Risk, Selection, and Productivity Differences

Wenbiao Cai

University of Winnipeg

Preliminary and Incomplete

1 Introduction

Two puzzles about international productivity differences are (1) across countries, the difference in PPP output per worker is much larger in agriculture than in non-agriculture. Between the richest 10 percent of countries and the poorest 10 percent of countries, the gap in PPP output per worker is a factor of 55 in agriculture, compared to only a factor of 4.6 in non-agriculture (Caselli, 2005); (2) within a country, the ratio of output per worker at domestic prices in non-agriculture to agriculture - the agricultural productivity gap - decreases systematically with income (Gollin et al., 2013).

These puzzles have attracted a large body of research. Papers that try to understand cross-country difference in real output per worker in agriculture include Restuccia et al. (2008), Donovan (2013), Lagakos and Waugh (2013) and Cai and Ravikumar (2014); Others focus on understanding why agriculture displays lower productivity relative to non-agriculture, e.g., Gollin et al. (2004), Gollin et al. (2013), Herrendorf and Schoellman (2014) and Cai and Pandey (2013). With the exception of Gollin et al. (2004), the existing literature appears to treat the two puzzles separately.

This paper develops a model that is capable of explaining these two empirical facts that existing research have difficulty reconciling. The model augments the classical Roy (1951)

Draft prepared for the AEA 2015 meeting. Preliminary and incomplete. Please do not circulate. Comments welcome. Email: we.cai@uwinnipeg.ca.

model of occupational choice with employment risks that differ across occupations. Specifically, I consider a setting where risk-averse individuals are endowed with different innate productivity in agricultural and non-agricultural production, and decide to work in agriculture or non-agriculture. In addition, I introduce employment risks in non-agriculture as in Harris and Todaro (1970). The employment risks take the form of idiosyncratic shocks to income of individuals working in non-agriculture. In contrast, individuals working in agriculture face deterministic payoffs that depend on their skill and prices.

I first analyze equilibrium division of labor between agriculture and non-agriculture under complete markets and incomplete markets. Under complete markets, there are state contingent contracts that insure away risks associated with idiosyncratic income shocks. As a result, individuals choose a sector with *higher expected income*. Absent complete markets, such insurance is unavailable. Hence, individuals choose a sector the provides *higher expected utility*. I show two key results. First, individuals working in non-agriculture are more specialized in non-agricultural production when market is complete. The reason is simple. Since individuals are risk-averse, stronger comparative advantage is required for individuals to self-select in to non-agriculture where there are income shocks that can not be insured against.

Second, I show that the gap in wage between agriculture and non-agriculture is decreasing with income. Because of the risk associated with employment in non-agriculture, equilibrium wage is higher in non-agriculture than in agriculture to compensate risk bearing. Such risk premium is higher when income is low. This is driven by the fact that preferences are non-homethetic and there is a minimum consumption requirement. When income is low, consumption is closer to the minimum level. In this case, a negative shock to income induces a larger loss in terms of utility. When income is high, a negative shock reduces utility by less because consumption is sufficiently far away from the minimum consumption.

To quantitatively analyze the model, I first calibrate the model to U.S. data. In particular, the model is parameterized such that it reproduces three moments in the data: the share of labor in agriculture, the distribution of wage in agriculture, and the distribution of wage in non-agriculture. The wage distribution for each sector is constructed using data from the March Supplements to Current Population Survey.

I use the model to study the productivity difference between rich and poor countries. As a first pass, I restrict the analysis to a representative poor country and a representative rich country. The former is the poorest 10 percent of countries in the world income distribution, and the latter is the richest 10 percent. All countries are assumed to have the same (unconditional) distribution of skills and also the same distribution of income shocks. The only difference between rich and poor country is TFP. The gap in real GDP per worker between the rich and the poor countries is 26-fold. I pick TFP for rich and poor country such that the model generates the same difference in real GDP per worker between them. Then I assess the model's implications about the employment, output per worker, and price against the data counterpart.

As in the data, the model generates a gap in real output per worker that is larger in agriculture (49-fold) than in the aggregate and in non-agriculture (12-fold). Also consistent with data, the model generates a much larger share of labor in agriculture (64 percent) for the poor country. Relative price of agriculture is more than twice higher in the poor country than in the rich country, an implication that conforms with data on international prices.

Output per worker measured at domestic price is lower in agriculture than non-agriculture, and more so in poor countries than in rich countries. Put differently, the agricultural productivity gap - the ratio of nominal output per worker in non-agriculture to that in agriculture - is a factor 3 for the poor country, and only a factor 2 for rich country. A relative implication is that average nominal wage is lower in agriculture than in non-agriculture, and more so in poor countries.

The results show that employment risks present in non-agriculture can have material impact on self-selection and sectoral productivity. The model with employment risk generates simultaneously (1) across countries much larger difference in real output per worker in agriculture than in non-agriculture and (2) in poor countries, lower productivity in agriculture than non-agriculture. That is, the model is capable of explaining the two empirical facts presented in the first paragraph that existing research have difficulty reconciling.

2 Model

The economy is populated by a continuum of individuals of measure one. Each individual is endowed with skill $X = (x_a, x_m)$, where x_a determines productivity in agriculture and x_m determines productivity in non-agriculture. These skills are random draws from a known distribution $G(X)$. Preferences are given by

$$U(c_a, c_m) = \eta log(c_a - \bar{a}) + (1 - \eta) log(c_m),$$

where c_a and c_m is consumption good produced in agriculture and non-agriculture. The parameter $\bar{a}$ represents the minimum consumption requirement of agricultural goods. If $\bar{a} > 0$, preferences are non-homothetic.

Output in agriculture (Y_a) and non-agriculture (Y_m) are produced by representative firms using the following

$$Y_a = A \int_{j \in \Omega_a} x_a^j dG(X),$$
$$Y_m = A \int_{j \in \Omega_m} x_m^j dG(X),$$

where Ω_a and Ω_m denote the set of individuals who work in agriculture and non-agriculture, respectively. A is total factor productivity (TFP). The production technology implicitly assumes that (1) each type of skill is useful only in one type of production; (2) skills of individuals working in the same sector are perfect substitutes.

There are perfectly competitive labor markets. If an individual with skill (x_a, x_m) chooses to work in agriculture, she earns wage $w_a = pAx_a$, where p is the price of agricultural good relative to non-agricultural good. If the individual chooses to work in non-agriculture, she faces idiosyncratic income shocks. Specifically, her wage is $w_m = Ax_m + \epsilon$, where ϵ represents the income shock that is distributed $H(\epsilon)$. Assume that $\int \epsilon dH(\epsilon) = 0$.

Optimization Conditional on observed skill endowment, individuals decide which sector to work in. Let $V_a(X)$ denote the value function of an individual with ability $X =$

(x_a, x_m) who chooses to work in agriculture, and $V_m(X)$ the value function associated with working in non-agriculture. We have

$$V_a(X) = \max_{\{c_a, c_m\}} U(c_a, c_m)$$
$$s.t. : pc_a + c_m = pAx_a$$

Similarly, for an individual working in non-agriculture, we have

$$V_m(X) = \max_{\{c_a, c_m\}} \int U(c_a, c_m) dH(\epsilon)$$
$$s.t. : pc_a + c_m = pAx_m + \epsilon$$

Finally, the individual chooses the sector that yields a higher value, i.e.,

$$\max_{\{\Pi\}} V_a(X)\Pi + V_m(X)(1 - \Pi)$$

where Π is the indicator function.

Complete Market Consider first the case when there is either no income shock or there is a complete set of contingent claims to insure away the idiosyncratic shocks. In such a case, the division of labor between sectors take a simple form. That is, individuals choose a sector that provides the highest income. And an individual with skill $X = (x_a, x_m)$ choose agriculture if and only if

$$pAx_a > Ax_m.$$

Incomplete Market When markets are incomplete, a sector that yields a higher expected income might not generate higher utility because individuals are risk-averse. Consider an individual working in agriculture with agricultural skill x_a. Her deterministic income is pAx_a, and the associated indirect utility function is

$$V_a(X) = log(pAx_a) + \eta log(\eta) + (1-\eta)log(1-\eta) - log(p).$$

Now consider an individual with non-agricultural skill x_m who works in non-agriculture. The associated indirect utility function is

$$\begin{aligned} V_m(X) &= \int [log(Ax_m + \epsilon) + \eta log(\eta) + (1-\eta)log(1-\eta) - log(p)]dH(\epsilon) \\ &= \int log(Ax_m + \epsilon)dH(\epsilon) + \eta log(\eta) + (1-\eta)log(1-\eta) - log(p). \end{aligned}$$

Holing price p fixed, consider an individual that is indifferent between working in agriculture and non-agriculture in the complete market case. For such an individual, it must be satisfied that $px_a = x_m$. It is easy to see that such an individual would strictly prefer agriculture over non-agriculture when markets are incomplete, i.e.,

$$\begin{aligned} V_m(X) &= \int log(Ax_m + \epsilon)dH(\epsilon) + \eta log(\eta) + (1-\eta)log(1-\eta) - log(p) \\ &< log(Ax_m + \int \epsilon dH(\epsilon)) + \eta log(\eta) + (1-\eta)log(1-\eta) - log(p) \\ &= log(Ax_m) + \eta log(\eta) + (1-\eta)log(1-\eta) - log(p) \\ &= log(pAx_a) + \eta log(\eta) + (1-\eta)log(1-\eta) - log(p) \\ &= V_a(X), \end{aligned}$$

where the first inequality follows from the Jensen's Inequality.

The following proposition establishes results related to labor allocation and productivity in each sector in two different economies: one with complete market and one with incomplete market. To simply the analysis, I assume that preferences are homothetic. The ability distribution $G(X)$ has independent marginal distribution $g(x_a)$ and $g(x_m)$, each is distributed according to a Fréchet distribution with a common shape parameter θ and a common location parameter 0. And the income shocks take the form as a random fraction of income.

Proposition 1. *Relative to an economy with complete market, an economy with incomplete market has the following properties:*
(i) The share of labor in agriculture is higher.
(ii) Average ability of individuals working in agriculture is lower, but average ability of individuals working in non-agriculture is higher.
(iii) Nominal output per worker in agriculture relative to that in non-agriculture is lower.

Proof: See Appendix.

3 Quantitative Analysis

3.1 Calibration

I calibrate the model to U.S. data. Aggregate TFP is normalized to 1. The marginal distribution of skill takes the form of a Fréchet distribution with scale parameter λ_a and λ_m, and a common location parameter that is zero. The distribution of income shocks follow a normal distribution with mean μ and variance σ^2. I approximate the continuous distribution over the support $(\underline{\epsilon}, \bar{\epsilon})$. For the U.S., set $\mu = 0$ and $\sigma = 1$. Hence, the expected income shock is zero. Set $\epsilon = \bar{a}$ such that the maximum negative shock to income is the same as the minimum consumption. Set consumption weight $\eta = 0.0046$ to be consistent with the values used in Restuccia et al. (2008) and Lagakos and Waugh (2013).

There remains three parameters whose values need to be chosen: two distributional parameters (λ_a, λ_m), the subsistence consumption parameter $(\bar{a})$. They are chosen such that the model replicates the share of labor in agriculture, the dispersion of log wage in agriculture, and the dispersion of log wage in non-agriculture. Figure 1 plots the distribution of wage in agriculture, both in the model and in the data. Figure 2 does the same for non-agriculture.

3.2 Productivity Difference

Now I use the model to think about sectoral productivity differences across countries. For the time being, I focus on the difference between two representative groups: rich and poor. The former is the 10 percent richest countries in the world income distribution, and the latter is the poorest 10 percent. In the data, the gap in real GDP per worker between rich countries and poor countries is 26-fold. The gap in real output per worker in agriculture is 55-fold, while that in non-agriculture is 4.6-fold.

Suppose all countries have the same unconditional distribution of skills. For simplicity, also suppose that the distribution of income shocks is the same across countries. The latter might seem rather restrictive as one might expect these random shocks differ both in magnitude and frequency across countries with different levels of development. Nevertheless, it is a useful benchmark. That leaves the level of TFP as only exogenous variable to the model. I pick the level of TFP such that the models produces real GDP per worker that is 1/26 that of the U.S. Then I compare the model's other implications against those in the data.

Table **??** summarizes the results. By construction, the model generates a 26-fold difference in real GDP per worker. As implication, the model generates a 49-fold difference in real output per worker in agriculture, and a 12-fold difference in real output per worker in non-agriculture. Hence, the model captures quantitatively the fact that productivity difference is larger in agriculture than in the aggregate and that in non-agriculture. The poor country also allocates a much larger share of labor to agriculture than in the rich country: 62 percent in the poor country vs. 2 percent in the rich country.

The asymmetric difference in sectoral productivity difference between rich and poor countries stem from the difference in skills. This in fact is the key insight provided in Lagakos and Waugh (2013). This can be easily seen by comparing the conditional mean of agricultural skills for those working in agriculture, and that of non-agricultural skills for those working in non-agriculture. The average agricultural skill is 2.98 for the rich countries, compared to 1.31 for the poor country. In contrast, the average non-agricultural skill is twice higher in the poor country than in the rich country. As a result, the model generates a productivity gap in agriculture that is larger than that in non-agriculture.

	Rich-Poor Ratio	
	Data	Model
-Real output per worker		
Aggregate	26	26
Agriculture	55	49
Non-agriculture	4.6	12
	Data Rich (Poor)	Model Rich (Poor)
-Nominal output per worker		
Non-agriculture/Agriculture	2 (5)	2 (3.3)
-Share of labor		
Agriculture	0.02 (0.83)	0.02 (0.62)

Another prediction from the model is that the ratio of nominal output per worker in non-agriculture to that in agriculture - the agricultural productivity gap - is also higher in poor countries. In the rich country, the agricultural productivity gap is 2. In the poor country, the agricultural productivity gap is 3. This result reflects that the fact that when income is low, the utility loss from income shocks is larger when the minimum consumption constraint is more stringent. As a result, two implications follow. Holding price and shocks fixed, an individual requires stronger comparative advantage in non-agriculture to self-select into that sector. Second, holding skills fixed, an individual commands a higher relative wage in order to compensate the risks associated with employment in non-agriculture.

3.3 The Importance of Risk

What is the importance of employment risks - in this case, shocks to income of those working in non-agriculture - in understanding sectoral productivity differences? A mechanical way to answer this question is to generate prediction in a risk-free economy and compare allocations to those in an economy with income shocks. For this purpose, set $\epsilon = 0$. The unconditional skill distribution is the same as before. I feed into the difference in TFP as before, and compare the model allocations. The results are summarized in Table **??**.

	Rich-Poor Ratio	
	With Risk	No Risk
-Real output per worker		
Aggregate	26	25.5
Agriculture	49	46
Non-agriculture	12	13
	With Risk Rich (Poor)	No Risk Rich (Poor)
-Nominal output per worker		
Non-agriculture/Agriculture	2 (3)	2 (1.8)
-Share of labor		
Agriculture	0.02 (0.62)	0.02 (0.57)

Table **??** illustrates two important implications associated with employment risk in non-agriculture. First of all, given the same difference in aggregate TFP, the model with risk generates a larger productivity difference in agriculture, and a small difference in non-agriculture. Hence, the model with risk does a better job explaining the sectoral productivity differences observed in the data. The reason is that in the presence of risk, individuals need a stronger comparative advantage in non-agricultural production to self-select into non-agriculture.

Second, the model with risk generates a larger agricultural productivity gap - the ratio of nominal output per worker in non-agriculture to agriculture. In fact, the model without risk generates a counterfactual prediction, i.e., the agricultural productivity gap is smaller in the poor country than in the rich country.

4 Conclusion

[TO BE WRITTEN]

References

Cai, W. and M. Pandey (2013). The Agricultural Productivity Gap in Europe. Departmental Working Papers 2013-05, The University of Winnipeg, Department of Economics.

Cai, W. and B. Ravikumar (2014). Skill Accumulation and Sectoral Productivity Differences Across Countries. *Working Paper, University of Winnipeg*.

Caselli, F. (2005). Accounting For Cross Country Income Differences. In P. Aghion and S. D. Durlauf (Eds.), *Handbook of Economic Growth*, pp. 679–741. ELSEVIER.

Donovan, K. (2013). Agricultural Risk, Intermediate Inputs, and Cross-Country Productivity Differences. Working paper, University of Notre Dame.

Gollin, D., D. Lagakos, and M. E. Waugh (2013). The Agricultural Productivity Gap. *Quarterly Journal of Economics. Forthcoming*.

Gollin, D., S. L. Parente, and R. Rogerson (2004). Farm work, Home Work and International Productivity Differences. *Review of Ecconomic Dynamics 7*(4), 827–850.

Harris, J. R. and M. P. Todaro (1970). Migration, unemployment and development: a two-sector analysis. *The American Economic Review 60*(1), 126–142.

Herrendorf, B. and T. Schoellman (2014). Why is Measured Productivity so Low in the Agriculture. *Forthcoming. Review of Economic Dynamics*.

Lagakos, D. and M. E. Waugh (2013, April). Selection, Agriculture, and Cross-Country Productivity Differences. *American Economic Review 103*(2), 948–80.

Restuccia, D., D. T. Yang, and X. Zhu (2008). Agriculture and Aggregate Productivity: A Quantitative Cross-country Analysis. *Journal of Moneytary Economics 55*(2), 234–250.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers 3*(2), 135–146.

A Appendix

A.1 Proof of proposition 1

Proof. Since preferences are homothetic, the result is independent of TFP. Hence, set $A = 1$. Suppose the income shock is τ, then an individual working in non-agriculture with skill x_m has realized income $x_m(1-\tau)$. I make the following assumptions about the distribution of income shocks and the distribution of abilities.

- The distribution of shocks is such that $\Delta < 1$, where $\Delta = exp\left(\int_{-\infty}^{1} log(1-\tau)dH(\tau)\right)$
- $\theta > 1$.

Complete Market In an complete market, an individual chooses agriculture if and only if $px_a > x_m$. As a result, the share of labor in agriculture is $n_a = Prob(px_a > x_m) = \int_0^\infty \int_0^{px_a} dg(x_m)dg(x_a)$. It is shown in Lagakos and Waugh (2013) that

$$n_a = \frac{1}{p^{-\theta}+1}.$$

Correspondingly, the share of labor in non-agriculture is $n_m = 1 - n_a = \frac{p^{-\theta}}{p^{-\theta}+1}$. The average ability of individuals working in agriculture and non-agriculture is given by

$$E(x_a|px_a > x_m) = (1+p^{-\theta})^{\frac{1}{\theta}}\gamma,$$
$$E(x_m|px_a < x_m) = (1+p^{\theta})^{\frac{1}{\theta}}\gamma,$$

where γ is the Gamma function evaluated at $\frac{\theta-1}{\theta}$. Hence, total output in agriculture and non-agriculture is given by

$$y_a = An_aE(x_a|px_a > x_m) = A(1+p^-\theta)^{(\frac{1}{\theta}-1)}\gamma,$$
$$y_m = An_mE(x_m|px_a < x_m) = A(1+p^{\theta})^{(\frac{1}{\theta}-1)}\gamma.$$

Imposing the market clearing condition $c_a = y_a, c_m = y_m$ yields the following expression for relative price

$$(1-\eta)pA(1+p^-\theta)^{(\frac{1}{\theta}-1)}\gamma = \eta A(1+p^{\theta})^{(\frac{1}{\theta}-1)}\gamma,$$

which yields $p = \left(\frac{\eta}{1-\eta}\right)^{\frac{1}{\theta}}$.

Incomplete Market Denote the price of agricultural output by $\tilde{p}$. In an incomplete market, an individual chooses agriculture if and only if $V_a(x_a) > V_m(x_m)$, i.e.,

$$\begin{aligned} log(\tilde{p}x_a) > \int_{-\infty}^{1} & log(x_m(1-\tau))dH(\tau), \\ &= log(x_m) + \int_{-\infty}^{1} log(1-\tau)dH(\tau), \\ &= log(x_m) + log(\Delta), \\ &= log(x_m\Delta), \end{aligned}$$

where $\Delta = exp\left(\int_{-\infty}^{1} log(1-\tau)dH(\tau)\right)$. Using the optimal consumption allocation and market clearing condition, the price of agricultural output in the incomplete market is given by the following equation

$$(1-\eta)\tilde{p}A(1+(p/\Delta)^-\theta)^{(\frac{1}{\theta}-1)}\gamma = \eta A(1+(p/\Delta)^{\theta})^{(\frac{1}{\theta}-1)}\gamma,$$

which yields $\tilde{p} = \left(\frac{\eta}{1-\eta}\right)^{\frac{1}{\theta}}\Delta^{\frac{\theta-1}{\theta}}$. The share of labor in agriculture ($\tilde{n_a}$) and the average ability of workers are given by

$$\begin{aligned} \tilde{n}_a &= \frac{1}{(\tilde{p}/\Delta)^{-\theta}+1}, \\ \tilde{E}(x_a|V_a > V_m) &= (1+(\tilde{p}/\Delta)^{-\theta})^{\frac{1}{\theta}}\gamma, \\ \tilde{E}(x_m|V_a < V_m) &= (1+(\tilde{p}/\Delta)^{\theta})^{\frac{1}{\theta}}\gamma. \end{aligned}$$

(i) $\tilde{p}/\Delta = \left(\frac{\eta}{1-\eta}\right)^{\frac{1}{\theta}}\Delta^{\frac{-1}{\theta}} > \left(\frac{\eta}{1-\eta}\right)^{\frac{1}{\theta}} = p$, hence it follows immediately that $\tilde{n}_a > n_a$.
(ii) It also follows immediately from the fact that $\tilde{p}/\Delta > p$.

(iii) The ratio of nominal output per worker between agriculture and non-agriculture is given by

$$\frac{py_a/n_a}{y_m/n_m} = \frac{pE(x_a|px_a > x_m)}{E(x_m|px_a < x_m)}.$$

In the complete market, the productivity ratio is

$$\frac{p(1+p^{-\theta})^{\frac{1}{\theta}}\gamma}{(1+p^{\theta})^{\frac{1}{\theta}}\gamma} = \frac{p}{(p^{\theta})^{\frac{1}{\theta}}} = 1.$$

In the incomplete market, the productivity ratio is

$$\frac{\tilde{p}(1+(\tilde{p}/\Delta)^{-\theta})^{\frac{1}{\theta}}\gamma}{(1+(\tilde{p}/\Delta)^{\theta})^{\frac{1}{\theta}}\gamma} = \frac{\tilde{p}}{(\tilde{p}/\Delta)} = \Delta < 1.$$

$\square$

A.2 Sectoral Wage

Wage data is mainly from March supplements to the Current Population Survey. To increase sample size, I pool together samples between from 2000 to 2006. The samples are restricted to include individuals between age 18 and 65 with positive wage and salary and hours. Additional information includes the industry, gender, age, years of schooling, and class of employment (self-employed or employees). I calculate wage dividing wage and salary income by hours of work.

Figure 1: Wage Distribution in Agriculture: U.S.

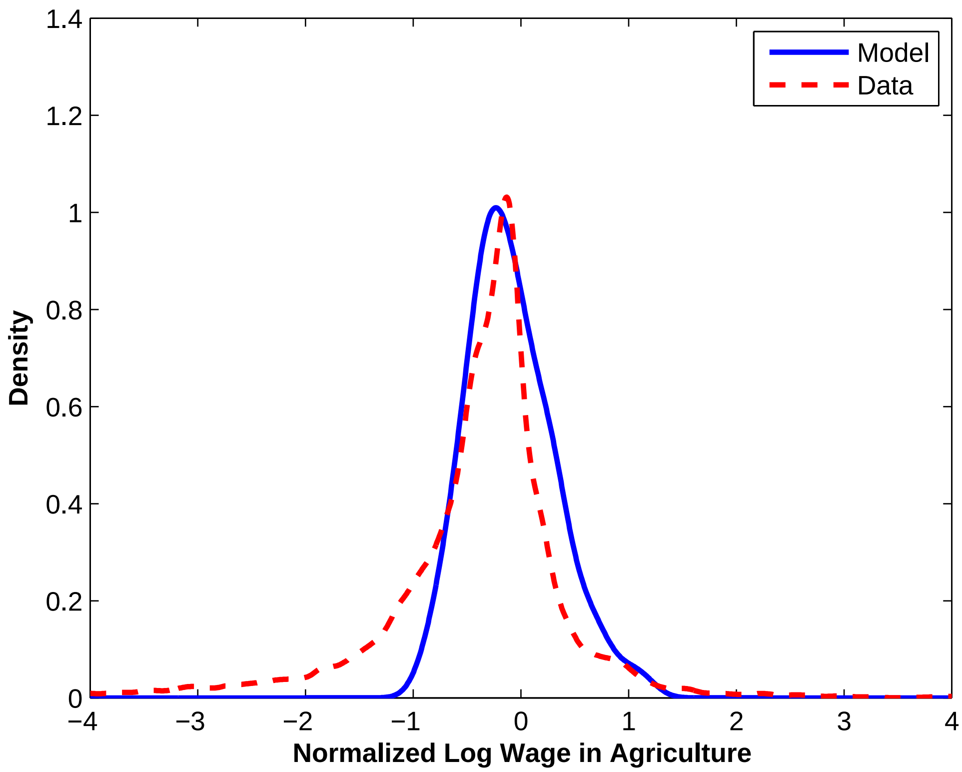

Figure 2: Wage Distribution in Nonagriculture: U.S.