wages, rather than all wage variation, is that 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. This distinction is important because transitory effects may be relatively more prevalent in agriculture, for example, as a result of weather shocks.

3.1. CPS Data

To estimate the variance of the non-transitory component of wages, we make use of micro-level data from the March Current Population Survey (CPS). We use data from 1996 to 2010, which are the most recent years available which allow for consistent matching of workers across years. We calculate each individual's wage as total labor income in the previous year divided by hours worked in the previous year. We define total labor income as the sum of salary income plus

Table 2: Summary Statistics of CPS Data: 2003-2010

Statistic	Value
Percent of Workers in Agriculture	1.55
Ratio of Average Wage in Agriculture / Non-agriculture	0.701
Variance of Log Wages, Agriculture	0.355
Variance of Log Wages, Non-Agriculture	0.380

0.66 of business income plus 0.46 of farm income, where the fractions of business and farm income assigned to labor are those estimated for the U.S. non-agricultural and agricultural sectors found by Valentinyi and Herrendorf (2008). We exclude all individuals who have missing hours or income data or whose wage is lower than the Federal minimum wage. We express all wages in 2010 dollars using the U.S. Consumer Price Index.

We make use of the short panel dimension of the CPS, which allows a subset of individuals to be matched in two consecutive years. We follow exactly the criteria of Madrian and Lefgren (2000) in eliminating any potentially spurious matches. We end up with 202,677 individuals total that can be matched in two consecutive years. We define agricultural workers to be those whose primary industry of employment in both years is agriculture, forestry or fishing. We define non-agricultural workers to be those in any other sector in both years.

Table 2 presents some summary statistics of the data. Agricultural workers constitute 1.55% of all workers, which is in line with estimates of agriculture's share in employment from other sources, e.g. Herrendorf and Schoellman (2011). The average hourly wage in agriculture is 0.701 times as high as in non-agriculture. The variances of log wages are 0.355 in agriculture, and slightly higher at 0.380 in non-agriculture. These values are consistent with those reported in Heathcote, Perri, and Violante (2010) from the CPS in their study of cross-sectional inequality in the United States using various micro-level data sources.

3.2. Specification and Estimation of Non-Transitory Components

To estimate the fraction of wage variance arising from the non-transitory component of wages, we assume that log wages for an individual in sector j at time t are given by

$$\log(w_{j,t}) = \log(z_j) + \epsilon_{j,t}$$

where z_j is the non-transitory component of wages, and $\epsilon_{j,t}$ is a transitory shock that is serially uncorrelated, independent of z_j, and distributed with mean zero and variance σ^2_ϵ. Given this specification, the variance of log wages can then be written as

$$Var[\log(w_{j,t})] = \sigma^2_{j,z} + \sigma^2_{j,\epsilon}, \tag{1}$$

where $\sigma^2_{j,z}$ captures the variance of the non-transitory component of wages in sector j. To obtain estimates of the two $\sigma^2_{j,\epsilon}$, we note that:

$$Cov[\log(w_{j,t}), \log(w_{j,t+1})] = E[(\log(w_{j,t}) - \mu_j)(\log(w_{j,t+1}) - \mu_j)] = \sigma^2_{j,z}, \tag{2}$$

where $\mu_j = E[\log(w_{j,t})]$. Thus, the covariance of log wages in periods t and $t+1$ is exactly equal to the variance of the non-transitory component. To estimate the $\sigma^2_{j,z}$ we subtract our estimates of $\sigma^2_{j,\epsilon}$ from the total log wage variance in each sector.[1]

We end up with estimates of the non-transitory component of wages of $\sigma^2_{a,z}$ = 0.144 and $\sigma^2_{n,z}$ = 0.224. These are the values for which we target in our calibration along with the ratio of average wages in agriculture to average wages in non-agriculture reported in Table 2.

There are two intuitive features of these results. First, while total variance of log wages is similar across sectors, after correcting for transitory and non-transitory components we find that there is more variance in non-transitory wages in non-agriculture than agriculture. Given the mapping from non-transitory wages to individual productivity in the model, this has the implication that there is more variation in individual productivity in non-agricultural work than in agricultural work, which seems reasonable given that non-agricultural work encompasses more types of economic activities. Second, this implies that estimates of the transitory component of wages are larger in agriculture relative to non-agriculture ($\sigma^2_{a,\epsilon}$= 0.106 relative to $\sigma^2_{n,\epsilon}$= 0.077). This is what one might be expect given the importance of transitory weather shocks in agricultural production.

D. Other Data Sources

The other data sources employed in the paper are as follows:

- **GDP Per Worker** – From the Penn World Table version 6.2., variable "rgdpch".

[1] An alternative approach to estimating the two $\sigma^2_{j,\epsilon}$ terms is to run, in each sector, a regression of log wages on a complete set of individual fixed effects, and then compute the variances of the residuals from the regressions. Using this approach we find a similar, and somewhat larger, quantitative importance of the paper's selection mechanism.

- **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. 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/).

E. Quantitative Results for Rich vs. Intermediate Income Countries

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 agricultural workers are much less pronounced then they are between rich and poor countries.

Table 3 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 agricultural productivity as non-agricultural 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? As in the first alternative experiment, 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, agricultural workers are highly selected based on agricultural productivity in both countries, and hence average worker productivity is only slightly lower in the

Table 3: 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	4.2	3.1	3.0	1.4
Without Selection	3.1	3.1	3.1	1.0

Note: The aggregate productivity differences are the ratios of GDP per worker between the 90th and 50th percentile countries. Sector productivity differences are the ratios of sector output per worker in the 90th and 50th percentile countries. The Ag/Non-Ag Ratios are the agricultural productivity differences divided by the non-agricultural productivity differences.

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.

F. Open-Economy Considerations

The benchmark model treats each economy as closed. This raises an important question: how would 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 (7): 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 (2011) 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-agricultural 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 sector productivity patterns.

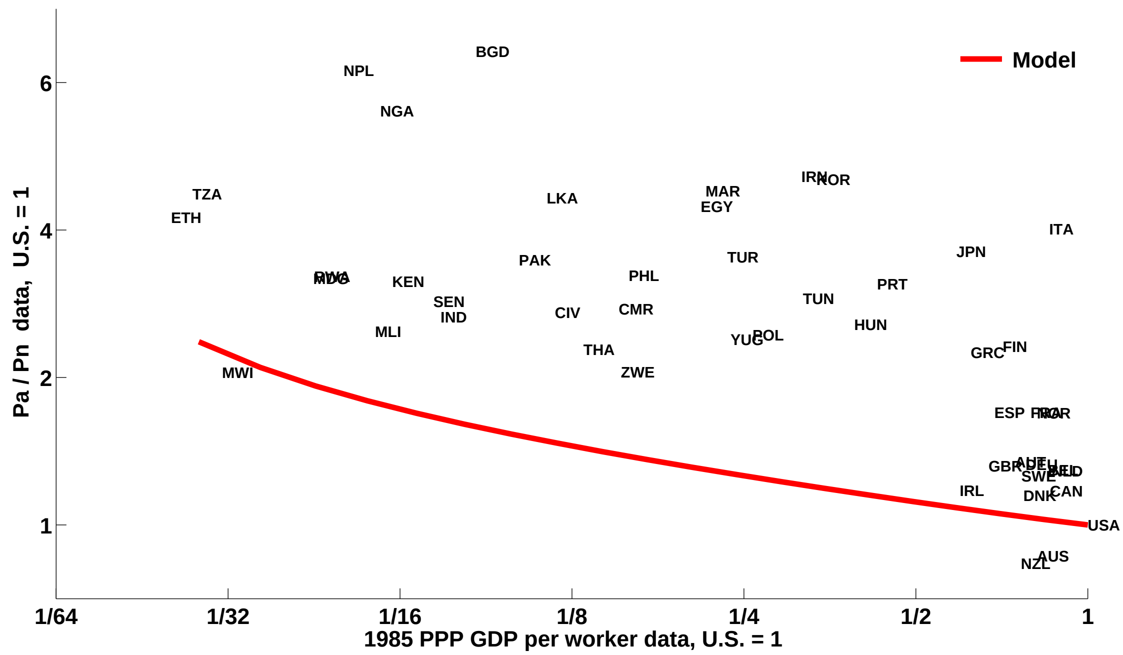

Figure 1: Relative Agriculture Producer Prices, Data and Model

G. Agricultural Producer Price Data

We show that the model's prediction that the relative price of agricultural goods is higher in poor countries is also consistent with data on producer prices. While in principle producer prices are more directly comparable to the prices in our model, since producer prices do not include a distribution margin, in practice producer prices for agricultural and non-agricultural goods are available for a much smaller set of countries. Nevertheless, we find that relative producer prices of agricultural goods behave very similarly to relative consumer prices of agricultural goods.

Our data source is the 1985 FAO food producer price data, explored in detail by Adamopoulos (2009), and used by Caselli (2005) and Restuccia, Yang, and Zhu (2008) to construct sector productivity measures. For the prices of non-agricultural goods we use the consumer price data for the corresponding countries available in the 1985 Penn World Tables. We end up with 60 countries with reasonably broad variance in per capita income.

Our results using producer prices of agriculture are in Figure 1. In the figure, one can see that relative prices of agricultural goods are still higher in poor countries than rich countries, with the 10th percentile of countries around 4 times as high as in the United States (again normalized to one in the figure.) Note that relative agricultural prices appear a bit higher in poor countries once producer prices are used. This is consistent with the finding of Adamopoulos (2009) that distribution margins for food are moderately higher in richer countries than poor countries.

References

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

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

CASELLI, F. (2005): "Accounting for Cross-Country Income Differences," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf., 679-741. Elsevier.

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

GOLLIN, D., D. LAGAKOS, AND M. E. WAUGH (2011): "Selection, Agriculture, and the Gains from Trade," Unpublished Manuscript, New York University.

HEATHCOTE, J., F. PERRI, AND G. VIOLANTE (2010): "Unequal we Stand: An Empirical Analysis of Economic Inequality in the UnitedStates, 1967-2006," *Review of Economic Dynamics*, 13(1), 15–51.

HERRENDORF, B., AND T. SCHOELLMAN (2011): "Why is Measured Productivity so Low in Agriculture?," Unpublished Manuscript, Arizona State University.

MADRIAN, B. C., AND L. J. LEFGREN (2000): "An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents," *Journal of Economic and Social Measurement*, 26(1), 31–62.

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.

VALENTINYI, A., AND B. HERRENDORF (2008): "Measuring Factor Income Shares at the Sectoral Level," *Review of Economic Dynamics*, 11, 820–835.