Why Is Agricultural Labor Productivity so Low in the United States?*

Berthold Herrendorf† **and Todd Schoellman**‡

February, 2011

Abstract

A big question in development economics is why developing countries are so unproductive in agriculture. This question is hard to answer because of limited data. In this project, we explore what we can learn from agriculture in US states where we have rich data. Focusing on US states has the advantage that there are no major barriers or institutional differences. We provide evidence that there are large labor productivity gaps between nonagriculture and nonagriculture; in some US states the difference between the two is larger than a factor of five. We show that these gaps are not likely to be the result of imperfect measurement. We also show that sectoral differences in human capital and capital shares account for most of these labor productivity gaps.

*We thank David Lagakos and Akos Valentinyi for helpful discussions.

†Address: Department of Economics, Arizona State University, Tempe, AZ 85287. E-mail: Berthold.Herrendorf@asu.edu

‡Address: Department of Economics, Arizona State University, Tempe, AZ 85287. E-mail: todd.schoellman@gmail.com

1 Introduction

It is well known that developing countries are much more unproductive in agriculture than in nonagriculture (Restuccia, Yang, and Zhu 2008, Caselli 2005). For example, the World Bank's World Development Indicators report that in Thailand during the 1990s the labor productivity gap between nonagriculture and agriculture was at least a factor of ten (measured in local currency). This raises three important questions:

(i) Why are developing countries so unproductive in agriculture?

(ii) Why do people in developing countries not leave agriculture and move to other sectors where labor productivity – and presumably wages – are much higher?

(iii) Why do developing countries not import more agricultural goods?

Although these questions have spurred a stream of interesting recent research, we are far from having convincing answers. One major obstacle to progress is that good, comparable data are difficult to find for many developing countries. To give just one relevant example, we lack data for most developing countries on the education of agricultural and non–agricultural workers. Therefore, we cannot assess quantitatively the fairly obvious hypothesis that labor productivity is higher in the nonagricultural sector because nonagricultural workers are more educated.

In this project, we take a different approach than the literature and study agriculture in the United States. Focusing on the United States has the advantage that we can draw on rich and well-developed data on the agricultural sector. Nonetheless, one might think that there is not much to learn from the United States about agriculture in developing countries. After all, the US is typically considered to be a relatively undistorted economy where workers and production factors are free to move across sectors and state borders.

We provide evidence that suggests otherwise: There are large labor productivity gaps between nonagriculture and nonagriculture at the level of US state (where labor productivity is measured as value added per worker in current dollars). We find that in some states the difference between the two is larger than a factor of five. Further, there are large differences in agricultural labor productivity across states. This shows that labor productivity in agriculture and non–agriculture need not be equalized even when there are no obvious barriers to leaving agriculture, no obvious trade barriers, and no obvious differences in the institutional environment.

We then set out to understand the reasons for the sectoral differences in labor productivity. In particular, we would like to know which roles are played by imperfect measurement and by sectoral differences in human capital and in labor shares. Only after understanding the quantitative importance of these factors can we hope to be able to distinguish between "normal" and "abnormal" sectoral productivity differences in developing countries. We find that imperfect measurement is not likely to be the reason for sectoral differences in labor productivity. In fact, once we use the best measurement available, the labor productivity gaps between nonagriculture and agriculture increase somewhat. We explain why that is the case. We also find that sectoral differences in human capital and in labor share close a considerable part of the labor productivity gaps. We then discuss the lessons from our results for US states for the macro development literature about labor productivity gaps in developing countries.

Our paper joins a growing literature that documents and seeks to explain sectoral productivity differences within a country and to learn which sectors make poor countries unproductive. While some of this literature explores multiple sectors (such as Herrendorf and Valentinyi (forthcoming)), most focus on the apparent gap between agriculture and non-agriculture productivity. Restuccia, Yang, and Zhu (2008) and Caselli (2005) first documented this gap. Our paper is closely linked to Gollin, Lagakos, and Waugh (2011) in that both try to measure the gap as carefully as possible given the existing data. Their paper uses surveys from developing countries, which offers advantages over aggregate data but are less detailed than some of the US data, such as time-use survey data. On the other hand they have the advantage of working directly with developing countries. Our paper is also linked to a line of research that seeks to explain the productivity gaps. Existing theories point to the scale or risk of farming (Adamopoulous and Restuccia 2010, Donovan 2011); to barriers between agriculture and nonagriculture (Restuccia, Yang, and Zhu 2008); to differences in factor endowments (Caselli 2005); to selection in the workers in the two sectors (Lagakos and Waugh 2010); and to home production (Gollin, Parente, and Rogerson 2004). Our paper is again most similar to Gollin, Lagakos, and Waugh (2011) in that we make use of the abundant and detailed to quantify the relative role for these and other potential explanations, in the spirit of an accounting exercise.

The paper proceeds as follows. Section 2 measures sectoral differences in labor productivity using conventional sources that would be available to international researchers. Section 3 shows that differences in labor productivity are actually larger when using the best available data. Section 4 provides our explanations for the differences in the labor productivity in the US. Section 5 concludes.

2 Sectoral labor productivity gaps

In this section, we document that in most US states labor productivity in non–agriculture is considerably higher than in agriculture when measured in current dollars. Since we are interested in the relative differences between the labor productivities in non–agriculture and agriculture, we focus on the ratio of the two. Since this ratio is unit free, we can compare it across different states and dates. If this ratio is not equal to one, then we will say that there is a *labor productivity gap*. If the labor productivity gap is considerably larger than one, then the question arises why there is no reallocation of workers from agriculture to non–agriculture, which would increase GDP and, under plausible assumptions, the wages of the relocating workers. In the context of developing countries, this question has frequently been asked by researchers who found large labor productivity gaps.

The labor productivity gaps that we compute here are not the only relevant statistic that the macro development literature has considered. Many studies have instead focused on labor productivity gaps within agriculture across countries. These gaps are typically computed by using agricultural output in international prices, which can be obtained from the Food and Agriculture Organization (FAO). The typical finding is that the gaps in agriculture are much larger than those at the aggregate level. For now, we are not computing the corresponding statistic for US states. A first reason is that we do not have sufficient information to compute a set of ppp prices of agricultural output that are invariant across states. A second reason is that our main motivation is the question why workers do not move between sectors. For answering this question, the labor productivity gaps we compute are the relevant statistic because they are based on the system of prices that workers actually face when they make their choices.

In order to ensure comparability with the literature on macro development, we start by constructing measures of labor productivity in agriculture and nonagriculture based on data that are typically available for developing countries. As a result, our baseline estimates of the productivity gap do not use the best and most reliable data that are available for US states. This raises the question to which extent these baseline estimates are a result of imperfect measurement. We address this question in the next section where we use the best available data and compare the results with our baseline. We find there that our baseline estimates are a lower bound of the best measures of labor productivity gaps.

We make the following choices regarding the data. We define the agriculture sector in a manner consistent with the FAO, so that it includes crop and animal production. Note that agriculture so defined does not include fishing, forestry, horticulture, hunting, or trapping. We define the nonagriculture sector as all industries outside of the farm sector. For agricultural value added, most studies about developing countries use the numbers from the World Development Indicators (WDI) in current prices, which come from the national accounts. The closest comparable numbers for US states are the value added numbers by industry and state from the BEA's regional accounts.[1] For employment in developing countries, the typical data source is the International Labor Organization (ILO), which reports labor forces based on censuses. The closest comparable numbers for US states are the workers by industry from the census. We obtain these numbers from the public–use version of the decennial population census made available through Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek (2010).[2]

Figures 1 show the raw data about labor productivity for the census years 1970, 1980, 1990, and 2000. Agricultural value added per worker is depicted on the x axis and non–agricultural value added per worker is depicted on the y axis. We can see that in all years and in most states, labor productivity is considerably higher in non–agriculture than in agriculture. Table 1 provides three summary statistics that quantify the resulting gaps in sectoral labor productivities: the median gap is 1.3, the gap at the 90th percentile is 2.2, and the maximum gap is 7.4. These gaps are sizeable, and the maximum gap has the same order of magnitude as the gaps found for some of the poorest countries in the world. This fact was surprising to us because there are no barriers to moving from agriculture to nonagriculture in the US and US states are typically thought to have similar institutional and legal environments.

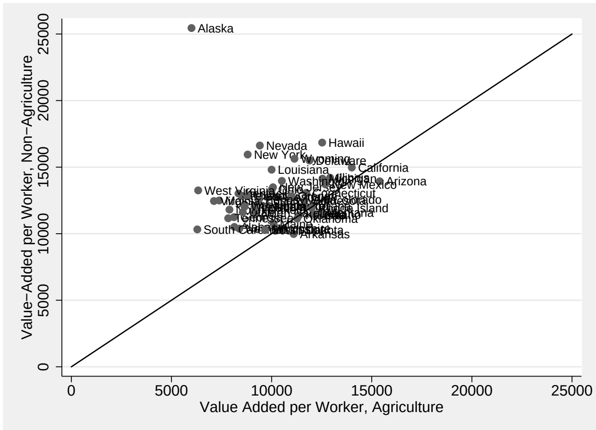
3 Robustness

In the previous section, we calculated baseline estimates of the productivity gap based on data that is typically available for developing countries. As a result, we did not use the best and most reliable data that is available for the US. This raises the question to which extent our baseline estimates of the productivity gap are due to bad measurement.

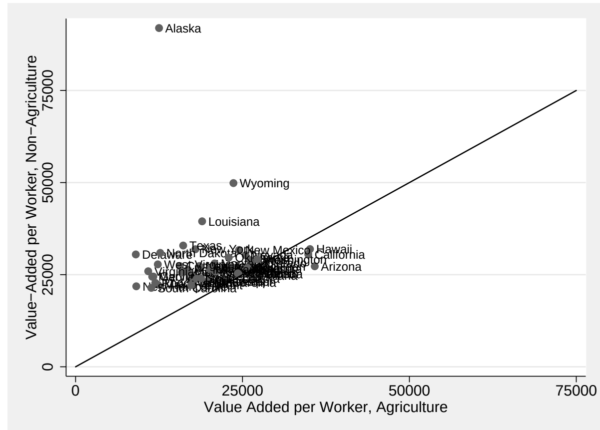
We address this question in this subsection by using the best available data to calculate

[1]One potential difficulty with using the regional accounts is that in 1997 the industry categorization switched from SIC to NAICS. Fortunately, the BEA published two sets of regional accounts in 1997, one using SIC and another one using NAICS. A comparison of the two shows that the series for agricultural value added are closely related. In what follows, we merge the two series, so that all estimates up to 1997 are based on SIC categorizations and all estimates from 1998 onward are based on NAICS categorizations.

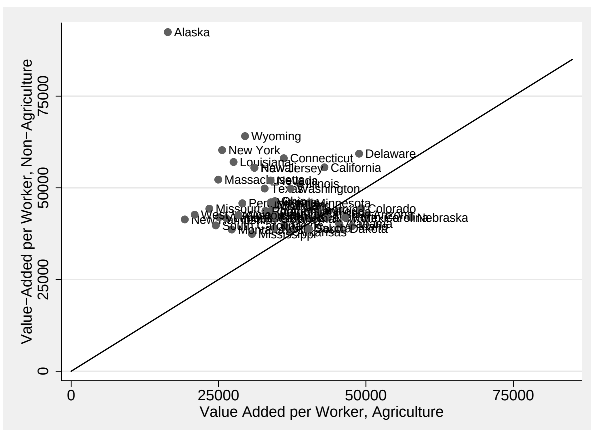
[2]In each year, the census uses different industry coding schemes, which also differ from the BEA. Appendix A.2 gives details and the crosswalk used to match census industries to the agricultural and nonagricultural sectors.



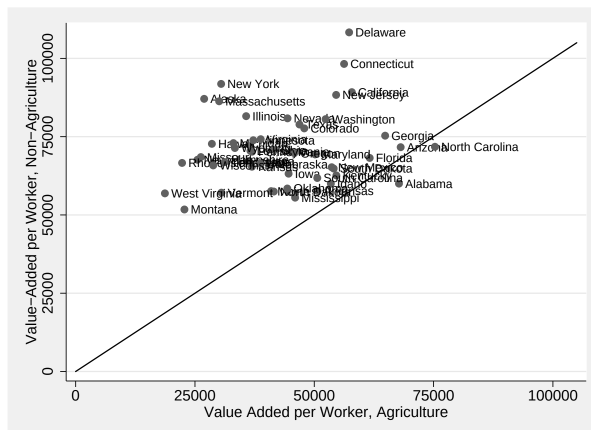
(a) 1970



(b) 1980



(c) 1990



(d) 2000

Figure 1: Raw Data – Sectoral Value Added per Worker

sectoral labor productivity. We first turn to the calculation of the sectoral value added. We discuss the shortcomings of the BEA's measure of value added in agriculture and derive a better measure. We then turn to sectoral employment and calculate hours worked rather than workers. We find that the productivity gaps calculated in the previous subsection increase for many states. We conclude from these findings that the gaps calculated in the previous section, though sizeable, are a lower bound for the actual gaps.

3.1 Sectoral value added

Above we have used the BEA's regional accounts to calculate value added in agriculture. While these accounts are a natural starting point, they are not the last word because the

Table 1: Summary Statistics for Labor Productivity Gaps – Baseline

Data Sources	
Value Added	BEA
Employment	Census
Results	
Median	1.3
90th Percentile	2.2
Maximum	7.4

Table notes: Distributional results are for the sectoral productivity gap of the fifty states for four years: 1970, 1980, 1990, and 2000.

BEA calculates value added in agriculture in a somewhat peculiar way as the value added produced by *farmers* where a farmer is a person that operates a farm.[3] To appreciate the implications of this notion of value added, an example may be helpful. Consider the wages paid to contract labor on farms or the rental payments that farmers pay to land owners who are not farmers. Clearly, both of these payments are factor payments that belong to the value added generated on farms. Nonetheless, since they do not lead to income of farmers, the BEA treats these payments as expenses which are not part of value added in the farm sector.

To see whether this matters, we construct a new measure of value added that includes factor payments generated on farms irrespective of whether they accrue to farmers or to people who are not farmers. To this end, we use the United States Department of Agriculture's state–net–value–added accounts, from which the BEA construct its measure of farm value added. Since these accounts include a detailed breakdown of the receipts, expenses, and factor payments in the farm sector at the state level, we have sufficient information to make the required adjustments. Since these accounts also have information about the net subsidies paid to the farm sector at the state level (i.e., farm subsidy received minus motor vehicle registration fees and property taxes paid), we produce agricultural value added with and without subsidies. The appendix describes in detail what we do.

Columns 3 and 4 of Table 2 report the results when we calculate the labor productivity gaps with this new measure of value added without and with subsidies – while as before employment is taken from the census. Perhaps surprisingly, the labor productivity gaps do not change much. In the most relevant case where agricultural value added includes

[3]Note that the USDA has the same notion of value added in agriculture. Both institutions want to measure the income of farmers.

Table 2: Summary Statistics for Labor Productivity Gaps – Robustness

Data Sources			
Value Added	BEA	USDA, no subs.	USDA, subs.
Employment	Census	Census	Census
Results			
Median	1.3	1.2	1.2
90th Percentile	2.2	2.1	2.1
Maximum	7.4	7.4	7.8

Table notes: Distributional results are for the sectoral productivity gap of the fifty states for four years: 1970, 1980, 1990, and 2000.

subsidies (column 4), the median and the gap at the 90th percentile go down by one percentage point to 1.2 and 2.1 whereas the maximum gap goes up by four percentage points to 7.8. One might wonder why the gaps do not change more. At a mechanical level, this must mean that in states with relatively unproductive agriculture farmers own most of the production factors used in agriculture. One might also wonder why the maximum gap actually goes up. The explanation lies in the fact that the vast majority of US farm subsidies go to states with relatively productive agriculture. In contrast, agriculture in relatively unproductive states pays more in taxes than it receives in subsidies so that the net transfers from the government are negative.

In passing, we mention that the situation is very different in states with relatively productive agriculture. In these states (many of which turn out to be in the Plains) the numbers for the labor productivity gaps change considerably depending on which measure one uses. For example, at the 10th percentile, the labor productivity of nonagriculture relative agriculture comes out at 0.96 for BEA value added, 0.81 for the new measure of value added without subsidies, and 0.75 for the new measure of value added with subsidies. Subsidies make a sizeable difference here because, as we have just mentioned, most of them go to states with relatively productive agriculture.

3.1.1 Sectoral employment

The census has two shortcomings that may be crucial for what we found above. First, it is administered during a particular month, typically March. If employment is seasonal then census employment may not be a good measure of annual employment. Second, the census asks respondents only about their primary job. If secondary jobs are important, then census employment may not be a good measure of actual employment.

A first way of trying to gauge the importance of these shortcomings is to use data from establishment surveys. In the US, we have the current employment survey of establishments (CES), which is administered by the BLS. The CES collects data during the whole year on the number of jobs per establishment (rather than the number of employees), and so it includes second jobs. The CES forms the basis of the BEA's employment–by–industry statistics, which we use at the state level.

Column 2 of Table 3 provides our three summary statistics for the same data sources as in the previous section. Column 3 provides them when we replace census employment by BEA employment based on the CES. To ensure comparability with some results that will come below, these statistics are calculated for the year 2000 only. We can see that the gaps in sectoral labor productivities increase considerably when we move from census employment to BEA employment. This suggests that seasonality and secondary jobs are potentially important for calculating labor productivity gaps. We choose to hedge this statement somewhat because the BEA employment figures also have potential problems. In particular, the BEA counts all jobs equally and does not attempt to adjust for full–time equivalent employment at the state level.[4] As a result, the BEA numbers for jobs at the state level exaggerate actual employment at the state level, and it is not clear in which sector this exaggeration is more pronounced.

The current population survey (CPS) offers a second and more reliable way of gauging the importance of the two shortcomings of the census. The CPS is a rotating panel survey that is administered by the BLS in each month of the year to a sample of households. A key advantage of the CPS is that starting in 1994 it asks one quarter of respondents about the identity of their first two jobs (if applicable) as well as their hours worked for each of the first two jobs. We use that information for the year 2000 to calculate hours worked by sector at the state level. We then use these hours to calculate labor productivity gaps by state.[5] Column 4 of Table 3 reports the results for the year 2000 when employment is calculated from the CPS. We can see that compared to column 2 the sectoral productivity gaps do go up considerably when we use CPS numbers for hours. But in most states they do not quite go up as much as in column 3.

At a mechanical level, the fact that the labor productivity gap is larger with CPS data than with Census data must mean that the CPS counts more agricultural and/or fewer

[4]For further discussion see Chapter IX of the BEA's "Local Area Personal Income and Employment Methodology", April 2010. Note that at the national level, the BEA attempts to adjust for full–time equivalent employment. However, there is some doubt how successful it is at that, see Cociuba, Prescott, and Ueberfeldt (2009) for further discussion.

[5]As with the census we have to construct a crosswalk from CPS industries into our agriculture–nonagriculture sectors. This point and other details are explained in Appendix A.3.

Table 3: Summary Statistics for Labor Productivity Gaps – Robustness for 2000

Data Sources			
Value Added	BEA	BEA	BEA
Employment	Census	BEA	CPS
Results			
Median	1.7	2.5	2.1
90th Percentile	2.8	3.7	3.9
Maximum	3.2	9.7	4.3

Table notes: Distributional results are for the sectoral productivity gap of the fifty states for year 2000.

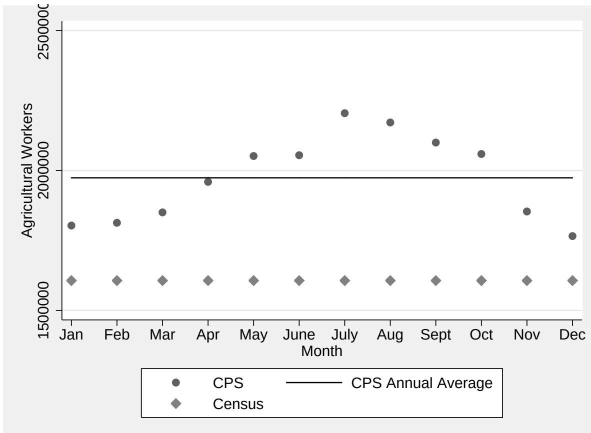
Table 4: Census over CPS Employment

	Total Hours	=	Workers	×	Hours per Worker
Agriculture	0.71		0.81		0.86
Non–Agriculture	0.93		0.97		0.96

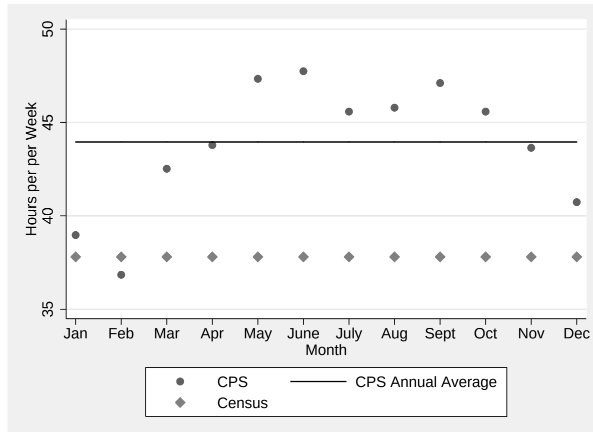
nonagricultural employment. To dig deeper into this, we look at hours in nonagriculture and agriculture at the national level in 2000. Table 4 shows that the census counts 93 percent of total CPS hours worked in the nonagriculture sector, but only 71 percent of CPS hours worked in agriculture. Moreover, the CPS counts both more agricultural workers and more hours per agricultural worker.

These differences between the CPS and the census are directly related to the fact that the census is taken in March and does not take account of secondary jobs. To see how seasonality matters, Figures 2a and 2b plot from the CPS the numbers of workers who report agriculture as their main sector and the number of hours worked per agricultural worker in the whole US economy over the course of the year. Both Figures also include the annual average and the March number from the census for reference. We can see that the number of agricultural workers and the number of hours worked per agricultural worker are highly cyclical, with a large increase in the summer and a large decrease in the winter. March is a month with below–average activity on both of these dimensions. So the fact that the census is taken in March implies that it counts too few agricultural workers and hours worked per agricultural worker.[6]

[6]The census asks respondents both how many hours they worked last week and how many hours they usually work per week. Most papers attempt to correct for seasonality by using usual hours worked, and we follow this practice. However, our explorations in the CPS suggest that this correction does not work



(a) Workers



(b) Hours per Worker

Figure 2: Agricultural Employment by Month

Figures 2 also show that the CPS counts more workers and hours worked even in the month of March when the census is taken. To understand the reason for this, recall that the census collects information only on the main job of each worker, which is defined as the job in which the worker earns most income. The census then attributes all hours worked to that main job. In contrast, the CPS collects information on up to two jobs and defines the main job as the one in which the workers work most hours. It then attributes hours in both jobs to the agricultural or nonagricultural sectors as appropriate.

To understand why the CPS counts more agricultural workers than the census even in the month of March, workers who have two jobs in different sectors are key. The census defines the primary job as the one in which the worker earns more income whereas the CPS defines it as the one in which the worker works more hours. Since, as we have documented above, nonagriculture is more productive than agriculture, the number of workers who earn more income in nonagriculture but work more hours in agriculture is larger than the number of workers who do the opposite. This fact implies that the CPS will count more workers in agriculture than the census. In addition, there are many more nonagricultural workers than agricultural worker who hold a secondary job in the other sector. This fact implies that again the CPS will count more hours worked per agricultural worker than the census.[7]

well for agriculture. Hours worked last week and usual hours worked are highly correlated over the course of the year in the CPS (coefficient = 0.91): in months when respondents work many hours they also report that they usually work many hours.

[7]One way to quantify this effect is to create an imperfect measure of labor input in the CPS that attributes all of a worker's hours to their main job. Doing so closes 72 percent of the gap between census and CPS hours worked per agricultural worker in the month of March, suggesting that the treatment of

Note that agricultural workers are much more likely than nonagricultural workers to have a secondary job in the other sector. However, the sheer number of nonagricultural workers relative to agricultural workers still implies a positive net flow of secondary job hours into agriculture.

Finally, we want to briefly address a possible concern with using the CPS, and employee surveys more generally. The concern is that workers do not correctly recall how many hours they work or the sectors in which they work them. To explore whether this is a problem for us, we compare the American Time Use Survey (ATUS) and the CPS between 2003 and 2009, which is the longest period of overlap between the two data sources. ATUS data are collected from time use diaries kept by participants who record their activities by minute for a 24–hour period. ATUS data are deemed very reliable. We find that, at the national level, ATUS hours worked per agricultural and nonagricultural worker were 5.4% and 4.3% lower, respectively. The fact that the differences between CPS and ATUS hours are small suggests that imperfect recall is not important for our benchmark measure of employment.

We began this paper by measuring sectoral labor productivity gaps using conventional data that would generally be available across countries. We documented that within the US there are large labor productivity gaps between nonagriculture and agriculture for several states. We then addressed the natural next question whether these gaps are the result of imperfect measurement. Our results suggest that they are not: When we carefully measure sectoral value added and and labor inputs, we find *larger* sectoral productivity gaps.

In what follows, we take the measured labor productivity gaps seriously and explore what may explain them. We choose to take the gaps measured conventionally with BEA value added and census employment as our baseline estimate. One might prefer that we used the larger labor productivity gaps calculated with CPS labor as our baseline estimates. We do not do this for two reasons. First, the CPS has asked about second jobs only since 1994 and we do not want to discard the information for the census years 1970, 1980, and 1990. Second, data of similar quality to the CPS are not available for most developing countries. So the data that most macro development people work with is similar to US census data.

4 Explaining the Gap in Measured Output per Worker

We have now documented the existence of large sectoral productivity gaps for states in the US. When measured using conventional data sources they are around a factor of 3 at the

second jobs is the main issue for this dimension.

Table 5: Sectoral Productivity Gap

Experiment	Baseline	No Mining	Human Capital		Cumulative
			Equal RR	Different RR	
Results					
Median	1.3	1.3	1.1	0.9	0.9
90th Percentile	2.2	2.1	2.0	1.4	1.4
Maximum	7.4	4.6	6.2	4.7	2.9

Table notes: Distributional results are for the sectoral productivity gap of the fifty states for four years: 1970, 1980, 1990, and 2000.

90th percentile; careful measurement only increases, rather than decreases the gaps. Now we turn to the question of whether the abundant data in the US can help explain these gaps.

4.1 Mining

In some states, mining is a very important component of the nonagricultural sector, particularly in the earlier years. Indeed, inspection of figure 1 shows that Alaska is an extreme outlier from 1970-1990; probably much of their high apparent output per worker in nonagriculture is due to oil. We want to note first that this problem is not special to the US. Many studies of sectoral productivity gaps include natural resource-rich countries. Nonetheless, high labor productivity in mining may be misleading given that mining is capital-intensive, implying that labor payments are much lower than value added. Further, it may not be a reasonable counterfactual to consider reassigning agricultural workers to the mining sector while holding value added per worker fixed.

To explore the quantitative magnitude of this effect we recalculate the labor productivity gap. We exclude the labor input and value added in mining from the nonagriculture sector and recompute our baseline statistics. These figures are given in column 3 of Table 5. Column 2 repeats the baseline statistics from our earlier sections. What we can see is that mining has little effect on the gap for the median state and a modest effect at the 90th percentile. However for a few states it makes a big difference, which results in a large reduction in the maximum gap from 7.4 to 4.6.

4.2 Human Capital

A second alternative is that the workers in agriculture and nonagriculture may be different in terms of the human capital they possess. The population censuses include information on the educational attainment of agricultural and nonagricultural workers. We convert these data to information on years of schooling and verify that agricultural workers typically have fewer years of schooling. To explore the significance of this channel, we translate years of schooling into measures of human capital in each state and sector. We explore two methods for doing so. The first draws on Caselli and Coleman (2006) and relates the human capital h_{ij} of workers in state i and sector j to their years of schooling s_{ij} via $\log(h_{ij}) = 0.1 s_{ij}$. We follow Caselli and Coleman in using a return per year of schooling of 10 percent, which is roughly the average return across countries but also close to the average return to schooling in the US in 2000.[8] Column 4 of Table 5 shows the statistics for the unexplained labor productivity gap after we adjust for human capital in this way. Adjusting for human capital results in a modest reduction of sectoral productivity gaps across the board. For example it falls from 2.2 to 2.0 at the 90th percentile.

This first approach to adjusting for human capital is based on the standard cross-country approach. The assumption is that while agricultural workers have less schooling, each year of schooling is equally valuable in either sector. An alternative assumption is that the value of a year of schooling itself differs across sectors.To be concrete, we allow sectors to differ in the relationship between years of schooling and human capital:

$$\log(h_{ij}) = \mu_j s_{ij}.$$

μ_j indexed the education-intensity of sector j, which we assume to be the same for all states.

To implement this idea we need to measure the education-intensity of agriculture and nonagriculture. Under the assumption of competitive factor markets, it is easy to do so. In this case, workers' wages are proportional to their human capital. To identify the relationship between human capital and schooling in sector j, we can estimate the relationship between wages and schooling in sector j, as in Bils and Klenow (2000). Empirically this involves regressing a Mincer log-wage regression of the form:

$$\log(w_{ij}) = \beta_j X_{ij} + \mu_j s_{ij} + \varepsilon_{ij}$$

[8] We obtain nearly identical results if we instead use the transformation suggested by Hall and Jones (1999).

where X_{ij} is a vector of control variables. We implement this regression separately for workers in agriculture and nonagriculture for the 1970-2000 population censuses. We measure wages as average hourly wage and control for age, sex, and state of residence using dummy variables. Details are available in the appendix.

We find that the return to a year of schooling μ_j is lower for agriculture than nonagriculture in each census. Averaging across censuses, the return to a year of schooling in agriculture is 3.9 percent while the return to a year of schooling in nonagriculture is 6.9 percent. From this fact we conclude that nonagriculture is more education-intensive, in addition to having more educated workers. In column 5 of Table 5 we adjust human capital in each sector using the sector-specific return to schooling we have just measured. The sectoral productivity gaps decrease still further. The labor productivity gap has essentially disappeared at the median and is cut by around one-third relative to the baseline.

Column 6 reports the summary statistics for our cumulative results, which add up the effect of our explanations. In this case it gives the joint effect of removing mining from nonagriculture and adjusting for human capital differences using sector-specific returns to schooling. The combined effect of these two changes is large, leaving us with no gap for the median state and a factor of just 1.4 for the 90th percentile state. However the largest gap is still quite large, nearly a factor of 3.

4.3 Factor Shares

4.4 Physical Capital

4.5 Quality–Adjusted Land

5 Conclusion

References

Abraham, K. G., S. M. Flood, M. Sobek, and B. Thorn (2008): "American Time Use Survey Extract System: Version 1.0 [Machine-readable database]," Maryland Population Center, University of Maryland, College Park, Maryland and Minnesota Population Center, University of Minnesota, Minneapolis, Minnesota, AVailable online at `http://www.atusdata.org`.

Adamopoulous, T., and D. Restuccia (2010): "The Size Distribution of Farms and International Productivity Differences," mimeo, University of Toronto.

Banerjee, A. V., and E. Duflo (2005): "Growth Theory Through the Lens of Development Economics," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. N. Durlauf, vol. 1A, chap. 7, pp. 473–554. Elsevier Science, North-Holland Publishers.

Bils, M., and P. J. Klenow (2000): "Does Schooling Cause Growth?," *American Economic Review*, 90(5), 1160–1183.

Caselli, F. (2005): "Accounting for Cross-Country Income Differences," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. N. Durlauf, vol. 1A, chap. 9, pp. 679–741. Elsevier.

Caselli, F., and W. J. Coleman, II (2006): "The World Technology Frontier," *American Economic Review*, 96(3), 499–522.

Cociuba, S., E. C. Prescott, and A. Ueberfeldt (2009): "U.S. Hours and Productivity Behavior Using CPS Hours Worked Data: 1947-III to 2009-III," mimeo, Arizona State University.

Donovan, K. (2011): "Risk, Farm Ownership, and International Productivity Differences," mimeo, Arizona State University.

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

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

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

Herrendorf, B., and A. Valentinyi (forthcoming): "Which Sectors Make Poor Countries So Unproductive?," *Journal of the European Economic Association*.

Lagakos, D., and M. Waugh (2010): "Specialization, Agriculture, and Cross-Country Productivity Differences," mimeo, Arizona State University.

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

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010): "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]," Minneapolis: University of Minnesota.

A Data Sources, Explanation, and Construction

In this section we list the raw data that we use. We also explain how to interpret these data and construct the statistics used in the paper.

A.1 Value Added

This section describes how we construct farm sector value added from USDA records. We define value added in agriculture as the value of output produced on farms minus the expenses for intermediate inputs used by farms. In contrast, as discussed in the text, the USDA and BEA define it as the value added that is produced by *farmers* (or farm operators). To construct our measure at the state level, we use the value–added spreadsheets that the USDA offers at the state level. These spreadsheets are available since 1949 for all states except for Alaska, Hawaii, and Washington, D.C. For Alaska and Hawaii they are available since 1960 and for Washington, D.C. they are not available.[9] We now explain in detail what we include and what we exclude.

We construct income in agriculture as follows:

- We include the value of crop production. The USDA reports values for eight types of crops, as well as total values for home consumption and inventory adjustment.

- We include the value of livestock production. The USDA reports values for four types of livestock, as well as values for home consumption and inventory adjustment.

- With respect to the revenues from services and forestry:
  - We include the value of machine hire and customwork, because this item includes payments for providing to others services closely related to the farm. Examples are planting, plowing, spraying, or harvesting for others.
  - We exclude the value of forest products sold.
  - We include the value of other income, because this item is closely related to agricultural operations. Examples include animal boarding, breeding fees, and energy generated on the farm.
  - We exclude the gross imputed rental value of farm dwellings.

We construct expenses in agriculture as follows:

[9] The spreadsheets are at `http://www.ers.usda.gov/Data/FarmIncome/Zip_filesXls.htm`.

- We include the value of what the USDA calls farm–origin expenses. In this category, the USDA reports feed purchased, livestock and poultry purchased, and seed purchased.
- We include the value of manufactured inputs. The USDA includes fertilizers and lime, pesticides, petroleum fuel and oils, and electricity in this category.
- We include the value of other purchased inputs if they correspond to expenses. We exclude them, if they correspond to factor payments. In particular:
  - We exclude the value of repair and maintenance of capital items, in line with usual NIPA procedures.
  - We include the value of expenses for machine hire and custom work.
  - We include the value of marketing, storage, and transportation expenses.
  - We exclude the value of contract labor. The reason is that the value of contract labor equals the payments made to contractors or crews that provide labor to farms, which is a factor payment.
  - We include the value of miscellaneous expenses. Examples include costs of animal health care and insurance.

The USDA spreadsheets also include data on what it calls net government transactions (direct government payments minus motor vehicle registration and licensing fees and property taxes). We use these data to construct our measure of value added in agriculture plus subsidies.

A.2 Census Labor

To calculate labor input derived from the U.S. population census, we use the public–use census data made available through IPUMS (Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek 2010). We use the largest publicly available sample for each census year: these are the five–percent samples from 1980 onward and the one–percent samples up to 1970. We impose as little sample selection as possible. We require that workers be in the labor force and employed (empstat = 1). Since this question was only asked of non–minors, we exclude workers under 14 years old (1940–1960) or under 16 years old (1970–). We also exclude workers with invalid or missing industry or occupation codes.

To assign workers to the agriculture and nonagriculture sectors, we need to construct a crosswalk from workers' self–reports of industry/occupation and to our agriculture/nonagriculture classification. One potential complication is that the U.S. census uses a different coding scheme for occupation and industry in every year. Fortunately, however, these schemes are reasonably detailed. Throughout we work with the original classification schemes from the censuses, rather than the re–codings performed by IPUMS ("occ" and "ind" rather than say "occ1950" and "ind1950").

Our primary form of classification draws on which industry workers report. Using the reported industries, we construct three measures of the labor force in the farm sector, each including progressively more workers:

- Narrow: animal and crop production industries.
- Intermediate: Narrow plus agricultural services.
- Broad: Broad plus fishing, forestry, and horticulture.

In practice, the 1950 and 1960 industry classifications were fairly coarse, with only three agricultural industries: agriculture, fishing, and forestry. In these early years our narrow and broad definitions of farm sector employment coincide. From 1970 onward the industries are coded more finely, allowing us to calculate all three measures. Table 6 gives the full crosswalk. It lists for each year all industries (and corresponding industry codes) that we use.

As a robustness check on our results, we also experiment with defining workers in the agriculture sector using their reported occupation rather than industry. That is, we identify workers who report being farmers, ranchers, farm managers, or farm laborers rather than those who report being in the animal and crop production industry. Again, the census includes a measure of occupation that varies with each census and generally becomes more detailed over time. Table 7 gives the occupation titles and corresponding codes that we associate with the agriculture sector for each census year.

After dividing the population into the agriculture and non-agriculture sectors, we calculate employment and hours worked by sector. We always arrive at the average using weights (perwt). To calculate our employment statistics, we add the number of individuals employed in each state and industry. We compute annual hours worked using information on the weeks worked in a year and the hours worked in a week. The information used varies by census. For weeks, the 1940, 1950, and 1980–2000 censuses, include the actual weeks worked in the previous year. In the 1960 and 1970 censuses, workers reported their weeks worked in an interval; we recode each value to the midpoint of the interval, so that workers reporting 50–52 weeks worked are coded as having 51 weeks. For hours, the 1940 and 1950

Table 6: Coding Census Industries to the Agriculture Sector

Year	Narrow	Intermediate = Narrow, Plus:	Broad = Intermediate, Plus:
1950	Agriculture (105)	N/A	Forestry (116)
			Fishing (126)
1960	Agriculture (016)	N/A	Forestry (027)
			Fishing (028)
1970	Agricultural Production (017)	Agricultural Services, except Horticultural (018)	Horticultural Services (019)
		Agriculture, Forestry, and Fisheries - Allocated (029)	Forestry (027)
			Fisheries (028)
1980	Agricultural Production, Crops (010)	Agricultural Services, except Horticultural (020)	Horticultural Services (021)
	Agricultural Production, Livestock (011)		Forestry (030)
			Fishing, Hunting, and Trapping (031)
1990	Agricultural Production, Crops (010)	Agricultural Services, n.e.c. (030)	Lanscape and Horticultural Services (020)
	Agricultural Production, Livestock (011)		Forestry (021)
			Fishing, Hunting, and Trapping (032)
2000	Crop Production (017)	Support Activities for Agriculture and Forestry (029)	Forestry Except Logging (019)
	Animal Production (018)		Logging (027)
			Fishing, Hunting, and Trapping (028)

Table notes: Numbers in parentheses correspond to codes for the variable ind in the IPUMS data.

Table 7: Coding Census Occupations to Agriculture Sector

Year	Agriculture	
1950	Farmers (Owners and Tenants) (100)	Farm Managers (123)
	Farm Foremen (810)	Farm Laborers, Wage Workers (820)
	Farm Laborers, Unpaid Family Workers (830)	Farm Service Laborers, Self–Employed (840)
1960	Farmers (Owners and Tenants) (200)	Farm Managers (222)
	Farm Foremen (901)	Farm Laborers, Wage Workers (902)
	Farm Laborers, Unpaid Family Workers (903)	Farm Service Laborers, Self–Employed (905)
1970	Farmers (Owners and Tenants) (801)	Farm Managers (802)
	Farmers and Farm Managers, Allocated (806)	Farm Foremen (821)
	Farm Laborers, Wage Workers (822)	Farm Laborers, Unpaid Family Workers (823)
	Farm Service Laborers, Self–Employed (824)	Farm Laborers and Farm Foremen, Allocated (846)
1980	Farmers, Except Horticultural (473)	Managers, Farms, Except Horticultural (475)
	Supervisors, Farm Workers (477)	Farm Workers (479)
	Supervisors, Related Agricultural Occupations (485)	Graders and Sorters, Agricultural Products (488)
1990	Farmers, Except Horticultural (473)	Managers, Farms, Except Horticultural (475)
	Supervisors, Farm Workers (477)	Farm Workers (479)
	Farm Laborers and Farm Foremen, Allocated (480)	Supervisors, Related Agricultural Occupations (485)
	Graders and Sorters, Agricultural Products (488)	
2000	Farm, Ranch, and Other Agricultural Managers (20)	Farmers and Ranchers (21)
	Supervisors/Managers of Farming, Fishing, and Forestry Workers (600)	Graders and Sorters, Agricultural Products (604)
	Miscellaneous Agricultural Workers, Including Animal Breeders (605)	

Table notes: Numbers in parentheses correspond to codes for the variable “occ” in the IPUMS data.

censuses record hours worked last week and the 1980–2000 censuses record usual hours worked in a week. In the 1960 and 1970 censuses, workers reported their hours worked in the last week in an interval; we again recode each value to the midpoint of the interval.

A.3 CPS Labor

Our baseline figures on employment by state and sector draw on the monthly Current Population Survey (CPS) data constructed by the Bureau of Labor Statistics. We use data from all months because sufficient data are available for our purposes and we want to study how employment and hours worked vary over the year. In contrast, many other studies use just the March file. CPS is a rotating panel, with each housing unit participating in the sample for four months, then dropped for eight months, and then returned for four months. Information on participants' second jobs was collected in their fourth and eighth month in the sample, so we restrict our sample to these months throughout. We multiply standard CPS weights by 4 to correct for the roughly 75 percent loss in sample from this restriction.

We now explain how we construct our measures using CPS data. We use the CPS Basic Monthly Data and the corresponding dictionaries provided on the NBER website.[10] Our sample selection is similar to that in the Census. We require that workers be in the labor force and employed, i.e., employed and at work the previous week or employed and away from work (pemlr = 1 or 2). The CPS does not collect labor force and employment status for workers in the armed forces, so we separately include adults in the armed forces (prpertyp = 3) in the sample regardless of labor force status. Finally, we exclude workers with invalid or missing industry or occupation codes for their first jobs.

The CPS also collects information on whether workers have more than one job. If they do, they collect the occupation and industry of employment for the first and second job. They use the same occupation and industry coding schemes to record information about the first and second jobs, so the crosswalks discussed below apply equally to first and second job.

As with the census, we construct four measures of workers employed in agriculture. The first are based on the workers' reported industry of occupation, and include the same narrow, intermediate, and broad categories. Over the ten year period that we study, the CPS used two different coding schemes for industries and occupations. From 1995–2002 they use a coding scheme very similar to the one used in the 1990 census; from 2003–2005 they use a coding scheme very similar to the one used in the 2000 census. For completeness,

[10]Data available at `http://www.nber.org/data/cps_basic.html`.

Table 8 gives the crosswalk from industry codes to agriculture/nonagriculture status for both coding schemes.

Again, as with the census we compute one alternative definition of workers in the farm/non–farm sectors that uses reported occupation rather than reported industry as its basis. The CPS coding scheme from 1995–2002 is almost identical to the 1990 census coding scheme, while the CPS coding scheme from 2003–2005 is almost identical to the 2000 census coding scheme. Table 9 gives the crosswalk from occupation codes to farm sector status.

We compute employment by state and sector using the above definitions. Employment is weighted by individual weights (pwsswgt).

A.4 ATUS Labor

Our last data source is the American Time Use Survey, which available online from IPUMS as ATUS–X (Abraham, Flood, Sobek, and Thorn 2008). ATUS is a time–use study funded by the BLS and fielded by the U.S. Census Bureau since 2003. It is organized as a follow–up to the CPS. Some members of each outgoing rotation group (those in their eighth month in the sample) are invited to continue into ATUS. If they choose to do so, they are interviewed again 2–5 months after their final month in the CPS. Some of the demographic information available in the CPS is collected again in ATUS. The key new data are collected in a time diary, in which respondents are asked to record how they use their time during one 24–hour period from 4am to 4am. They record their time usage in minutes, selecting from over 400 detailed activities.

Our extract includes data from the years 2003–2009. For the most part our sample selection is the same as in the CPS. We require that workers be employed, whether at work or absent from it at the time of the interview (empstat = 1 or 2). We exclude workers with invalid or missing industry or occupation codes for their first job.

A complication arises from the fact that ATUS does not collect the industry and occupation of the second job. More specifically, at the time of the final CPS interview, they collect information on whether or not the respondent works multiple jobs, as well as the industry, occupation, and hours worked in up to two jobs. At the time of the ATUS interview they collect information on whether or not the respondent works multiple jobs and what industry and occupation are for the first job. This presents a challenge for our work because we do not know from ATUS records whether a second job is in agriculture or nonagriculture. To minimize the impact of this problem, we choose to keep only workers who either (i) work one job at the time of the ATUS interview, or (ii) who work multiple jobs at

Table 8: Coding CPS Industries to the Agriculture Sector

Year	Narrow	Intermediate = Narrow, Plus:	Broad = Intermediate, Plus:
	Agricultural Production, Crops (010)	Agricultural Services, N.E.C. (030)	Landscape and Horticultural Services (020)
1995-2002	Agricultural Production, Livestock (011)		Forestry (031)
			Fishing (032)
	Crop Production (0170)	Support Activities for Ag. and Forest. (0290)	Forestry Except Logging (0190)
2003-2005	Animal Production (0180)		Logging (0270)
			Fishing, Hunting, and Trapping (0280)

Table notes: Numbers in parentheses correspond to codes for the variable peio1icd and peio2icd in the monthly CPS extracts.

Table 9: Coding CPS Occupations to the Agriculture Sector

Year	Farm	
1995–2000	Farmers, Except Horticulture (473)	Managers, Farms Except Horticulture (475)
	Supervisors, Farm Workers (477)	Farm Workers (479)
	Supervisors, Agricultural Operations (485)	Graders and Sorters, Agricultural Products (488)
2003–2005	Farm, Ranch, and Other Agricultural Managers (0200)	Farmers and Ranchers (0210)
	Mangers of Farming, Fishing, and Forestry Workers (6000)	Animal Breeders (6020)
	Graders and Sorters, Agricultural Products (6040)	Miscellaneous Agricultural Workers, Including Animal Breeders (6050)

Table notes: Numbers in parentheses correspond to codes for the variable peio1ocd and peio2ocd in the monthly CPS data.

the times of the CPS and the ATUS interviews and list the same industry and occupation for their first job. In this case we assume that the industry and occupation for the second job are also unchanged. The crosswalk from industry and occupation codes in the ATUS to farm/nonfarm status is the same as in the CPS for 2003–2005 (see Tables 8 and 9).

In addition to the standard demographic variables, we use two pre–defined time–use variables. The first is "work, main job", which includes the minutes in the diary day spent on activities at work, checking e–mail, making phone calls to/from clients, putting in overtime, or receiving on–the–job training. This category explicitly excludes "work–related" activities, such as socializing, eating, drinking, or playing golf with coworkers or clients. It also excludes travel for most workers, unless the travel is an essential part of business, such as for taxi–drivers or traveling salespersons. The second pre–defined time–use variable we use is "work, other job(s)", which includes the same information as "work, main job" but for all other jobs.

We aggregate this information to obtain the time spent working on first and second jobs in an average week. To do so we translate minutes per day into hours. We then take the average hours per week spent in work activities in the population, using individual–level weights (wt06). The resulting numbers are directly comparable with those from the CPS.