

Why is Measured Productivity so Low in Agriculture?*

Berthold Herrendorf and Todd Schoellman
Arizona State University[†]

December 12, 2011

Abstract

It is well known that poor countries are much less productive in agriculture than in the rest of the economy, and that it is hard to account for these productivity gaps. In this paper, we study US states during 1980–2009. We find that there are large productivity gaps between agriculture and non-agriculture. These productivity gaps are not at all accounted for by gaps in real wages per efficiency unit, which are similar in the two sectors. Instead, they are accounted for by two key factors: human capital is much higher in non-agriculture; and value added is seriously mis-measured in agriculture.

Keywords: agriculture; measurement error; productivity gaps.

JEL classification: O1.

*For helpful discussions and suggestions, we thank Roozbeh Hosseini, Moritz Kuhn, David Lagakos, Patrick Minford, Edward Prescott, Richard Rogerson, Ákos Valentinyi, and Gustavo Ventura, as well as seminar participants at Arizona State University, the Free University of Berlin, the 2011 Midwest Macroeconomics Meetings, the 2011 Meetings of the Society for Economic Dynamics, the 2011 Cologne Workshop on Macroeconomics, and the Universities of Toronto and Western Ontario. For financial support, Herrendorf thanks the Spanish Ministry of Education (Grant ECO2009-11165).

[†]Address: Department of Economics, W.P. Carey School of Business, Arizona State University, Tempe, AZ 85287-9801, USA. E-mails: berthold.herrendorf@asu.edu and todd.schoellman@gmail.com

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 2 | Measuring Productivity Gaps | 5 |
| 2.1 | Conventional measurement for 2000 | 5 |
| 2.2 | Improved measurement | 7 |
| 2.2.1 | Hours worked | 7 |
| 2.2.2 | Averages by decade | 9 |
| 3 | Measuring Wage Gaps | 11 |
| 3.1 | Wages and human capital | 12 |
| 3.2 | Consumer price level | 14 |
| 3.3 | Real wages per efficiency unit | 15 |
| 4 | Connecting Productivity Gaps with Wage Gaps | 16 |
| 4.1 | An accounting framework | 16 |
| 4.2 | Implied labor shares | 18 |
| 5 | Mis-measurement of Agricultural Productivity | 19 |
| 5.1 | The BEA's notion of farming | 19 |
| 5.2 | Under-reporting of proprietors' income | 22 |
| 6 | Conclusion | 26 |
| A | Data Appendix: Sectoral Value Added | 29 |
| A.1 | BEA | 29 |
| A.2 | USDA | 29 |
| B | Data Appendix: Sectoral Labor Inputs | 31 |
| B.1 | Quarterly Census of Employment and Wages | 31 |
| B.2 | Population Census | 31 |
| B.3 | Current Population Survey | 34 |
| C | Data Appendix: Proprietors' Income | 37 |
| C.1 | Census of Agriculture | 37 |
| C.2 | March supplement of the CPS | 38 |
| D | Data Appendix: Consumer Price level | 38 |

1 Introduction

Poor countries have the majority of their labor forces in agriculture while being much less productive there than in the rest of the economy; see for example, Caselli (2005) and Restuccia, Yang and Zhu (2008). This fact raises two obvious questions: Why are there large sectoral productivity gaps between non-agriculture and agriculture? Why do people not leave agriculture? There have been many attempts to answer these questions. Recent examples point to the scale or risk of farming (Adamopoulos and Restuccia 2011, Donovan 2011); to barriers between agriculture and non-agriculture (Restuccia et al. 2008, Herrendorf and Teixeira 2009); to differences in factor endowments (Caselli 2005); to selection of the workers in the two sectors (Lagakos and Waugh 2010); and to home production (Gollin, Parente and Rogerson 2004). Although these contributions advance our understanding of what is special about agriculture, it is probably fair to say that to date we do not have a conclusive answer to the questions posed above. The main reason for this is the scarcity of direct evidence from poor countries, which is required to connect the different theories to the data.

In this paper, we go down a different path and study productivity gaps between non-agriculture and agriculture in US states. Focusing on US states has the advantage that we can draw on richer data than are generally available for developing countries. Further, each piece of information (such as value added) is collected by a single statistical agency, ensuring cross-state comparability. Nonetheless, some readers might think that there is not much to be learned from studying US states. After all, agriculture produces only a small percentage of total US value added and there should only be a small productivity gap between non-agriculture and agriculture that can easily be accounted for by obvious explanations. In what follows we will present evidence to the contrary. To begin with, although US agricultural value added is small relative to GDP, its absolute size is large; e.g., relative to the total population agricultural value added in the United States is of the same order of magnitude as in Africa. More surprisingly, standard measurement generates large productivity gaps between non-agriculture and agriculture in many US states, and, as we will show, these gaps are surprisingly hard to account for.

As a first pass at the evidence, we choose US data sources that are as similar as possible to those that are available for developing countries. In particular, for sectoral value added we use the regional accounts of the BEA (which form the basis for NIPA) and for sectoral employment we use workers from the Population Census or the BEA (which bases its employment numbers on the Quarterly Census of Employment and Wages). We start

by calculating “conventional measures” of labor productivity (measured as value added per worker in current dollars) for non-agriculture and agriculture for the year 2000. These conventional measures suggest that in most US states, labor productivity was considerably higher in non-agriculture than in agriculture. Our baseline statistic is the productivity gap, which we define as the ratio of labor productivity in non-agriculture to the labor productivity in agriculture measured in current dollars. In several states the productivity gap was more than a factor of three and in the maximum state it was almost a factor of ten. We also find that the productivity gaps calculated based on employment from the Population Census tend to be considerably smaller than those calculated based on employment from the BEA, suggesting that measured productivity gaps are fairly sensitive to the underlying data source for employment. This suggests to be cautious about employment numbers from developing countries, which are based on data sources that are similar to the two we have used.

There are three obvious concerns with our first pass at the evidence: the two different sources of sectoral employment produce rather different productivity gaps; 2000 could have been an unusually bad year for US agriculture; in some states the agricultural sector is relatively small so that our calculations are based on few observations only. To address these concerns, we measure sectoral hours worked (instead of employment) from the Current Population Survey, we extend our measurement to the thirty year period 1980–2009, we report averages by decade (the 1980s etc), and we exclude the five states with the smallest agricultural sectors. We find that sizeable productivity gaps remain, and that these gaps do not decline over time. To give some concrete numbers, for the median state the productivity gap is 1.9 while for the 90th percentile the productivity gaps is 3.0. So even our considerably improved measurement leads to the surprising conclusion that during 1980–2009 there were large and sustained productivity gaps in most US states. We therefore take these numbers seriously and devote considerable effort to exploring what might account for them.

We start by asking whether the large measured productivity gaps suggest that there is scope for beneficial reallocation of workers from agriculture to non-agriculture. To answer this question, we calculate real wages per efficiency unit for non-agriculture and agriculture at the state level. In this context, the real wage per efficiency unit in a sector is defined as the average nominal wage per hour (i.e., in current dollars) divided by the product of the average consumer price level and the average human capital of the workers in the sector, where human capital is calculated in the standard Mincer way. We find that at the state level there are sizeable differences in human capital between non-agriculture and agriculture; in the median state, an average worker in non-agriculture has about twice the

human capital than an average worker in agriculture. The consumer price level is also higher in non-agriculture, but in comparison with human capital the difference in the consumer price level is not important quantitatively. After taking into account the differences in human capital and the consumer price level, the gaps in real wages per efficiency unit are small only. In other words, there is little scope for beneficial reallocation of workers from agriculture to non-agriculture. This finding is of interest for the development literature because it provides an example where large measured productivity gaps do not imply that there is scope for beneficial reallocation which is prevented by large moving costs (some sort of “barriers”).

The obvious next question is how large gaps in productivity may co-exist with small gaps in real wages per efficiency unit. To address this question, we derive an accounting identity that links the two gaps and shows that except for sectoral gaps in real wages per efficiency unit (which we found to be small), productivity gaps result from sectoral differences in human capital, the consumer price level, and the labor share in value added. To build some intuition for how these differences matter, suppose that in each state non-agriculture and agriculture pay exactly the same real wage per efficiency unit. Then productivity gaps arise for two reasons. First, productivity will be higher in non-agriculture than in agriculture if human capital and the consumer price level are higher in non-agriculture. As we reported above, this is the case. Moreover, the difference in human capital is sizeable, implying that in the US a considerable part of the productivity gap between non-agriculture and agriculture comes from the fact that workers in non-agriculture have more human capital. Second, productivity will be higher in non-agriculture than in agriculture if the labor share is lower in non-agriculture. The reason for this is that, given the assumption that both sectors pay the same real wage per efficiency unit, the sector with the lower labor share needs higher value added to be able to pay that real wage per efficiency unit. However, the existing evidence suggests that the opposite is the case, that is, US agriculture has a lower labor share than non-agriculture, as it uses capital and land more intensively. We show that taken together with the differences in human capital and the consumer price level, this implies that standard accounting cannot reconcile the large gaps in productivity with the small gaps in real wages per efficiency unit that we found. This suggests that at least one of the two gaps is mis-measured.

We provide evidence that the productivity gaps are mis-measured. In particular, we demonstrate that the BEA’s under-estimates value added in agriculture along three dimensions: (i) it does not include several factor payments that conceptually belong to agriculture; (ii) it does not include agricultural subsidies at the state level; (iii) it does not correct suf-

ficiently for under-reporting of proprietors' income in agriculture. We show that making appropriate corrections to address these flaws reduces the measured productivity gaps by enough that we can now account for them, at least at the median state. We take this to suggest that one should be skeptical of the large productivity gaps that many developing countries report, at least to the extent that developing countries have lower data quality than the US. We conclude that improving the measurement of agricultural productivity in developing countries is of first-order importance in order to establish robust facts that can guide future research on sectoral productivity gaps.

Our paper joins a recent literature about sectoral productivity differences across countries, which aims to identify the sectors that make poor countries unproductive; see for example Herrendorf and Valentinyi (2011). A growing part of this literature focuses on the apparent gap between non-agricultural and agricultural productivity. Our paper is most closely related to Gollin, Lagakos and Waugh (2011), which also seeks to measure and account for gaps between non-agricultural and agricultural productivity. The difference between the two papers is that Gollin et al use evidence from survey data for poor countries whereas we use evidence for US states. The approach of Gollin et al has the obvious appeal of providing more direct evidence regarding productivity gaps in poor countries. Our approach has the obvious appeal of allowing us to bring to bear a wealth of detailed, harmonized, and well-documented data that do not exist in developing countries. With regards to the conclusions, the two papers nicely complement each other along at least two dimensions. First, they also find that in countries for which reliable estimates exist, sectoral differences in labor input and human capital are important explanatory factors of productivity gaps. Second, they are also unable to account for a sizeable part of the productivity gaps with the standard explanatory factors. Our work suggests that this may be due to persistent mis-measurement of agricultural productivity.

The remainder of the paper proceeds as follows. Section 2 establishes that there are large measured productivity gaps in US states and Section 3 establishes that there are small gaps in real wages per efficiency unit. Section 4 presents several accounting identities, connects them with the data, and shows that the findings of the previous two sections are inconsistent. Section 5 establishes that there is severe under-measurement of agricultural value added, which explains a large part of the productivity gaps. Section 6 concludes. An appendix contains a detailed description of our data sources.

2 Measuring Productivity Gaps

In this section, we document that measured productivity in non-agriculture is higher than in agriculture in most US states, and that the difference can be large quantitatively. Productivity is defined as value added in current dollars per worker or per hour. Since the ratio of productivity in non-agriculture to productivity in agriculture is unit free, we can compare it across states and time. If this ratio is not equal to one, then we say that there is a *productivity gap*. We will first measure productivity gaps using the conventional data sources that are typically available for developing countries. We will then use better data on hours worked that is typically not available for developing countries.

2.1 Conventional measurement for 2000

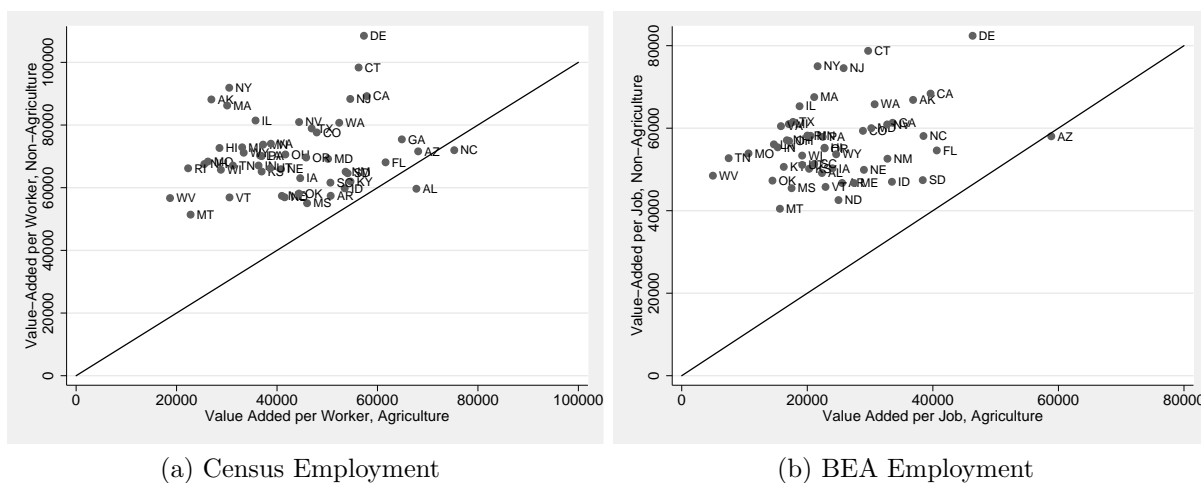


Figure 1: Sectoral Value Added per Worker in 2000

We define the agricultural sector as farming, that is, agriculture comprises the farm industries crop and animal production. Our definition of agriculture is consistent with that of the Food and Agricultural Organization (FAO). It does not include fishing, horticulture, hunting, or trapping, and it includes forestry only to the extent to which it is done on farms. We define the non-agriculture sector as all industries other than the farm industries and the military. The reason for excluding the military is that we do not have employment data by state for it.

We start by using data sources for US states that are as similar as possible to those that are available for developing countries. For agricultural value added in current prices, the

Table 1: Productivity Gaps in 2000

| | Census | BEA |
|-----------------------------|--------|-----|
| Median | 1.7 | 2.5 |
| 90 th Percentile | 2.8 | 3.7 |
| Maximum | 3.3 | 9.8 |

standard sources are the United Nations National Accounts and the World Development Indicators (WDI), which are both based on national accounts. The closest comparable numbers for US states are industry value added by state from the BEA’s regional accounts. For employment, the standard data source is the International Labor Organization (ILO), which is based on population censuses or employment surveys. The closest comparable numbers for US states are the employment numbers from the Population Census or from the BEA, which are based on the Quarterly Census of Employment and Wages. We obtain the Census numbers from the public–use version made available through Ruggles, Alexander, Genadek, Goeken, Schroeder and Sobek (2010) and the BEA numbers from the BEA’s regional accounts.¹

To take a first look at the data, we initially focus on the year 2000. Figure 1 shows the productivity gaps for 2000, with panel (a) referring to Census employment and panel (b) referring to BEA employment. Agricultural value added per worker is depicted on the x–axis and non–agricultural value added per worker is depicted on the y–axis. For almost all states productivity is higher in non–agriculture than in agriculture. Moreover, the resulting productivity gaps are sizeable and change considerably depending on whether we use the employment numbers from the Census or the BEA. Table 1 reports key summary statistics, confirming the impression conveyed by Figure 1: the productivity gap in the mean state equals 1.7 or 2.5 depending on which measure of agricultural employment we use. The table also shows that all summary statistics of productivity gaps are larger with BEA labor than with Census labor. Interestingly, with BEA labor, the maximum gap is close to 10, which is a number often mentioned in the context of developing countries.² It is very surprising, at least to us, that in 2000 there was such large sectoral productivity gaps in some US states.

In sum, we find that there are sizeable productivity gaps between non–agriculture and agriculture for the year 2000, and that these gaps depend crucially on the measure of

¹The Appendix contains a detailed discussion of the data sources and what we do to construct the non–agricultural and agricultural sectors.

²Gollin et al. (2011) document for 112 developing countries that the maximum productivity gap is around ten.

Table 2: Sectoral Employment in 2000 (in millions)

| | Census | CPS | BEA |
|-----------------|--------|-------|-------|
| Agriculture | 1.6 | 2.4 | 3.1 |
| Non-Agriculture | 127.8 | 136.9 | 159.5 |

Table notes: Employment in millions at the national level. CPS numbers measured as full-time equivalents.

sectoral employment that we use. In the next subsection, we explore whether these gaps are robust to better measurement.

2.2 Improved measurement

2.2.1 Hours worked

We measure sectoral hours worked using the monthly data from the *Current Population Survey (CPS)*, which we access through the National Bureau of Economic Research. The CPS is a rotating panel survey administered by the BLS to a sample of households that contains questions about the identity of workers' jobs and their hours worked. The survey has always asked about the main job and total hours worked. From 1994 onward it also asks workers in outgoing rotation groups about the identity of their first two jobs and their hours worked at each. Further, the CPS included supplements on multiple job holding for the month of May in 1979, 1980, 1989, and 1991 that collected similar information. We use the monthly files in combination with the information from outgoing rotation groups and earlier supplements to estimate the hours worked in agriculture and non-agriculture for each state, month, and year, accounting for first and second jobs. The details are available in the Appendix.

We start by comparing employment by sector at the national level from the CPS with that from the Population Census and the BEA in the year 2000. We represent the CPS numbers as full-time equivalent employment, assuming 48 weeks worked per year and 40 hours per week. Table 2 shows that CPS employment lies between Census and BEA employment in both sectors. To the extent that the CPS is deemed to be a very reliable source of US employment data, this implies that the Population Census underestimates and the BEA overestimates sectoral employment. The table also shows that the relative discrepancies between the Population Census and the BEA numbers are considerably larger in agriculture than in non-agriculture. This explains why in the previous subsection the measured productivity gaps came out larger when we used BEA employment.

There are two main reasons for the differences between the employment numbers from the Population Census and the BEA. The first one is that the Population Census is taken during the month of March. This is more important in agriculture than non-agriculture, because, as Figure 2 shows, employment in agriculture is rather seasonal and March is a month with below average activity in agriculture. The figure also shows that non-agricultural employment is hardly seasonal.³

The second reason for the differences between the employment numbers from the Population Census and the BEA comes from the way in which the two data sources deal with secondary jobs. The Population Census records only primary jobs, implying that it underestimates employment in both sectors. In contrast, the BEA numbers are based on the Quarterly Census of Employment and Wages, which is an establishment survey and counts all jobs equally. Since the BEA does not adjust secondary jobs for full-time equivalent employment at the state level, it overestimates sectoral employment.⁴ While in principle the different treatments of second jobs could be equally distortive in both sectors – or more important in either one of them – it turns out that the relative discrepancies between the two data sets are larger in agriculture than in non-agriculture. The reason for this is that in the US the ratio of secondary to primary jobs is much larger in agriculture than it is in non-agriculture.

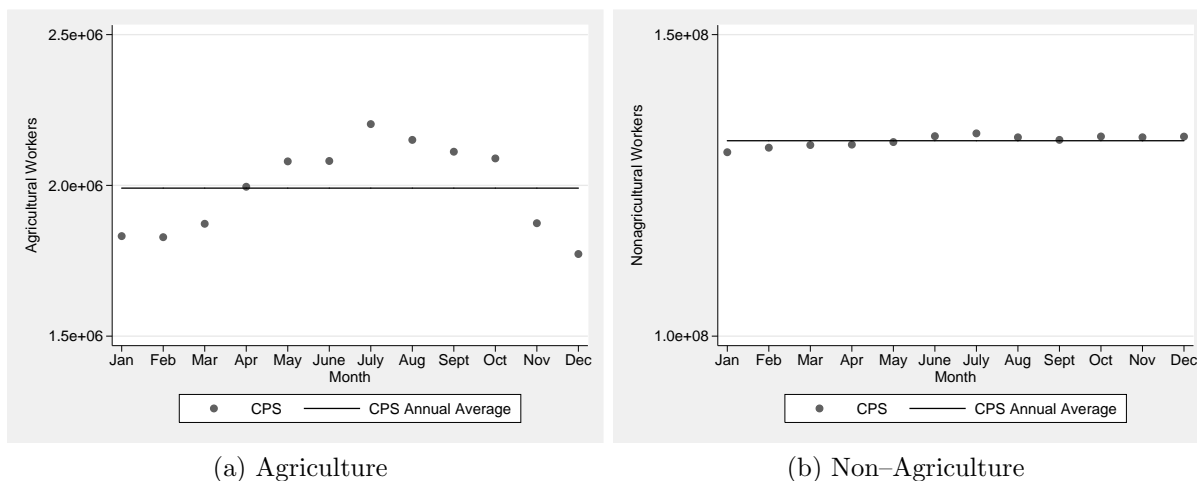


Figure 2: Sectoral employment during 2000

An additional issue with the employment numbers from the Population Census and the

³Note that hours worked and usual hours worked in agriculture show similarly strong seasonal variation.

⁴For what the BEA does at the national level, see the discussion in Cociuba, Prescott and Ueberfeldt (2009).

Table 3: Productivity Gaps with Different Employment Data in 2000

| | Census | CPS | BEA |
|-----------------------------|--------|-----|-----|
| Median | 1.7 | 2.1 | 2.5 |
| 90 th Percentile | 2.8 | 3.5 | 3.7 |
| Maximum | 3.3 | 5.0 | 9.8 |

BEA is that they refer to workers instead of hours worked. If hours per worker are roughly the same in both sectors, then using workers instead of hours worked does not affect the calculations of productivity gaps. The information on hours worked by sector from the CPS allows us to assess whether this is the case for the United States. We find that it is not, and that workers in agriculture tend to work more hours. In particular, restricting attention to workers with only one job, the average farmer works as much as 40.4 hours whereas the average non-farmer works just 37.2 hours. Not taking the sectoral difference in hours worked into account leads to an underestimation of sectoral productivity gaps.

Table 3 compares the summary statistics for value added per worker based on employment with those based on hours worked. The first and the third column are repeated from Table 1 above for convenience. The second column reports the sectoral productivity gaps calculated as the ratio of BEA value added (as before) and CPS hours worked (new). We can see that, as expected, using hours worked increases the productivity gaps compared to using Census employment and reduces them compared to using BEA employment.

2.2.2 Averages by decade

The previous findings were all for the year 2000, which raises the question whether 2000 was an unusually bad year for US agriculture. To address this question, we extend our measurement to 1980–2009. To make sure that we have enough observations in the CPS, we group the data in three non-overlapping ten-year bins: 1980–1989, 1990–1999, and 2000–2009, which we refer to as the 1980s, 1990s, and 2000s. All numbers that we report are the ten-year averages of the sectoral productivity gaps by state within the respective bin. Even with ten-year bins, the CPS contains only few agricultural workers for states that have small agricultural sectors. To make sure that our results are not driven by these states, we require that for all states in our sample the CPS have complete hours information for at least 90 agricultural workers in each decade. This criterion leads us to exclude Alaska,

Connecticut, Massachusetts, Rhode Island, and West Virginia.⁵

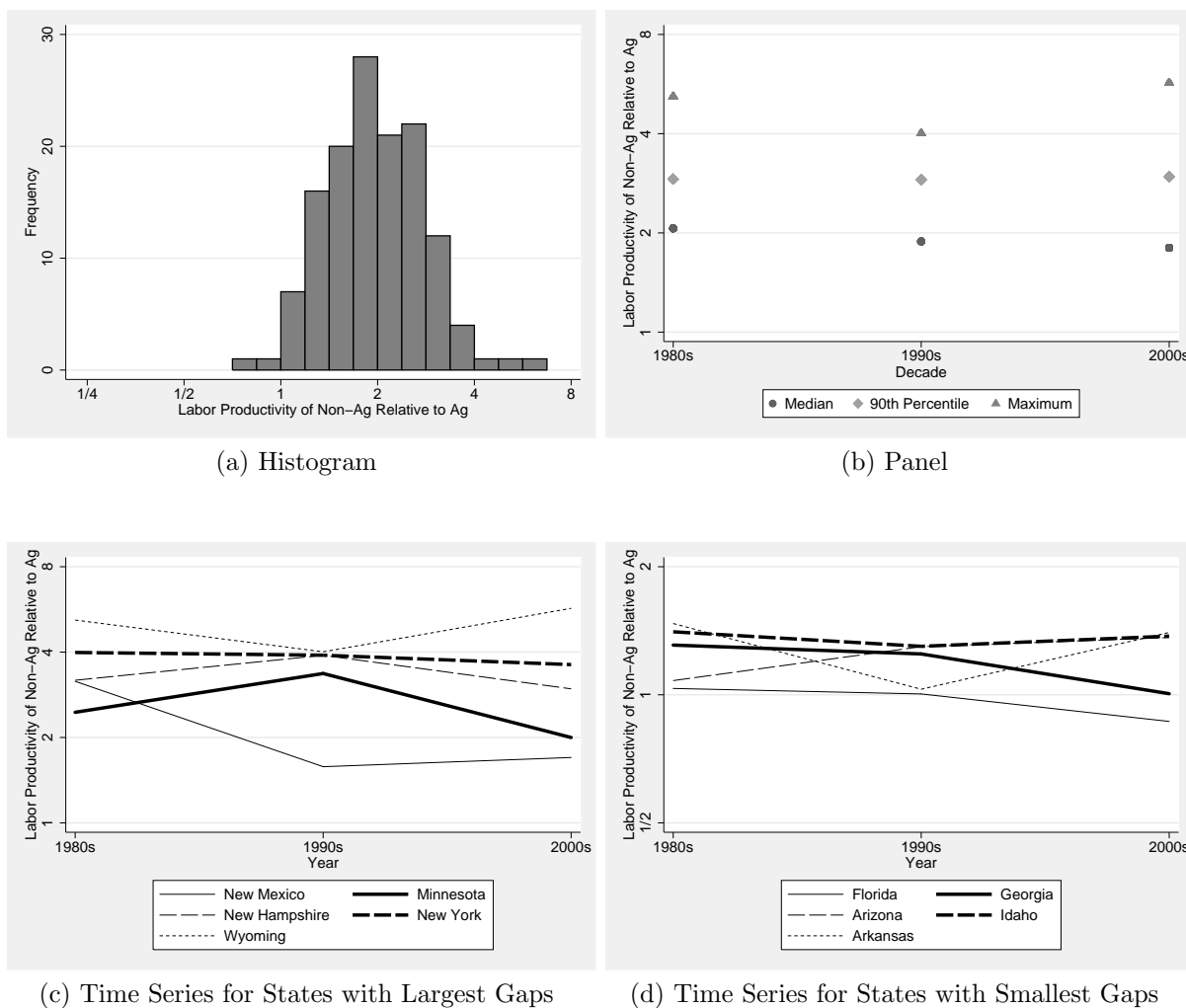


Figure 3: Improved Measurement of Labor Productivity Gaps 1980–2009

Panel (a) of Figure 3 gives the results in the form of a histogram, showing that with the best possible measurement and thirty years of data the stylized fact of Subsection 2.1 survives; in most states and years, there are considerable productivity gaps, that is, productivity in non-agriculture is considerably higher than in agriculture. Panel (b) of Figure 3 shows that the productivity gaps do not decline over time. This, and the fact that we average over ten-year bins, addresses the concern that our results above reflected a bad harvest during the year 2000. Panels (c) and (d) of Figure 3 show the time series for the

⁵Note that although the CPS started before 1980, only since 1978 does it have information for each individual state, which is crucial here. We start in 1980 because this allows us to form natural ten-year bins.

Table 4: Productivity Gaps with CPS Hours 1980–2009

| | |
|-----------------------------|-----|
| Median | 1.9 |
| 90 th Percentile | 3.0 |
| Maximum | 5.7 |

Table notes: Results are for 45 states, excluding five states with small samples: Alaska, Connecticut, Massachusetts, Rhode Island, and West Virginia.

five states with the largest and smallest productivity gaps (of course after excluding Alaska, Connecticut, Massachusetts, Rhode Island, and West Virginia), suggesting that each state has consistently large or small productivity gaps. Table 4 gives the summary statistics for the productivity gaps in the remaining US states: the median gap is 1.9, the gap at the 90th percentile is 3.0, and the maximum gap is 5.7. These gaps are sizeable, which is particularly remarkable given the fact that we have excluded five states from the sample. Moreover, the summary statistics come out surprisingly close to those in developing countries. For example, Gollin et al. (2011) document for a set of 112 developing countries that the median productivity gap is 3 and the 95th percentile is 8.8.⁶

3 Measuring Wage Gaps

In the last section, we have documented that there are large measured productivity gaps in US states. This raises the question whether workers want to relocate from the unproductive sector agriculture to the productive sector non-agriculture, and whether large moving costs (“barriers”) prevent them from doing so. This line of thought seems implausible for the US because it is unclear what form these moving costs should take there. The advantage of studying US states instead of developing countries is that we can go further and investigate empirically whether or not there are barriers. This is the goal of the current section.

We view each sector–state pair as a location sj , where s indexes the state and j the sector ($j \in \{a, n\}$, that is, agriculture or non-agriculture). Households choose one location, which means that they work and buy a final consumption good in that location. We assume that the key determinant of the households’ location choice is the real wages per efficiency

⁶Some readers might wonder whether across US states productivity gaps show the same relationship with state GDP per capita as they do across countries, that is, whether poorer states have higher productivity gaps than richer states. We find that this is not borne out by the data, that is, across US states there is no strong relationship between productivity gaps and state GDP.

unit:

$$w_{sj} \equiv \frac{W_{sj}}{h_{sj}P_{sj}}. \quad (1)$$

W denotes the nominal wage per hour (i.e., the wage in current dollars), h human capital, and P the price of a bundle of final consumption goods that households care about. We will refer to P as the consumer price level. In what follows, we will calculate the different components of the real wage per efficiency unit, starting with the nominal wage and human capital.

3.1 Wages and human capital

To construct a measure of the nominal wage and human capital in each sector and state, we use the CPS. We follow the standard practice and focus on the restricted sample of workers who work at least 30 hours per week, are employed for wages and salaries, and have between zero and fifty years of potential experience (which is defined as age minus years of schooling minus six). For each decade (i.e. 1980s, 1990s, 2000s) and sector (i.e. non-agriculture, agriculture) we run the following regression:

$$\log(W_{sj}^{it}) = \mu_{dj}Z_{sj}^t + \beta_{dj}X_{sj}^{it} + \varepsilon_{sj}^{it},$$

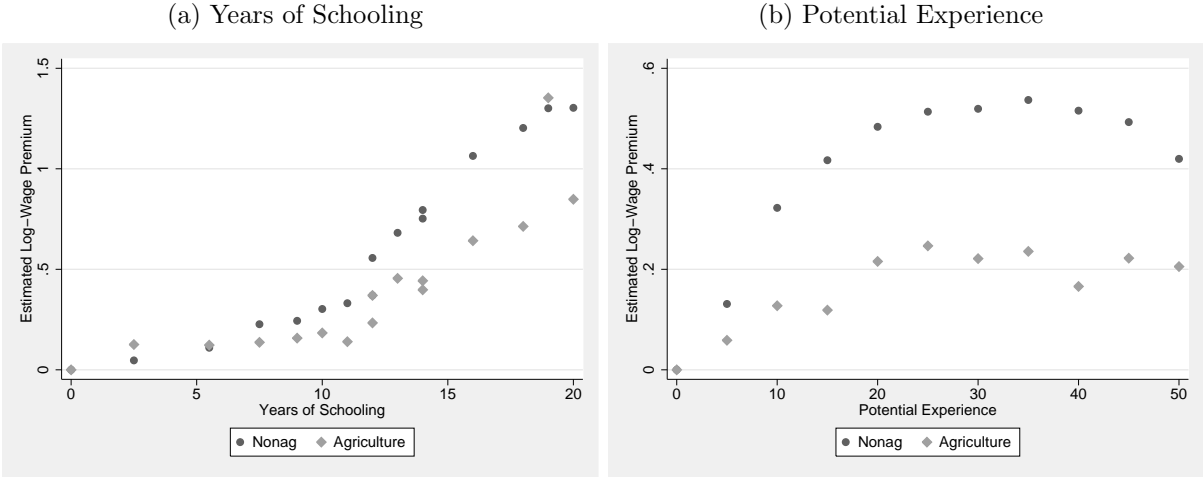
where W_{sj}^{it} is the hourly wage of worker i in year t , state s , and sector j ; Z_{sj}^t is a vector of state-year interaction terms; μ_{dj} captures the average wage by sector, after controlling for other factors; X_{sj}^{it} is a vector of the worker's characteristics, which includes a full set of education dummies, potential experience dummies, and a gender dummy; and β_{dj} measures the market return to these characteristics, which again may vary by sector.

A few remarks are in order. First, running this regression separately by decade and sector allows us to capture differences in the wage structure and in the return to observed factors over time and across sectors. For example, the market return to skills may have risen over time or it may be higher in one sector than in the other. Figure 4 suggests that the latter is indeed the case over the period 1980–2009; the rates of return to schooling in agriculture are lower than in non-agriculture and the potential experience profile is flatter in agriculture. Second, we implement the regression only for workers who meet our sample selection criteria. For the other workers, we use the estimated coefficients of this regression to predict a wage. That is, for workers who are self-employed or work only few hours we impute the wage by assuming that they earn the same wage as other people in their

state, sector, and year with their demographic characteristics. We then take the observed or imputed wages and calculate the (average) wage in each sector and state.

We use the regression coefficients β_{dj} to construct a measure of (average) human capital by sector. We follow the standard Mincer practice and define human capital as the part of wages that can be attributed to $\beta_{dj}X_{sj}^{it}$. The reasoning behind this definition is that human capital corresponds to the differences in workers' innate characteristics, evaluated at the observed market rate of return. In contrast, $\mu_{dj}Z_{sj}^t$ measures the determinants of wages that are independent of workers' characteristics. An example would be if unions in some parts of non-agriculture increased the wage above the competitive wage.

Figure 4: Wage Profiles by Sector, 2000 CPS



Following the procedure explained above, we find that there are relatively large differences between wages and human capital in non-agriculture and agriculture at the state level. Table 5 shows that the summary statistics of the two distributions are surprisingly similar though. We can see that for the states at the median of each distribution, the gaps in nominal wages and human capital are both equal to 1.9. This means that in the median state, an average worker in non-agriculture has almost twice the nominal wage and the human capital of an average worker in agriculture. The gaps at the 90th percentile and the maximum are even larger.⁷ The fact that we find sizeable gaps in human capital across non-agriculture and agriculture for US states is remarkable because in the United States human capital is fairly high in all sectors (even in agriculture most people have high

⁷Note that in Table 5 the states for which a given summary statistic of nominal wages and of human capital are reported may not be the same, implying that the table does not yet establish that gaps in human capital do account for gaps in nominal wages.

Table 5: Gaps in Nominal Wages and Human Capital

| | Nominal Wages | Human Capital |
|-----------------------------|---------------|---------------|
| Median | 1.9 | 1.9 |
| 90 th Percentile | 2.2 | 2.0 |
| Maximum | 2.8 | 2.2 |

school degrees). This is not the case in poor countries where illiterate workers tend to be in agriculture, suggesting that gaps in human capital should be at least as large in poor countries than in the US. The evidence collected by Gollin et al. (2011) suggests that this is the case.

3.2 Consumer price level

The next step in constructing real wages per efficiency unit is to measure the consumer price level in the different locations. We define the consumer price level for a state–sector pair as the average price level that the workers of that state–sector pair pay for their consumption basket. To calculate this price level for each worker, we need to know the sector in which he works and the price level where he lives. The information about the sector and the residence comes from the Population Censuses.⁸ The information about the price level in the residence comes from recent research by Aten (2006), Aten and D’Souza (2008), and Aten (2008), who constructed consumer price levels for 363 metropolitan areas as well as for the rural area of each state for the year 2006.⁹ We combine this information with the residence information to obtain a price level for each worker. We then average across the price levels of the workers of each sector and state so as to obtain the average price level in that sector and state. Appendix C contains a more detailed description of how we calculate the price levels.

We find that across the entire United States the gap in the price level between metropolitan and non–metropolitan areas is 1.4. While this gap is fairly large, a considerable part of it comes from price variation across states. Moreover, for the current purpose, we have to take into account that some agricultural workers live in metropolitan areas (typically

⁸After our prior discussion, one might think that it would be preferable to use CPS data instead of Census for this, but before 1986 the geographic detail in the CPS is too limited and the sample size is too small to produce reliable estimates for smaller metropolitan areas. Since this is not an issue with the Population Censuses, we use them instead.

⁹Note that the CPI is not useful in the current context, because it covers only the major metropolitan areas and is normalized to 100 for every city in the base year, making level comparisons impossible.

Table 6: Gaps in the Consumer Price Level

| | |
|-----------------------------|-----|
| Median | 1.1 |
| 90 th Percentile | 1.2 |
| Maximum | 1.3 |

Table 7: Gaps in Real Wages per Efficiency Unit

| Gaps in | Productivity | Real Wages peu |
|-----------------------------|--------------|----------------|
| Median | 1.9 | 1.0 |
| 90 th Percentile | 3.0 | 1.1 |
| Maximum | 5.7 | 1.4 |
| 90/10 Ratio | 2.3 | 1.4 |
| SD(log(.)) | 0.4 | 0.1 |

smaller ones) and many rural residents do not work in agriculture. When we factor this in, we find that gaps in the consumer price levels between non-agricultural and agricultural workers are sizeable (around 1.3) only in the few states that have large, expensive urban areas: Illinois, New York, Texas, and Virginia. In contrast, in the median state the gap in the consumer price level between non-agriculture and agriculture merely equals 1.1, which is negligible compared to the gaps in nominal wages and human capital that we found in the previous subsection.

3.3 Real wages per efficiency unit

We are now ready to calculate the real wage per efficiency unit as defined in (1). Figure 5 shows the histogram of the resulting gaps in the real wage per efficiency unit between non-agriculture and agriculture. The most noticeable feature of the histogram is that the distribution of the gaps is very closely centered around 1. Table 7 reports the summary statistics; in the second column are the gaps in productivity from Table 4 for comparison and in the third column are the gaps in real wages per efficiency unit. The median gap in real wages per efficiency unit is exactly one whereas the maximum gap is only 1.4. Moreover, the gaps in real wages per efficiency unit are small compared to the gaps in productivity.

In sum, we have found that the gaps of real wages per efficiency unit are small at the level of US states. This finding suggests that there is little scope for beneficial reallocation

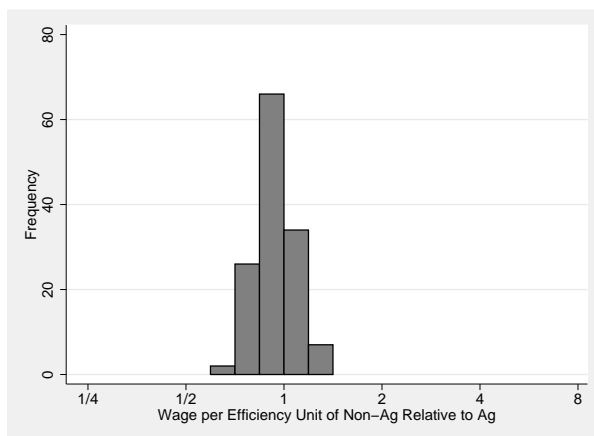


Figure 5: Real Wage per Efficiency Unit Gaps

of workers from agriculture to non-agriculture within US states, which is not surprising in light of the fact that the US economy is relatively undistorted and efficient. What is surprising is that there are only small gaps in real wages per efficiency unit and large gaps in measured productivities. This raises the question as to how these two findings may arise at the same time. In the next section, we attack this question. We begin by deriving several useful accounting identities, which we then connect to the data from US states.

4 Connecting Productivity Gaps with Wage Gaps

4.1 An accounting framework

To establish the link between productivity and wage gaps, we start with a basic identity:

$$W_{sj} = LS_{sj} \frac{Y_{sj}}{L_{sj}}, \quad (2)$$

where, as above, W denotes the nominal wage, $LS \equiv (WL)/Y$ denotes the labor share and Y/L is the nominal value added per hour.¹⁰ In words, identity (2) says that the nominal wage in location sj equals the labor share times labor productivity in that location. From the definition of the real wage per efficiency unit, the left-hand side of identity (2) also satisfies:

$$W_{sj} = w_{sj} h_{sj} P_{sj}.$$

¹⁰As always, nominal means in current dollars. Note that with Cobb–Douglas production functions and competition, (2) is also the first-order condition for the choice of labor. The reason for this is that for Cobb–Douglas production functions average product equals marginal product.

Combining the last two equations, we obtain a new identity that links labor productivity and real wages per efficiency unit:

$$\frac{Y_{sj}}{L_{sj}} = w_{sj} \frac{h_{sj} P_{sj}}{LS_{sj}}. \quad (3)$$

Identity (3) says that labor productivity equals the real wage per efficiency unit times a factor that depends on human capital, the consumer price level, and the labor share. To build some intuition for how these terms enter, consider two sectors that pay the same real wage per efficiency unit. Then labor productivity is higher in a sector if human capital is higher or the consumer price level is higher. This is fairly intuitive, because given that both sectors pay the same real wage per efficiency unit, the nominal wage is going to reflect the higher human capital or a higher consumer price. Moreover, labor productivity is higher in a sector if it has a lower labor share than the other sector. The reason for this is that a sector with a lower labor share needs a higher value added per hour to be able to pay the same real wage per efficiency unit.

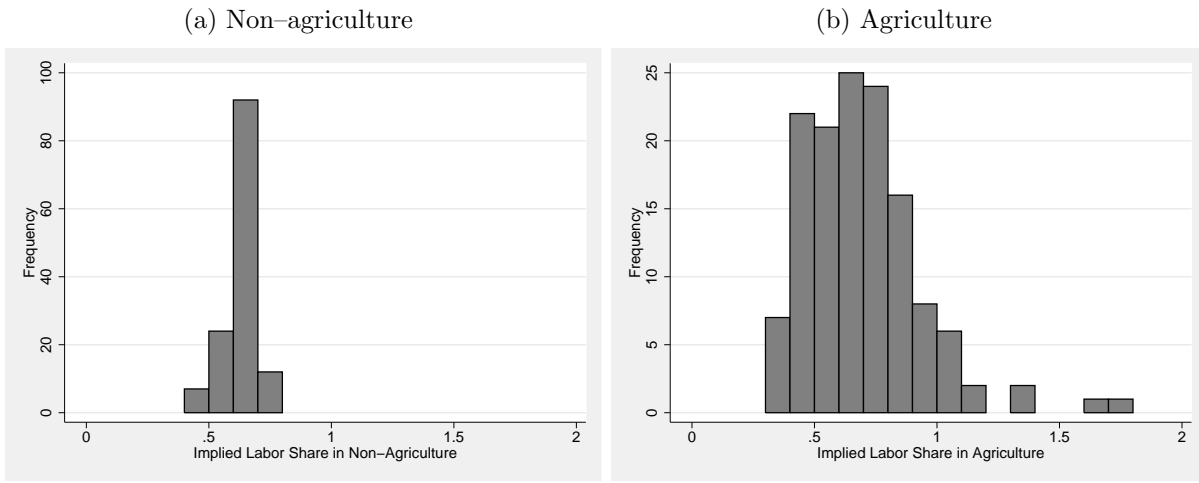
One implication of identity (3) is that if w , h , P , and LS are the same in the two locations of a given state, then labor productivities are the same there too. In this case, there are no productivity gaps. Conversely, if w , h , P , or LS differ across the two locations of a give state, then labor productivities differ and productivity gaps arise. This can be highlighted by dividing (3) for non-agriculture and agriculture and rearranging, which results in a third identity that links gaps in productivity to gaps in real wages per efficiency unit:

$$PG_s \equiv \frac{Y_{sn}/L_{sn}}{Y_{sa}/L_{sa}} = \frac{w_{sn}}{w_{sa}} \times \frac{h_{sn}}{h_{sa}} \times \frac{P_{sn}}{P_{sa}} \times \left(\frac{LS_{sn}}{LS_{sa}} \right)^{-1}. \quad (4)$$

This identity shows that there are three reasons why large productivity gaps may arise even when there are only small gaps in real wages per efficiency unit: sectoral differences in human capital, the consumer price level, and the labor share.

An important lesson to take away from the previous discussion is that there is no reason to view $PG_s = 1$ as a natural benchmark case that ought to hold in a developed economy such as the US. Instead, as we have seen, even if a similar real wage per efficiency unit is paid in the two sectors of a given state, then the sector with the higher human capital and the higher costs of living needs to pay a higher nominal wage. Moreover, the sector with the lower labor share needs to have higher productivity in order to be able to pay the common real wage per efficiency unit.

Figure 6: Implied Labor Shares



4.2 Implied labor shares

Now, we are ready to ask whether empirically our previous findings that there are large gaps in productivity and small gaps in real wages per efficiency unit are consistent with each other. We will establish that they are not. The easiest way of doing this is by using identity (2) from the previous subsection. We observe W and Y/L for each location, and so the identity implies a labor share for each location. Figures 6 plot these implied labor shares for non-agriculture and agriculture. It turns out that while they are very concentrated and within the plausible range in non-agriculture, they are all over the map in agriculture. More specifically, in the median state, the implied labor share in non-agriculture is 64%, which is a standard value. In contrast, in agriculture around 10% of the implied labor shares are larger than one, which obviously is impossible, and in the median state the implied labor is 63%. This is implausibly large compared to standard estimates. For example, according to BEA data the average labor share during 1980–2009 was 44%.¹¹

The fact that the implied labor shares in agriculture are implausibly large suggests that either measured wages in agriculture are too large or measured productivity in agriculture is too small, or both. Since wages are calculated with micro data from the CPS, we have considerable confidence in them. Moreover, we have found small gaps in real wages per efficiency unit which is what one would expect in the relatively undistorted US economy. This points to measurement of the productivity in agriculture as the potential problem. In the

¹¹To calculate this, we use the methodology of Gollin (2002), that is, we calculate the labor share in agricultural value minus proprietors' income and then assume that the labor share in proprietors' income is the same as in value added without proprietors' income.

next section, we will establish that in fact the BEA seriously under-estimates agricultural productivity.

5 Mis-measurement of Agricultural Productivity

To establish that agricultural productivity is mis-measured, we have to establish that either the numerator or the denominator is mis-measured. We start with the denominator. To assess whether it is mis-measured, we compare hours worked in the CPS with those in American Time Use Survey (ATUS). ATUS data are collected from time use diaries in which the participants record their activities by minute for a 24-hour period. ATUS data are deemed very reliable. We focus on the period 2003–2010 during which the ATUS and the CPS overlap. Given that ATUS does not have many observations in agriculture, we only calculate hours at the national level. We find that there are discrepancies between ATUS and CPS estimates of hours worked. However, the size of the discrepancy is roughly the same in the two sectors; while ATUS hours are 9.5% lower than CPS hours in agriculture, they are 8.4% lower non-agriculture.¹² This suggests that mis-measurement of hours worked is not an important factor behind the productivity gaps that we calculated above. In the following subsections, we will establish that instead mis-measurement of agricultural value added is the reason for the large measured productivity gaps.

5.1 The BEA's notion of farming

Above, we have used farm value added from the BEA's regional accounts. Although these data underlie the construction of NIPA, it turns out that they have two shortcomings when it comes to measuring agricultural value added at the state level. First, the BEA's agricultural value added at the state level does not include subsidies and taxes. This is potentially important for our findings because agriculture receives higher than average subsidies.¹³

¹²Note that the fact that ATUS hours are smaller than CPS hours is expected because ATUS uses a strict definition of time spent at work. For example, ATUS does not count time spent at business meals or commuting, while respondents in the CPS may implicitly include such time in their responses.

¹³In its various documentations, the BEA is not very explicit about whether or not it includes net subsidies in sectoral value added. We established in two ways that it does not do this at the state level. First, we personally checked with people in the BEA who are familiar with the relevant NIPA procedures. Second, we compared the BEA numbers for agricultural value added at the state level with value added without subsidies that we constructed from the State-Net-Value-Added Accounts provided by the USDA. We found that the two numbers are very similar.

The second shortcoming is that the BEA follows the System of National Accounts (SNA) and views agricultural value added as the value added that is produced by *farmers*, i.e., persons who operate a farm or who are employed on a farm, instead of the value added that is produced on *farms*. To appreciate the significance of using farmers instead of farms, an example may be helpful. Consider the payments that are received by farm contractors or the rental payments that are received by land owners who are not farmers. Clearly, these are factor payments that are generated on farms, and conceptually they belong to the value added produced on farms. However, the BEA does not report them as part of agricultural value added because they do not lead to income of farmers. Instead it reports the payments to farm contractors as value added in agricultural services and the payments to land owners as value added in real estate.

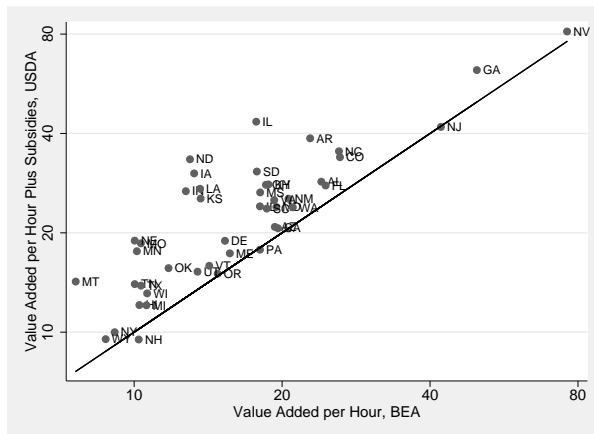


Figure 7: Agricultural Value Added in 2000 from Different Data Sources

To see how much these shortcomings matter for measuring productivity gaps, we construct a new measure of value added that includes subsidies and all factor payments generated on farms irrespective of who they accrue to. To this end, we use the State–Net–Value–Added Accounts provided by the USDA, which is the original data source of the BEA. Since these accounts include a detailed breakdown of the receipts, expenses, and factor payments in the farm sector at the state level, they have sufficient information to make the required adjustments. Note that 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). The Appendix describes in detail how we use this information to calculate the value added produced on farms. We obtain non–agricultural value added as the difference between state GDP and state agricultural value added including subsidies.

Table 8: Productivity Gaps with Different Value Added in 2000

| | BEA | USDA | |
|-----------------------------|-----|-------------------|----------------|
| | | without net subs. | with net subs. |
| Median | 1.9 | 1.6 | 1.6 |
| 90 th Percentile | 3.0 | 2.8 | 2.7 |
| Maximum | 5.7 | 5.7 | 5.8 |

Figure 7 plots on the x-axis the agricultural value added from the BEA, which we used above, and on the y-axis the agricultural value from the USDA, which includes all factor payments and subsidies. All value added numbers in the figure are for the year 2000 and per CPS hour worked. As expected, most observations are above the 45 degree line, that is, value added per hour worked based on the USDA is larger in all states than that based on the BEA. Moreover, for states with low agricultural productivity the difference can be large quantitatively. This suggests that using the BEA numbers for value added led to a downward bias in the calculation of productivity in agriculture, and thereby inflated the measured productivity gaps between non-agriculture and agriculture. Table 8 reports the summary statistics for productivity gaps based on BEA and USDA value added. As expected, the productivity gaps shrink when we replace BEA value added with USDA value added. Somewhat surprisingly, however, it does not matter much whether or not we include subsidies and taxes. The explanation lies in the fact that the vast majority of US farm subsidies go to large farms in states with relatively productive agriculture. In contrast, agriculture in relatively unproductive states actually pays more in taxes than it receives in subsidies so that the net transfers from the government are negative. This explains why the maximum gap with net subsidies is larger than without net subsidies.

In sum, using USDA value added closes part of the productivity gaps that we have found using BEA value added, but the remaining gaps are still sizeable. Since BEA value added is similar to the standard data sources that are available for developing countries, this suggests that part of the large productivity gaps that people find for developing countries may be the result of mis-measurement. This is particularly true with respect to agricultural value added that is not generated by farmers, which is counted in non-agriculture in all countries that adopt the SNA. In the next subsection, we provide further evidence of mis-measurement of agricultural value added, which arises because proprietors tend to under-report their income severely.

Table 9: Actual divided by Reported Proprietors' Income

| | Non-farm | Farm |
|-------|-----------|-----------|
| 1980s | 1.4 – 1.5 | 1.4 – 1.5 |
| 2001 | 2.3 | 3.6 |

5.2 Under-reporting of proprietors' income

The Internal Revenue Service (IRS) periodically conducts tax audits to assess the degree of tax compliance. These audits find that proprietors severely under-report their income. Table 9 lists the ratio of actual proprietors' income (as determined by tax audits) to reported proprietors' income for the 1980s and for 2001. The numbers are based on Internal Revenue Service (1996) and Internal Revenue Service (2007), respectively. They range from 1.4 to 3.6, suggesting that proprietors under-report on a massive scale. Moreover, the degree of under-reporting appears to be more severe in farming than in non-farming. This is likely to be a serious issue for calculations of agricultural productivity because proprietorships are very common in farming.

The BEA reacts to the evidence of under-reporting by adjusting reported proprietors' income in non-agriculture, but not in agriculture. The reason for this discrepancy likely stems from the method the BEA uses to construct value added. In the non-agricultural sector it uses the income approach. When it adds in the income of non-farm proprietors it draws on the evidence from the IRS and adjusts the reported income, roughly doubling it. In the agricultural sector it uses the production approach. Here it takes farm revenues and subtracts farm expenses. Since the production method does not directly add the income of proprietors, it is not obvious that the BEA should make any adjustment for the under-reporting of farm proprietors' income.¹⁴

Nonetheless, our estimates indicate that the BEA should adjust for the under-reporting of farm proprietors' income. The reason is that the value added of farm proprietors as calculated under the production approach only modestly exceeds the reported income of farm proprietors. We calculate each of these objects for 2002 and 2007 using the Census of Agriculture (to calculate value added via the production approach) and the March supplement to the Current Population Survey (to calculate farm proprietors' income). Table 10 reports the national figures for the production approach divided by farm proprietors' income,

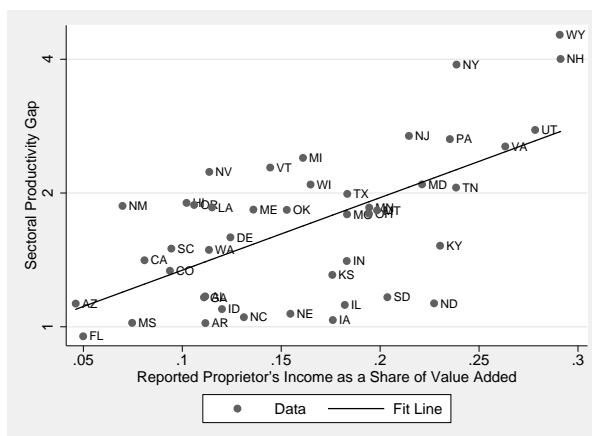
¹⁴The preceding paragraph draws on Bureau of Economic Analysis (2009) for details on how the BEA actually constructs state-industry value added.

Table 10: Revenues minus Expenses divided by Proprietors' Income

| | |
|------|-----|
| 2002 | 1.1 |
| 2007 | 1.2 |

showing that using revenues minus expenses does not solve the problem of under-reporting; while it is true that using the production method results in a figure that is 10 – 20% higher than reported farm proprietors' income, the IRS estimates indicate that a much larger correction is warranted.

Figure 8: Proprietors' Share in Agricultural Value Added versus Productivity Gaps (averages over 1980–2009)



If under-reporting of proprietors' income in agriculture is part of the reason why we have such large productivity gaps in some states, then states with a larger proprietors' income share in agricultural value added should have larger productivity gaps. Figure 8 shows that this is indeed the case (the regression coefficient is significantly different from zero at the 99% significance level). This suggests that under-reporting of proprietors' income should account for a sizeable part of the large measured productivity gaps. It is somewhat hard to be more specific though. To begin with, over the years the IRS studies found fairly different degrees of under-reporting. Moreover, we have no information on whether the degree of under-reporting in agriculture is uniform or varies by state. This is an issue to the extent that the degree of under-reporting varies with the type of agriculture which differs considerably across states. For example, while coastal states like California and Florida specialize in fruit and vegetable production, the Great Plains specialize in grain production, and Texas and the Western states specialize in animal production. Since these

types of agriculture have different production structures (e.g., they have different degrees of mechanization and reliance on intermediate inputs), the scope for under-reporting of proprietors' income may well differ also.

Given the previous discussion, we do not take a firm stand on the degree by which proprietors under-report their income in agriculture, but experiment with the range of possibilities (40 – 260%) under the assumption that the adjustment factor is the same in all states. This means that our results on the importance of under-reporting of proprietors' income should be viewed as suggestive rather than definitive. They are listed in Table 11. The second column reports the productivity gaps from Table 8 for comparability. Columns 3–5 report the productivity gaps for three different adjustment factors from Table 9: 1.5 is the average over the range 1.4 – 1.5, whereas 2.3 and 3.6 are the suggested adjustment factors for non-farming and farming in 2001.

We focus our attention on column 4 for three reasons. First, it adjusts farm proprietors' income by a factor of 2.3, which is roughly the same adjustment that the BEA uses to adjust non-farm proprietors' income. Intuitively, column 4 asks how far we can get simply by performing the BEA's adjustment also in the farm sector. The second reason we focus on this column is that it implies an aggregate labor share of 0.40. This labor share is close to the value implied of 0.44 implied by BEA data for the period 1980–2009.

The last reason why we focus on column 4 is that it resolves the entire productivity gap puzzle, at least for the median state. To see why, it is important to recall that a labor productivity gap of 1 is *not* the natural benchmark. Instead, we recall identity (2), which implies:

$$PG_s = \frac{W_{sn}}{W_{sa}} \times \left(\frac{LS_{sn}}{LS_{sa}} \right)^{-1}. \quad (5)$$

Productivity gaps should be consistent with observed wage gaps and labor share gaps. Indeed, we will now establish that for column 4 this is the case. In section 3 we found that the wage gap between agriculture and non-agriculture was 1.9 in the median state, explained almost entirely by a gap in human capital between the two sectors. The gap in labor shares in the aggregate is $0.63/0.44 = 1.43$. If we apply this labor share gap to the median state, then we find that the natural benchmark for productivity gaps in the median state is $1.9 \times (1.43)^{-1} = 1.3$. But this is exactly the productivity gap that we do find for the median state in column 4. What this tells us is that simply applying the BEA's correction for non-farm proprietors' income to farm proprietors' income leads us to measured productivity gaps that are perfectly consistent with observed wage and labor

Table 11: Corrected Productivity Gaps 1980–2009

| | USDA with Subs. | Adj. for Underrep. | | |
|-----------------------------|--------------------|--------------------|-------------|------|
| | | 1.5 | 2.3 | 3.6 |
| Median | 1.6 | 1.5 | 1.3 | 1.1 |
| 90 th Percentile | 2.7 | 2.5 | 2.0 | 1.7 |
| Maximum | 5.8 | 4.8 | 3.6 | 2.9 |
| Impl. Agg. Lab. Share | 0.49 | 0.46 | 0.40 | 0.34 |

share gaps.

In sum, we have established that there are three reasons why the BEA under-estimates agricultural value added at the level of US states: the SNA classifies some of the value added generated on farms in other industries; the BEA does not include subsidies in farm value added at the state level; the BEA does not properly adjust for under-reporting of proprietors' income in agriculture. This implies that the productivity gaps between non-agriculture and agriculture that are based on BEA value added data are artificially large. Caselli and Coleman (1998), which is the unpublished appendix to Caselli and Coleman (2001), reports a related fact: At the US national level during 1940–1990, nominal wages from the Population Census imply a larger reduction in the wage gap between non-agriculture and agriculture than nominal wages from NIPA. Moreover, it turns out that nominal wages in non-agriculture from both data sources are similar and that the difference is due to the behavior of the wages in agriculture. The authors argue that the explanation is that NIPA under-estimates agricultural wages. This is consistent with our finding that NIPA under-measures agricultural value added during the later period 1980–2009 that we study here.¹⁵

The findings of this section are relevant to the development literature to the extent that similar measurement problems are likely to be present also in poor countries. While the rules for obtaining net subsidies are particular to the US (and subsidies turn out not to matter quantitatively anyways), the other two measurement issues are potentially important for poor countries too. In particular, the procedures of the SNA are followed by many countries and under-reporting of proprietors' income is a widespread problem. Moreover, if anything, tax evasion and proprietorships are more common in poor countries than in rich countries.

¹⁵Note that the wage gaps that Caselli and Coleman report are not strictly comparable with our wage gaps. The reason for this is that they compute earnings per worker whereas we compute wages per hour. Since agricultural worker tend to work more hours, our labor productivity in agriculture is smaller and our wage gaps are larger than theirs.

To the extent that poor countries do not make adjustments either for under-reporting of agricultural value added by proprietors, this may explain in large part why they appear to be so unproductive in agriculture and why authors like Gollin et al. (2011) find that standard explanatory factors account only for part of the sectoral productivity difference.

6 Conclusion

We have studied productivity in agriculture and non-agriculture in US states during 1980–2009. We have found that: (i) there are large measured gaps in productivity between the two sectors, whose size depends critically on which data source we use for sectoral employment; (ii) there are sizeable gaps in human capital, which account for part of the productivity gaps; (iii) there are hardly any gaps in real wages per efficiency unit; (iv) the gaps in productivity and real wages are inconsistent with each other. We have provided evidence that the gaps in productivity are mis-measured. We have shown that making appropriate corrections to address the mis-measurement reduces the productivity gaps considerably, and that the corrected productivity gaps are consistent with the gaps in real wages per efficiency unit at least for the median state.

Our results have several important implications for the development literature that seeks to account for large productivity gaps between non-agriculture and agriculture in poor countries. First, the fact that measured productivity gaps change considerably depending on whether we measure labor input as bodies from the Population Census or the BEA, or as hours from the CPS suggests to be cautious about employment numbers from developing countries, which are not typically based on hours worked from high quality data sets such as the CPS. Second, the fact that we find that at the level of US states, there are large differences in human capital between non-agriculture and agriculture suggests that this is a promising candidate explanation for large productivity gaps also in poor countries. Third, the fact that we find large measured productivity gaps at the same time as which real wages per efficiency unit are similar across sectors provides an example where large measured productivity gaps do not imply that there is scope for beneficial reallocation which is prevented by large moving costs (some sort of “barriers”). Fourth, the fact that even the BEA does not measure agricultural productivity precisely suggests that there are likely to be severe measurement problems also for poor countries, and that more evidence from poor countries is required to establish convincingly that indeed the productivity gaps between non-agriculture and agriculture are as large as they appear.

References

- Adamopoulos, Tasso and Diego Restuccia**, “The Size Distribution of Farms and International Productivity Differences,” 2011. Manuscript, University of Toronto.
- Aten, Bettina H.**, “Interarea Price Levels: an Experimental Methodology,” *Monthly Labor Review*, 2006, 129 (9), 47–61.
- , “Estimates of State and Metropolitan Price Parities for Consumption Goods and Services in the United States, 2005,” October 2008. Manuscript, Bureau of Labor Statistics.
- and **Roger J. D’Souza**, “Regional Price Parities: Comparing Price Level Differences Across Geographic Areas,” *Survey of Current Business*, 2008, 88 (11), 64–74.
- Bureau of Economic Analysis**, “State Area Personal Income and Employment Methodology,” Technical Report October 2009.
- Caselli, Francesco**, “Accounting for Cross–Country Income Differences,” in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 1A, Elsevier, 2005, chapter 9, pp. 679–741.
- and **Wilbur John Coleman II**, “How Regions Converge: Alternative Data Sources,” June 1998. Manuscript.
- and —, “The U.S. Structural Transformation and Regional Convergence: A Reinterpretation,” *Journal of Political Economy*, 2001, 109 (3), 584–616.
- Cociuba, Simona, Edward C. Prescott, and Alexander Ueberfeldt**, “U.S. Hours and Productivity Behavior Using CPS Hours Worked Data: 1947-III to 2009-III,” November 2009. Manuscript, Arizona State University.
- Donovan, Kevin**, “Risk, Farm Ownership, and International Productivity Differences,” 2011. Manuscript, Arizona State University.
- Gollin, Doug, David Lagakos, and Mike Waugh**, “The Agricultural Productivity Gap in Developing Countries,” 2011. Manuscript, Arizona State University.
- , **Stephen L. Parente, and Richard Rogerson**, “Farm Work, Home Work and International Productivity Differences,” *Review of Economic Dynamics*, 2004, 7, 827–850.

- Gollin, Douglas**, “Getting Income Shares Right,” *Journal of Political Economy*, April 2002, 110 (2), 458–474.
- Herrendorf, Berthold and Akos Valentinyi**, “Which Sectors Make Poor Countries So Unproductive?,” *forthcoming: Journal of the European Economic Association*, 2011.
- and **Arlton Teixeira**, “Barriers to Entry and Development,” *Forthcoming: International Economic Review*, 2009.
- Internal Revenue Service**, “Federal Tax Compliance Research: Individual Income Tax Gap Estimates for 1985, 1988, and 1992,” Publication 1415 (Rev. 4-96), Washington, DC 1996.
- , “Reducing the Federal Tax Gap: A Report in Improving Voluntary Compliance,” Technical Report, Washington, DC 2007.
- Lagakos, David and Mike Waugh**, “Specialization, Agriculture, and Cross-Country Productivity Differences,” 2010. Manuscript, Arizona State University.
- Restuccia, Diego, Dennis T Yang, and Xiaodong Zhu**, “Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis,” *Journal of Monetary Economics*, March 2008, 55 (2), 234–250.
- Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek**, “Integrated Public Use Microdata Series: Version 5.0,” Minneapolis: University of Minnesota 2010.

A Data Appendix: Sectoral Value Added

In this part of the Appendix, we explain in more detail how we construct our measures of agricultural value added at the state level. Given these measures, it is straightforward to obtain measures of nonagricultural value added by subtracting agricultural value added from state GDP as reported by the BEA.

A.1 BEA

The BEA numbers for sectoral value added are taken straight from the BEA's regional economic accounts. In particular, for value added in agriculture, we use item 10010 (value added of farms) for years with the SIC classification and item 4 (value added of crop and animal production) for years with the NAICS classification; for GDP at the state level we use item 0 in the SIC and item 1 in the NAICS, minus value added in the military (item 112000 in the SIC and item 80 in the NAICS).

A.2 USDA

To construct a measure of agricultural value added from USDA data, we use the USDA's value-added spreadsheets at the state level.¹⁶ We construct income produced on farms as follows:

- The value of crop production is farm income. The USDA reports values for eight types of crops, as well as total values for home consumption and inventory adjustment.
- The value of livestock production is farm income. The USDA reports values for four types of livestock, as well as values for home consumption and inventory adjustment.¹⁷
- Revenues produced from miscellaneous farm activities may or may not be counted as farm income. Considering each in turn:
 - The value of machine hire and customwork is farm income, because it includes payments for providing services closely related to the farm. Examples are planting, plowing, spraying, or harvesting for others.
 - The value of forest products sold from the farm is farm income. Ideally we would exclude this revenue from agriculture and include it in forestry. However,

¹⁶The spreadsheets are available at http://www.ers.usda.gov/Data/FarmIncome/Zip_filesXls.htm.

¹⁷Note that NAICS uses animal production and USDA uses livestock production for the same industry.

some of the expenses of farms and some of the labor on farms are devoted to generating this revenue. Since we cannot isolate the relevant expenses and labor, we include the revenue as farm income.

- Other income is farm income, because it is closely related to farm operations. Examples include animal boarding, breeding fees, and energy generated on the farm.
- The gross imputed rental value of farm dwellings is not farm income, because it is not closely related to farm operations.

We construct expenses for intermediate inputs used by farms as follows:

- Farm–origin expenses are farm expenses. The USDA reports feed purchased, livestock and poultry purchased, and seed purchased in this category.
- Manufactured inputs are farm expenses. The USDA includes fertilizers and lime, pesticides, petroleum fuel and oils, and electricity in this category.
- Other purchased inputs may be farm expenses or factor payments, in which case they are not counted as farm expenses. In particular:
 - Repair and maintenance of capital items are farm expenses, in line with usual NIPA procedures.
 - Expenses for machine hire and custom work are farm expenses.
 - Marketing, storage, and transportation expenses are farm expenses.
 - Contract labor is a factor payment to contractors or crews that provide labor to farms, and so is counted as a factor payment and not as farm expenses.
 - Miscellaneous expenses are farm expenses. Examples include the costs of animal health care and insurance.

In addition to a “raw” value added measure, we also construct a value added measure that includes subsidies and property taxes. To do so we take value added and add the line “net government transactions”. Note that the government also provides indirect support for farmers through price supports and similar programs; the effects of these indirect supports are already counted in value added.

B Data Appendix: Sectoral Labor Inputs

This section provides the details of how we construct sectoral labor inputs. It also explains how we define agricultural and nonagricultural workers, hours worked, wages, and so on.

B.1 Quarterly Census of Employment and Wages

The BEA reports employment figures for each state and industry. These figures are drawn from the Quarterly Census of Employment and Wages. The BEA uses the same industry classification scheme for both value added and employment. Hence, we categorize workers into agriculture or non-agriculture in the same manner as with value added; see Appendix A.1.

B.2 Population Census

To calculate labor input from the US Population Census, we use the public-use census data made available through the IPUMS data service, (Ruggles et al. 2010). We use the 5 percent sample for 1980, 1990, and 2000, which is the largest publicly available sample for these years. We impose as little sample selection as possible. We require that workers be in the labor force and employed ($\text{empstat} = 1$). The Census asks about the employment status of those aged 16 and older, so we restrict our sample to this age group. We exclude workers with invalid or missing industry or occupation codes.

To assign workers to agriculture and non-agriculture, we need to construct a crosswalk from workers' self-reported industry/occupation to our agriculture/nonagriculture classification. One potential complication is that the US 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 (e.g., "occ" and "ind" rather than "occ1950" and "ind1950").

Our primary form of classification draws on which industry workers report. Using the reported industries, we construct the labor force in the farm sector, i.e., the animal and crop production industries. Table 12 gives the full crosswalk. It lists for each year all industries (and corresponding industry codes) that we use. We construct non-agriculture as the residual, that is, non-agricultural workers are all workers with valid industry reports that do not work in an agricultural industry or in the military.

As a robustness check on our results, we also experiment with two alternative methods

Table 12: Coding Census Industries to the Agriculture Sector

| Year | Narrow | Broad |
|------|--|---|
| | | = Narrow, Plus: |
| 1980 | Agricultural Production, Crops (010) Agricultural Production, Livestock (011) | Agricultural Services, except Horticultural (020) |
| 1990 | Agricultural Production, Crops (010) Agricultural Production, Livestock (011) | Agricultural Services, n.e.c. (030) |
| 2000 | Crop Production (017) Animal Production (018) | Support Activities for Agriculture and Forestry (029) |

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

Table 13: Coding Census Occupations to Agriculture Sector

| Year | Agriculture |
|------|--|
| 1980 | Farmers, Except Horticultural (473) Managers, Farms, Except Horticultural (475) |
| | Supervisors, Farm Workers (477) Farm Workers (479) |
| 1990 | Supervisors, Related Agricultural Occupations (485) Graders and Sorters, Agricultural Products (488) |
| | 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) |
| 2000 | Graders and Sorters, Agricultural Products (488) |
| | 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.

of defining workers in the agriculture sector. In the first, we still use workers' reported industry codes, but we take a somewhat broader view of which industries should be counted as agriculture. In particular, we include industries identified as agricultural services or support activities for agriculture. The codes are reported in the third column of Table 12. In the second alternative, we use workers' reported occupations to identify agricultural and non-agricultural workers. 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 industries. Again, the Census includes a measure of occupation that varies with each Census and generally becomes more detailed over time. Table 13 gives the occupation titles and corresponding codes that we associate with the agriculture sector for each Census year. In general both of these methods lead to similar or higher employment figures in agriculture, which in turn implies lower value added per worker in agriculture and larger sectoral productivity gaps. Results are available upon request.

After dividing the population into agricultural and non-agricultural workers, we calculate employment and hours worked by state and sector. For all calculations, we restrict the samples to individuals with valid responses. We use the reported state of residence (statefip) and weight all variables with individual weights (perwt). We compute sectoral employment as the number of workers in each sector.

B.3 Current Population Survey

The Current Population Survey (CPS) administered by the BLS is our principal data source for the number of workers and hours worked and the wages and earnings received.¹⁸ We restrict our attention to those workers in the CPS who have a job and valid occupation and industry codes. We use information about age, education, employment, gender, and state of residence from the CPS Basic Monthly Data, which we take from the NBER's CPS data repository. We also use information about hours worked in primary and secondary jobs by industry, which is available in the May supplements of the 1979, 1980, 1989, and 1991 CPS (again taken from the NBER's CPS data repository) and in the Outgoing Rotation Groups during 1994–2009. Lastly, we use information about wages and earnings from the NBER Matched Outgoing Rotation Groups (MORG) during 1980–1993 and the Outgoing Rotation Groups during 1994–2009.

To measure the number of workers, we use the CPS Basic Monthly Data. The CPS uses the coding schemes from the Population Censuses for both first and second jobs. The codes

¹⁸Data are available at http://www.nber.org/data/cps_basic.html.

for agriculture are 017 during 1980–1982, 010–011 during 1983–2002, and 0170–0180 during 2003–2009. We assign each worker to the sector of his primary job, which the CPS defines as the job with most hours worked. We then count the total number of hours worked by this worker in all jobs, regardless of the sector. We cap total hours worked at 99 (consistent with earlier CPS procedures and a reasonable limit on the work week), and weight using the provided weight (`pwsswgt`). This gives us a measure of hours worked by primary workers in each sector, state, month, and year. We multiply this figure by 4.33 to generate monthly hours worked.

This measure of hours worked does not account for the time allocation of workers who have secondary jobs in a different sector than their primary job. In other words, these data are analogous to what we would find in the US Population Census, which does not distinguish between primary and secondary jobs. To account for secondary jobs in a different sector, we draw on data on primary and secondary jobs from the CPS from 1994 onward, and from the May supplements of the 1979, 1980, 1989, and 1991 CPS. We use this information in the following way:

1. For the months where the data are available, we calculate the fraction of the total hours worked that is devoted to farming by workers whose first jobs are in farming and non-farming. We aggregate this information to the state-month-year level.
2. We use a regression with state and month fixed effects as well as a linear time trend to predict the time allocation of the workers with a primary job in farming and non-farming for years in which the data are not available.
3. We combine our information on the hours worked by those with primary jobs in farming and non-farming with our predicted time allocations of hours to calculate labor in farming and non-farming.¹⁹

We measure wages per hour using CPS data to be consistent with our other work. We perform wage regressions using the wage data from the NBER Matched Outgoing Rotation Group (MORG) during 1980–1993 and from the outgoing rotation groups in the CPS Basic Monthly Data during 1994–2009. We again use the outgoing rotation group weights and we multiply top-coded wages by 1.4 as suggested by the CPS. As is standard, we run the wage regressions for a selected sample of workers who meet the following criteria:

¹⁹Note that this means that we use the predicted hours also when we actually have hours in the CPS. We do this for logical consistency and to smooth the data in states with small farming samples.

- They work for wages and salaries (the reason for this restriction is that the reported wages of self-employed or unpaid workers are considered unreliable, and are usually not even collected).
- They have a valid hourly wage, that is, they either report a positive hourly wage or a positive weekly wage and provide a positive estimate of their usual weekly hours worked; in the latter case we compute the hourly wage as weekly wage/hours per week.
- They are strongly attached to the labor market, which is measured as working at least 30 hours per week.
- They have between 0 and 50 years of potential experience, which is defined as age minus years of schooling minus 6.

We use the following controls in our wage regression: the state of residence, gender, potential experience, and education. We transform potential experience into 5-year bins (0–4 years, 5–9 years, and so on) and run wage regressions with dummies. Education data are straightforward except that there is a shift in the coding scheme for education in the middle of this period. Until 1991 the scheme counted years in school (such as four years of college), while from 1992 onward it measured degree attainment (such as bachelor’s degree). We run wage regressions with dummies for years before 1991 and with dummies for degree from 1992 onwards.

One final issue to address is that monthly CPS wage data do not include all the compensation that labor receives. First, they do not include irregular compensation, such as bonuses. This type of income is better captured in retrospective questions, such as the question in the March CPS supplement on total labor income earned in the last year. We construct the ratio of retrospective March CPS income to the sum of all monthly income in the prior year, and use this ratio as a correction for irregular compensation. The second component of labor compensation missing is benefits. Fortunately, NIPA includes information on wages and on total compensation (wages plus benefits) by state, year, and industry. We correct CPS wages by the ratio of NIPA total compensation to NIPA wages to adjust for benefits. This correction is generally slightly larger for the non-agricultural sector, indicating that benefits are more generous there. All our figures on wages in the paper include both of these corrections so that they represent total labor compensation.

C Data Appendix: Proprietors' Income

C.1 Census of Agriculture

We calculate revenues minus costs for farm proprietors using the 2002 and 2007 Censuses of Agriculture. The Census of Agriculture distinguishes between family or individual farms, partnerships, corporations, and other farms (which includes cooperatives, trusts, institutional farms, and other unusual arrangements). We calculate farm proprietor income as farm income less farm expenses and factor payments for family or individual farms and partnerships. In theory, the result should be identical to farm income from the March CPS, which measures the income from owning and operating one's own farm, unless one is incorporated. In practice, it is quite close.

We construct income produced on farms as follows:

- The value of total sales is farm income. This includes the sales of crops and livestock of many types.
- The value of government payments is farm income. This category captures subsidies from the federal government.
- The revenue from several services is farm income:
 - The value of machine hire and customwork is farm income, because it includes payments for providing services closely related to the farm. Examples are planting, plowing, spraying, or harvesting for others.
 - The sale of forest products is farm income, for the same reasons as it is counted in the USDA.
 - Patronage dividends and refunds from cooperatives are farm income, which occur when farmers paid for inputs through cooperatives but price realizations come out lower than expected.
 - Agri-tourism payments, insurance payments, and payments from state and local government programs are farm income.
 - Other farm-related income is farm income.
- We do not count one other type of farm income, namely gross cash rent or share payments, which should be reported as a factor payment elsewhere.

We construct expenses for intermediate inputs used by farms as follows:

- The purchases of products produced in the farm sector (including seeds, livestock, and feed) are farm expenses.
- The purchases of manufactured inputs (including fertilizer, lime, chemicals, gasoline and other fuels, and utilities) are farm expenses.
- We count repair and maintenance of capital items, machine hire and custom work, and miscellaneous expenses as farm expenses.
- We do not count as farm expenses several items that are factor payments, such as rental payments for land and machinery; payments to landlords; interest payments; and payments to workers.

Finally, we take the measure of property taxes directly from the line “property taxes paid”. Our measure of revenues minus costs of proprietors then is farm income minus farm expenses, farm factor payments, and property taxes paid.

C.2 March supplement of the CPS

The March supplement to the Current Population Survey asks about the pre-tax income of farmers who own and operate their own farm or are self-employed on their own farm. It does not contain information on the income of those who work as salary or wage employees, including those who work on their own incorporated farm; this information is collected elsewhere. Moreover, it does not contain information on the earnings of owner non-operators, such as those with shares in a farm corporation. Thus the pre-tax income of farmer from the March supplements corresponds closely to the figure computed from the Census of Agriculture above.

We download the relevant data from the IPUMS data service, (Ruggles et al. 2010). The relevant variable is `incfarm`. We add this income to the national total using the provided weights (`wtsupp`).

D Data Appendix: Consumer Price level

To construct the consumer price level in agriculture and non-agriculture in each state, we use the price parities that Aten (2006), Aten and D’Souza (2008), and Aten (2008) estimated for the year 2006 for 363 metropolitan areas and for the rural area of each state. The execution of this approach is complicated by the fact that the Census obscures

the exact metropolitan area of residence for workers in small metropolitan areas or on the edges of some large metropolitan areas to protect their privacy. As a result of this restriction, the residence information from the Population Census falls into one of three categories: (i) the worker lives in an identified metropolitan area; (ii) the worker lives in the rural (i.e., non-metropolitan) area of an identified state; (iii) the worker lives in an unidentified metropolitan area of an identified state. We assign the following price levels to these three residence categories: (i) the price level for the metropolitan area of residence; (ii) the rural price level for the state of residence; (iii) the average price level for the unmatched metropolitan areas for the state of residence. An additional technical complication arises with metropolitan areas that span multiple states. We apportion these metropolitan areas (and their price levels) to the individual states using county-level employment data. We weight using nominal compensation at the county level, as did the original research.