How Regions Converge: Alternative Data Sources*

Francesco Caselli

Graduate School of Business

University of Chicago

Chicago, IL 60637

W. John Coleman II

Fuqua School of Business

Duke University

Durham, NC 27708-0120

June 1998

Abstract

Caselli and Coleman (1998) use data from Easterlin (1957) and the microdata samples of the US census of population to document a large increase in annual earnings of agricultural workers relative to non-agricultural workers. This appendix discusses alternative data sources, and argues that the upward trends in the relative farm wage is indeed a robust finding.

The first two columns of Table 1 report the series for the US-wide relative agriculture/non-agriculture wage generated by the data used in Caselli and Coleman (1998). Recall that our sources are Easterlin (1957) for the 1880-1950 period, and the microdata files from the US Census for the 1940-1990 period (also recall from Section 2.2 that, as a means of correcting for the difference in income concepts, we have forced the two series to coincide in 1950). The methods underlying these series are detailed in Section 2.2 and in Appendix 1 of Caselli and Coleman (1998).

An alternative series for the US-wide relative agricultural wage can be constructed from *Historical Statistics,* which provide historical average annual earnings per full-time employee in agriculture (Series D739) and the other major industries (D Series 740, 741, 745, 746, 750, 753, 754, 755, and 761). The latter can be combined with the corresponding employment figures (D Series 128, 129, 130, 133, 134, 137, 138, 139) to create a weighted average non-agricultural annual wage. It is then possible to obtain a measure of the relative farm wage by dividing the agricultural wage in Series D739 by the non-agricultural wage. The resulting series is presented in Table 1, under the column heading "Hist. Stat." It is immediately apparent that the alternative series based on the *Historical Statistics* data differs markedly from ours. In particular, it features no (or very little) upward trend in the relative farm wage. The rest of this appendix investigates the causes of this discrepancy.

According to the accompanying notes in *Historical Statistics*, the wage data in series D739-D761 originate in Table A-16 in Lebergott (1964) for the period 1900-1929, and in National Income and Product Account publications for the post-1929 period. Accordingly, we discuss the discrepancy in Table 1 separately for the these two sub-periods.

Table 1: Different Series for the Relative Farm Wage

Year	Easterlin[1]	Census[2]	Hist. Stat.[3]	NIPA[4]
1880	0.20			
1900	0.21		0.35	
1910			0.34	
1920	0.32		0.37	
1930			0.26	0.31
1940		0.35	0.29	0.36
1950	0.39	0.39	0.41	0.49
1960		0.51	0.34	0.41
1970		0.64	0.40	0.54
1980		0.69		0.54
1990		0.68		0.60

Sources. (1): Computed by the authors from data in Easterlin (1957) as described in Section 2.2 of Caselli and Coleman (1998); (3): Computed by the authors using the micro data files from the population census, as described in Appendix 1 of Caselli and Coleman (1998); (3): Computed by the authors from data in *Historical Statistics,* as described in this appendix; (4) Computed by the authors from data in *National Income,* as described in this appendix.

Pre-1929: Lebergott vs. Easterlin

It is clear that for this sub-period the only relevant discrepancy between our data –
from Easterlin (1957) – and those in *Historical Statistics* – from Lebergott (1964) –
concerns the year 1900. It turns out that the 1900 Lebergott estimate of the agricultural
wage is obtained by dividing total "Labor Expenditures" from the 1900 Census of
Agriculture by the number of "Farm and Plantation Laborers" from the 1900 Census
of Population.[1] The latter number is 2.047 millions. The problem with this estimate is
that it greatly undercounts the number of hired laborers in farming, and thus overstates
the farm wage in 1900. The reason is that, as discussed by Brainerd and Miller (1957), a
very large number of farm laborers are included, in the same table from the Population
Census, under the heading "Laborers (not specified)," whose total number is 2.588
millions. Hence, the correct number of farm laborers is the number used by Lebergott
plus the fraction of the 2.588m unspecified laborers that is deemed to work in the
agricultural sector. Brainerd and Miller's best estimate for this fraction is 40.3 percent,
i.e. about 1 million. When we correct Lebergott's estimate of the farm wage using
this figure, the relative farm wage falls to 0.23, which is remarkably close to the figure
we independently computed from Easterlin's numbers. Miller and Brainerd also report
attempts by previous authors to estimate the share of "not-specified" laborers who
work in farming. The lowest published estimate of this share is 17 percent, and it
implies a relative farm wage of 29 percent. In conclusion, after correcting the 1900
Historical Statistics figures for Lebergott's undercount of farm laborers the relative
farm wage is *at most* 29 percent and, using the most recent estimate, 23 percent.

---

[1]Specifically, the first figure is in *Twelfth Census,* Volume 5, Agriculture, Part I, Table 12, p. 145; the
second figure is in *Twelfth Census,* Volume 2, Population, Part II, Table 91, p. 505.

4

Post-1929: Census vs. NIPA

For dates after 1929 the sources for the *Historical Statistics* data are National Income and Product Accounts (NIPA) publications by the Bureau of Economic Analysis (BEA). It turns out that BEA has considerably revised its historical series for agricultural earnings since the publication of *Historical Statistics.* In the last column of table 1 we report the relative agricultural wage as it can be estimated from the most recent version of the National Accounts, which supersedes the version in *Historical Statistics.*[2] The revised NIPA numbers go some way towards reconciling the original *Historical Statistics* numbers with our estimates based on census data, and they do re-establish the impression of an upward trend in the relative agricultural wage. However, there is still a substantial discrepancy.

Recall that for the period 1940-1990 our figures are obtained from direct estimation of the agricultural and non-agricultural annual wage from the census microdata files, as documented in Appendix 1. A close comparison of the wage series reveals that, as for the pre-1929 period – the discrepancy is entirely due to the agricultural wage: our non-agricultural wage estimate differs from the one desumed from *National Accounts* by only 2 percent in 1960, 4 percent in 1980, 1 percent in 1990, and less than 1 percent in the other years. Instead, our farm wage series is 14 percent below the corresponding series from *National Accounts* in 1950, and more than 20 percent above in each of the subsequent decadal observations.

Agricultural wages are obtained in the NIPA with procedures that are quite different

---

[2]Specifically, for each year we divide the Agriculture, forestry and fishery wage and salary figure from Table 6.6 of *National Income* by a weighted average of the figures for Mining, Construction, Manufacturing, Transportation and public utilities, Wholesale trade, Retail trade, Finance, insurance, and real estate, Services, and Government, where the weights are the corresponding full-time equivalent employment numbers in Table 6.5.

from those underlying the data for almost all other industries. Wages and salaries for non-agricultural industries are normally computed using payroll information submitted by employers to State Unemployment Insurance Agencies. This insures that there is extremely wide coverage, as well as consistency of definition and sources across non-agricultural industries.

In this respect, it is very reassuring that our non-agricultural wage and the one obtained from payroll records are so close. On the other hand, the agricultural sector is not as well covered by Unemployment Insurance, so that the NIPA agricultural wage is constructed using a variety of alternative sources. We devote the rest of this appendix to a description of the sources and procedures for the NIPA agricultural wage but it is already clear that our census-based measures has the advantage over the NIPA measure of sharing the same source and the same methods underlying the non-agricultural wage.

We have devoted considerable time and effort to an attempt of establishing the origin of the discrepancy in the agricultural wage series between *National Accounts* and our own census-based estimates.[3] Written documentation for the agricultural wages in *National Accounts* is scarce and fragmentary. The following description is based mainly on verbal communications with several BEA and USDA officers. (For a description of our own census-based series we again refer the reader to Appendix 1 in the paper).

The NIPA measure of the agricultural wage – wage and salary per full-time equivalent employee in Agriculture, forestry and fishery - is an employment-weighted average of the earnings of two groups of workers: (i) employees of Farms, and (ii) employees in the Agricultural services, Forestry and Fishing industries.

---

[3]One difference is that the former is actually for "Agriculture, forestry, and fisheries," while ours is only for agriculture. Clearly, however, this cannot be the source of the large discrepancy.

(i) Employees of farms. The source of these data is the United States Department of Agriculture (USDA). USDA annualy surveys a sample of farming establishments and collects data on a variety of farm expenditures, including various forms of compensation paid to employees.[4] Sample weights are then used to construct an estimate of the US-wide bill for each form of compensation. This information is then conveyed to BEA, where a national farm wage bill is computed that essentially includes all cash wages and in-kind payments made to employees. In certain years BEA also made some further adjustments for underreporting associated with tax evasion, and an estimate of the voluntary 401(k) contributions paid by workers, but these adjustments are always minuscule. The resulting total wage and salary bill - Wage and Salary Accruals by Industry – appears in *National Income* in Table 6.3. USDA also provides an estimate of the number of full-time and part-time employees of Farms, which is computed from a separate survey of farm establishments.[5] BEA combines these data with information on hours worked by part-time workers to convert these numbers into estimates of full-time equivalent employees, which appears in Table 6.5 of *National Accounts*. Wage and Salaries for full-time equivalent employees of farms are obtained by dividing the wage and salary bill by the number of full-time equivalent employees.

(ii) Agricultural services, forestry, and fishing wages are computed by BEA using a variety of sources. For forestry and fishing the methodology is the same as for the typical non-agricultural industry. Namely, total wage and salary bill and employment numbers come from payroll data filed by employers with the authorities overseeing State Unemployment Insurance programs. This is also true for some, but not all categories of

---

[4]The survey has changed name several times. Currently it is known as the Agricultural Resource Management Survey. Previous incarnations include the Farm Costs and Returns survey, and the Farm Production and Expenditure survey.

[5]The Agricultural Labor Survey?

workers in agricultural services. For those categories of agricultural-service workers for which there is insufficient coverage in Unemployment-Insurance records BEA resorts once again to special estimates and calculation by USDA.

One important caveat is that this reconstruction of the agricultural wage numbers applies to the most recent years. We have found that the "institutional memory" on how these data were constructed as recently as the 1970s is fragmentary at best, so we are less than certain that the same methods and criteria applied in the past.[6] That qualification aside, however, our reconstruction does not reveal any basic conceptual difference between our definition of the agricultural wage and the one implied by the (recent) NIPA methodology. This suggests that ultimately an evaluation of the relative merit of the two measures must rest on a judgement of the quality of the underlying data. In this respect, it seems to us that there is a basis to deem our census-based estimates more satisfactory. First, our estimates are based on worker-level information, which is intrinsically more appropriate for a study of labor earnings, while the NIPA figures are based on establishment-level surveys. Second, our source being the census of population, there is no possible question as to whether the sample is representative. Instead, USDA surveys appear not to have been based on probability sampling until 1974, and at any rate they have a limited coverage. Third, because workers in the census report exactly the figure we are interested in estimating, we did not need to resort to the battery of judgmental modifications, additions and subtractions, that have been imposed in the NIPA. Lastly, our census data have the advantage of deriving from a unique source, while the NIPA figures require to combine several separate data sources.

As a last remark it is also important to note that the above constructed *Historical*

---

[6]The lack of an institutional memory, as well as the essential absence of written documentation, are probably attributable to the high degree of fragmentation in the data construction process, with responsibility for various stages of data handling divided between several different individuals at both USDA and BEA.

Statistics and *National Income*-based relative farm wage series appear to be in conflict with other information in *Historical Statistics*. For example, according to Series D697 compensation per man-hour in all industries grew 68 percent between 1950 and 1960, while according to series D698 compensation per man-hour in non-farm industries only increased 63 percent. Given the small weight of farm employment, this seems to indicate a large increase in relative farm wages (consistent with our data) and not a large decline as implied by the NIPA series in Table 1. This should cast further skepticism on the quality of the NIPA farm wages, or at the very least on the implausibly large agricultural wage estimate they imply for 1950.