Online Appendix: Labor Earnings Inequality in Manufacturing During the Great Depression

Felipe Benguria
University of Kentucky
fbe225@uky.edu

Chris Vickers
Auburn University
czvickers@gmail.com

Nicolas L. Ziebarth
Auburn University and NBER
nicolas.lehmannziebarth@gmail.com

February 28, 2020

1 Representativeness of Our Sample

Our dataset is not a random sample from the Census of Manufactures (COM). Instead, particular industries were selected for particular reasons and not for the purpose of creating a representative sample overall. It is natural to ask to what extent the final sample with which we work is representative of manufacturing as a whole during this period.

Our first set of comparisons examines the average size of establishments in the industries in our sample versus those industries not in our sample. Using the published volumes, this comparison can be done for revenue per establishment as well as blue collar employment and earnings. We can also calculate revenue per blue collar worker as the ratio of total industry revenue to total industry blue collar employment.

Figure 1 plots the difference in means between these two groups of industries for 1929 measured in units of the standard deviation of the variable. For example, the mean employment of industries in our sample is about 0.1 standard deviations smaller than for industries not in our sample. There is basically no difference for revenue and while the differences are larger for revenue and blue collar earnings per worker (though still smaller than 0.4 standard deviations), these differences are not statistically significant. We conclude that along this (admittedly limited) set of variables, our set of industries looks similar to industries not included in our sample.

We next consider whether our sample is drawn from parts of the country with particular economic and demographic characteristics. Figure 2 maps the percent of manufacturing wage earners covered by our sample by county in 1929 compared to the totals reported in the published volumes on the 1929 COM.[1] Figures 3 and 4 redo these maps based on coverage of the number of establishments and blue collar income. The patterns are similar

[1] These data were graciously given to us by Dave Donaldson, Richard Hornbeck, and Jamie Lee. We note

to the figure based on employment. Coverage is high in South Carolina and Texas, for example, and more spotty in the West.

Figure 5 compares some county-level economic and demographic characteristics as a function of the share of total manufacturing establishments covered by our sample. We report the change in units of the standard deviation of the dependent variable of an increase in the percentage covered from the 25th to 75th percentile. So, for example, increasing the share covered from the 25th to 75th percentile is associated with an increase of slightly more than 0.2 standard deviations of the log change in county population between 1920 and 1930. Similarly, the vote share for the Democratic presidential nominee between 1896 and 1928, the illiteracy rate, and the share of African Americans are positively associated with coverage by our sample. Turning to variables related to the severity of the Depression, we find statistically significant effects on the share of banks that fail between 1929 and 1933, one measure of the local severity of the Great Depression. On the other hand, the effect on the log change in retail sales, another measure local severity, is not statistically significant and the economic magnitude is very small. Figure 6 shows that these quantitative differences are much smaller if we calculate the percentage of manufacturing covered using revenue instead of number of establishments. Overall, we would argue that the demographic differences are reflective of the fact that our sample covers up a large fraction of manufacturing in the South. While this does translate into some demographic differences, on the balance, we do not think our sample is drawn from areas that had particularly harsh (or easy) experiences during the Depression.

2 Quality of the Datasource

We address two questions related to the quality of the underlying data.[2] The first question is whether the establishments in our sample represent a complete canvassing of all manufacturing establishments within the industries we consider. There are a number of reasons why establishments could be missing. One possibility is that for some reason, the Census was not able to enumerate all establishments in some given year. It is difficult to answer this question since it requires access to some external data to validate the Census enumeration. For the cement industry, Chicu, Vickers, and Ziebarth (2013) use industry trade publications that listed all establishments in operating at this time and found that the Census was successful in enumerating all such establishments. Hansen and Ziebarth (2017) compare the Census records from Mississippi to records from Dun & Bradstreet and find close agreement between the two as well. It is also possible that some records were lost after being tabulated. For example, Ziebarth (2015) highlights one case where it does appear some schedules were lost for the manufactured ice industry in Texas between being tabulated and transferred to the National Archives for storage. In general, this worry is easy to address by comparing total

that the published volumes did not necessarily report the totals for each county for each census year. So just under 10% of our sample does not have a corresponding match in the published volumes for 1929. This might be due to confidentiality reasons on the part of the Census Bureau or just that it was simply was not tabulated.

[2]We note that the papers by Ziebarth (2015) and Vickers and Ziebarth (2019) discuss in detail a number of checks on the quality of this datasource.

derived from our samples with those in the published volumes.

Underenumeration or the loss of schedules are very extreme forms of measurement error. Even assuming these are not major worries, there is a question about whether the information reported on the forms is accurate. These values are self-reported and, as far as we know, not checked in any way against some administrative data source. Like the problem of underenumeration, it is difficult to answer this question since it requires some (trusted) external data source to validate the Census records. One industry where such an external data source does exist is the cement industry. Chicu, Vickers, and Ziebarth (2013) take advantage of a court case that required cement establishments to report their production totals for a number of years that overlap with the Census. They find that production totals reported by the establishments on the schedules are quite close to those reported by the establishments to the court.

In a slightly different direction, Vickers and Ziebarth (2019) study whether edits made by the Census Bureau to the forms are non-random. Many of the original schedules exhibit markings that are clearly edits of what the establishments originally filled in. These markings in some cases fix obvious mathematical mistakes on the part of the establishment. For example, some of the fields on the schedules are supposed to be the sum of two other fields and, in some cases, it is clear that the Census Bureau corrected a mistake by the establishment in summing up the components. In other cases, it is harder to understand what the mistake was in the first place since the edit simply involves crossing out one number and writing in a new one with no explanation for the change. One potential problem is if these edits were more likely to be made as, say, a function of an establishment's size, which would generate heteroskedastic measurement error. In the end, Vickers and Ziebarth (2019) find no evidence that edits were more likely to occur for large establishments, and, conditional on an edit being made, the probability of an upward revision is no different for large or small establishments.

We conduct a new check on the quality of the data by examining the rates of rounding ("value heaping") over time.[3] The idea is similar to the analysis of age heaping in historical demographic data. If values of variables are too frequently rounded, this suggests a lack of effort on the part of the person filling out the schedule and a higher chance of measurement error. To operationalize this idea, we define a rounded value of a variable as one that ends in 0 or 5. Bedford's Law, an empirical regularity on the distribution of digits, predicts that the distribution of digits after the first is close to uniform. Figure 7 shows the percentages of rounded values for white and blue collar total earnings as well as total employees and revenue in 1929, 1933, and 1935. There are differences in this percentage across the variables, and, for 3 out of the 4 variables, there is a higher frequency of rounded values than predicted by Bedford's Law. This is *prima facie* evidence for "value heaping" and perhaps non-classical measurement error. However, these percentages are quite stable over time. So by focusing on differences over time, this form of measurement error does not contaminate the estimation of trends.

[3] We thank the editor for suggesting this.

3 Educational Attainment by Worker Skill

We interpret the distinction between salaried workers and wage earners as reflecting differences in educational attainment, a common measure of skill in the literature. To provide supporting evidence for this interpretation, we turn to the 1% sample from the 1940 Population Census. While the COM did not collect information on the educational attainment of workers, the occupational description and industrial affiliation of these white collar and blue collar worker categories can be matched to those in the Census of Population to obtain information on the characteristics of workers in these different types of jobs. In particular, IPUMS has coded occupations into 3 digit categories based on the 1950 occupational classification. We consider occupational codes 1 through 300 as white collar jobs and those with codes 301 to 496 as blue collar.

Pooling together workers in all manufacturing industries and in all states, we find that in 1940 only 17.2 percent of blue collar workers had years of schooling corresponding to a high school degree or more. In contrast, 61.6 percent of white collar workers had at least a high school degree. The median white collar worker has an educational attainment of 12 years of school (exactly a high-school degree) while the median blue collar worker had 8 years of schooling.[4] We put this into a regression framework to predict educational attainment based on the white versus blue collar distinction. We consider two definitions of educational attainment: (1) high school graduate and (2) some college. We also consider two specifications: (1) with no controls and (2) with controls for sex, race, industry and state. Table 1 shows that there is a statistically and economically significant relationship education and white collar employment. While not as good as directly observing educational attainment of a worker, this shows that the "collar color" of a worker's job carries considerable information about his educational attainment.

4 JMP Decomposition: Robustness Checks

4.1 The Choice of Base Year

The results of the JMP decomposition are potentially sensitive to the choice of the base or reference year. In the paper, we use the "first" year as the base year, meaning 1929. For robustness, we show the results from using the "last" year, meaning 1935, as the reference year.[5] This also means we normalize values to 0 in 1935 instead of 1929 and so the figures are "shifted up" relative to the those in the paper. Figure 10 shows the decomposition with this reference year. Overall, the qualitative results are not much changed relative to the results reported in the main body of the text. It is the case when using 1935 as a base year versus 1929 that residuals play a more important role in explaining changes in the 90-10 differential. This comes at the "expense" of a smaller role played by prices.

[4]Note that the Census only asked about years of schooling not educational attainment directly. We interpret 12 years of schooling or above as someone with a high school degree though this need not necessarily be the case.

[5]We could also, in principle, use the average of the coefficients over the three years as the base "year" or 1933.

4.2 Using Employment as the Measure of Establishment Size

In the text, we used the total revenue as our measure of establishment size. Figure 11 shows the results of the JMP model using total employment instead as our measure of establishment size. This is consistent with the approach taken by Atack, Bateman, and Margo (2004). The differences between the results using this measure of size and those in the text are not meaningful.

4.3 Winsorizing the Earnings Variables

In the main body of the text, we winsorized the earnings variables to mitigate the effect of outliers on the regression results. Here we consider the effects of this choice by reporting the results of the JMP regression without winsorizing the earnings variables. Figure 12 shows the results using unwinsorized earnings variables. The results are nearly identical to those using the winsorized earnings variables.

5 JMP Decomposition: Extensions

5.1 Durables vs. Nondurables

One interesting margin along which to split the sample is by whether the major product of the industry produces a durable good or not. Figures 8 and 9 plot the distributions of earnings for blue and white collar employees, respectively, separating industries that produce a durable good from those that do not. Only for blue collar workers in durable industries do we observe clear shifts in the earnings distribution over these years.[6]

Because the sharpest changes in the distribution of earnings were for blue collar workers, and particularly in the durable goods industries, we perform the same JMP decomposition for blue collar workers only, separately by nondurable and durable industries. Tables 2 and 3 show the regression results for nondurable and durable industries, respectively. The striking differences are in the regional effects with larger penalties for being in the South and smaller ones for other parts of the country (relative to New England) in non-durable goods industries and the reverse for durable goods industries. In non-durable goods industries, the coefficients on low income areas all increase from 1929 to 1933, after which there is a smaller further increase. That is, there was regional convergence in these industries. In durable goods industries, by contrast, there is a decline in the coefficients on low-income regions from 1929 to 1933, but some reversal of this trend in two of the three regions, with the other constant.

Turning to the decomposition, we plot the results for non-durable goods industries and durables in Figures 13 and 14 respectively. The 90-10 difference is flat for non-durables, while the 75-25 difference declines from 1929 and then is stable to 1935. The decomposition suggests a complicated pattern for inequality: Prices are associated with an increase in the

[6] This finding casts doubt on one interpretation of all these results that the changes in the overall earnings distribution are simply driven changes in measurement due to some quirk in how the 1933 COM was collected. That theory could not explain why the differences in 1933 are only there for blue collar workers in durable goods industries.

90-10 difference and a decrease in the 75-25 difference. In contrast, residuals are associated with a slight decline in inequality by 1935. Turning to durables, the prices are associated with declines both from 1929 to 1933 and from 1933 to 1935. In contrast, residuals are associated with an increase in inequality measures, particularly the 90-10 difference, before returning to roughly the same level as in 1929 by 1935.

5.2 Earnings per Hour

For blue collar workers, it is possible to infer hourly earnings using questions on the "typical" number of hours worked per week (assuming a constant number of weeks worked). An issue is that the wording of these questions regarding hours are quite heterogeneous from year to year. They include questions on the number of shifts, number of hours in operation, and the "normal" number of hours worked in a week. Even the wording of the workweek question changes from year to year, subtly affecting what is actually being reported. In 1929, the Census asks establishments to report the "normal number of hours for the individual wage earner" with the figures based on "practice followed during the year." In 1933, the question instead asks for the "number of hours the plant was operated (day shift only) during the week including December 15."[7] That is, the question asked about one particular week rather than the normal practice, as well as asking for the amount the establishment was operated for the day shift as opposed to the length of the workweek per se. The 1935 form asks for the "normal number of hours per week (day shift only)", though the context of the form suggests it is referring to establishment operating hours rather than individual workweeks. Because of these differences in the questions across years, we only calculate earnings per hour in 1929 and 1935 for blue collar workers. To do this, we assume that all workers within a given year work the same number of weeks. In this case, once we take the log transformation, this constant will be irrelevant for measuring inequality. It would matter obviously if we were interested in the mean of earnings per hour, but that is not our focus.

We now turn to examining the changes in the distribution of earnings per hour, as opposed to annual earnings per worker. Panel A of Table 4 reports the changes in the distribution of blue collar earnings per hour for 1929 and 1935, the only years where we have information on hours worked. We find that this distribution became more compressed over this period, for both the 75-25 and 90-10 differentials. The changes were larger in magnitude than those observed for the per worker earnings distribution. Such a disconnect could be due to either an increase in the dispersion of hours or an increase in the covariance between hours and earnings per worker. Panel B reports the changes in the distribution of hours per wage earner in 1929 and 1935. In 1935, the interquartile range is zero: Both the 75th percentile and 25th percentile workweeks are 40 hours long. In 1929, there is more dispersion as measured by the interquartile range. However, when the broader distribution is considered, the 90-10, 90-50, and 50-10 gaps are all unchanged from 1929 to 1935. Changes in the dispersion of the workweek seem unlikely to explain the disconnect between the income and wage measures. As we show below, the change in hours per worker was not uniform across the country: The workweek declined in the low wage South more than in other regions. This larger decline

[7] This number is not representative of the year as a whole. For manufacturing, the average workweek in 1933 was 36.4, but only 33.8 in December (Beney, 1936, Table II). For this reason, we exclude 1933 when we focus on hourly wages.

in hours for the South offsets the relative gain in hourly wages for this area, so the overall effects on the earnings distribution of these regional changes are somewhat muted.

Table 5 shows the regression results of specifications identical to the ones we estimated for earnings. Notable here are the changes in the regional differences in earnings per hour. In particular, earnings per hour in the South Atlantic and East South Central both increase by 29 log points while the West South Central region increases by 25 log points. This generates substantial degree of regional convergence, more so than in annual earnings, with regions all across the country catching up with New England. As with annual earnings, larger establishments are associated with higher earnings per hour.

One way to reconcile the larger degree of regional convergence in earnings per hour compared to that in earning per worker is if the hours worked changed differentially across geographic regions. Table 6 shows the results of regressing the length of the log workweek on the same set of variables as above. It is clear that regional convergence in *workweeks* is crucial in explaining the disconnect between the results for earnings per worker compared to those for earnings per hour. Note that in 1929, the South Atlantic, East South Central, and West South Central regions, conditional on industry and observable firm characteristics, had workweeks 6 to 10 log points longer than those in New England. By 1935, two of these three regions had essentially no difference in workweeks, and the South Atlantic had workweeks 4 log points shorter than that in New England.

One interpretation of these results ccomes from the paper by Bernanke (1986), which develops a model based on that in the paper by Lucas (1970). At the heart of the model is an assumption that there are "economies or diseconomies of 'bundling' of worker-hours." (pg. 219) So declines in hours worked driven by declines in labor demand unambiguously reduce the *marginal wage* but the effect on the *average wage*, which is defined as total earnings divided by total hours, is unclear. As Bernanke writes on pg. 217, "The rate at which firms can reduce hourly earnings as the workweek falls depends on workers' preferences and reservation utilities. Especially at low levels of work and earnings, when consumption is highly valued relative to leisure, it may not be possible to cut hourly earnings as sharply as hours... Thus the wage, in other words, hourly earnings, may rise even as labor demand and workweeks fall." Bernanke quoting on pg. 217 from earlier work by Daugherty, Chazeau, and Stratton (1937) reports on the experience of the iron and steel industry during the 1930s: "[F]irms' desires to maintain their work forces relatively intact led them to adopt 'staggered' or 'spread-work' schedules under which many workers worked only a few days a week... Firms must have recognized that their ability to keep cutting total earnings as the workweek shortened was limited, since if workers could not attain a subsistence level in the mill town they would be forced to try elsewhere. Thus real hourly earnings in iron and steel rose... as the workweek was cut." So as demand falls, the remaining employees have their hours reduced to this practical minimum and, therefore, the earnings distribution becomes more compressed.

6 Extensive Margin Analysis: Robustness Checks

Here we consider some robustness checks for our extensive margin analysis.

6.1 Counterfactual Exit and Entry Distributions

In the text, we only reported the differences in the measures of inequality between the counterfactual and actual distributions. For completeness, Figure 17 computes the counterfactual blue and white collar earnings distributions that includes the distribution of earnings of incumbent establishments and adds back in predicted earnings in 1933 for all establishments that exited between 1929 and 1933. Figure 18 does the same calculation for 1935. Figures 19 and 20 show the entry counterfactual distributions for blue and white collar workers in 1933 and 1935, respectively.

6.2 State and Industry Fixed Effects Only

In the text, we imputed earnings and employment based on observables in the preceding census. The idea was to regress earnings (and employment) growth rates for continuing establishments on observable characteristics and then use the predicted values based on exiting establishments observables. The assumption underlying such a method was that exits were due completely to random events once we condition on these observables. Here we consider a more limited set of observables: just state and industry fixed effects. Figure 21 shows the results for this approach. The difference between these results and those in the text are negligible. For brevity, we just plot these distributions for 1933. Similar results hold for 1935.

6.3 Employment Adjustment in Entry Counterfactual

In the text of the paper, we constructed a counterfactual distribution that removed the earnings penalty associated with entry. We did not make any adjustments to the employment of entrants to adjust for potential penalties there. Here we make those adjustments by running regressions for employment on the same set of controls as earnings. For brevity, we just plot these distributions for 1933. Similar results hold for 1935. We then construct counterfactual levels of employment for entrants and use these as weights in estimating the counterfactual density. Figure 22 shows this counterfactual density for 1933. This additional adjustment does not make a substantial difference.

References

Atack, Jeremy, Fred Bateman, and Robert A. Margo. 2004. "Skill Intensity and Rising Wage Dispersion in Nineteenth-Century American Manufacturing." *Journal of Economic History* 64:172–192.

Beney, M. Ada. 1936. *Wages, Hours and Employment in the United States, 1914-1936.* New York: National Industrial Conference Board.

Bernanke, Ben S. 1986. "Employment, Hours, and Earnings in the Depression: An Analysis of Eight Manufacturing Industries." *American Economic Review* 76:82–109.

Chicu, Mark, Chris Vickers, and Nicolas L. Ziebarth. 2013. "Cementing the Case for Collusion Under the NRA." *Explorations in Economic History* 50:487–507.

Daugherty, Carroll R., Melvin G. De Chazeau, and Samuel S. Stratton. 1937. *The Economics of the Iron and Steel Industry.* New York: McGraw-Hill.

Hansen, Mary E. and Nicolas L. Ziebarth. 2017. "Credit Relationships and Business Bankruptcies During the Great Depression." *American Economic Journal: Macroeconomics* 9:228–255.

Lucas, Robert E. 1970. "Capacity, Overtime, and Empirical Production Functions." *American Economic Review Papers & Proceedings* 60:23–27.

Neumann, Todd C., Price V. Fishback, and Shawn Kantor. 2010. "The Dynamics of Relief Spending and the Private Urban Labor Market during the New Deal." *Journal of Economic History* 70:195–220.

Vickers, Chris and Nicolas L. Ziebarth. 2019. "The Census of Manufactures: An Overview." In *Handbook of Cliometrics*, vol. 2, edited by Claude Diebolt and Michael J. Haupert. Springer.

Ziebarth, Nicolas L. 2015. "The Great Depression through the Eyes of the Census of Manufactures." *Historical Methods* 48:185–194.

Table 1: Educational Attainment by White Collar Status

	(1)	(2)	(3)	(4)
	High school graduate		Some college	
White collar	0.44***	0.41***	0.21***	0.20***
	(0.01)	(0.01)	(0.00)	(0.00)
Controls	No	Yes	No	Yes
Observations	30056	30056	30056	30056

Notes: These data are from the 1% sample of the 1940 Census of Population. Estimates are based on a linear probability model restricting to individuals employed in manufacturing industries and between ages 20 and 65. The set of controls includes sex, race, age, and age squared as well as state and industry fixed effects. Standard errors are robust.

Table 2: Regressions of Blue Collar Earnings per Worker: Non-durable Goods Industries

	Blue collar earnings per worker		
	(1)	(2)	(3)
	1929	1933	1935
Incorporated	-0.01	0.10**	-0.02
	(0.02)	(0.05)	(0.02)
Multiplant firm	0.00	-0.02	0.02
	(0.01)	(0.01)	(0.01)
Revenue	0.06***	0.05***	0.06***
	(0.01)	(0.01)	(0.01)
Mid-Atlantic	0.04	0.04	0.02
	(0.03)	(0.04)	(0.03)
East North Central	-0.02	-0.07*	-0.02
	(0.03)	(0.03)	(0.03)
West North Central	-0.02	-0.09**	-0.03
	(0.03)	(0.04)	(0.03)
South Atlantic	-0.39***	-0.26***	-0.21***
	(0.02)	(0.02)	(0.02)
East South Central	-0.43***	-0.23***	-0.23***
	(0.03)	(0.03)	(0.03)
West South Central	-0.27***	-0.26***	-0.20***
	(0.05)	(0.04)	(0.03)
Mountain	0.10*	0.07	0.07*
	(0.06)	(0.07)	(0.04)
Pacific	0.03	0.07*	0.08***
	(0.06)	(0.04)	(0.03)
Observations	11322	8607	9713
Year	1929	1933	1935

Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. The set of controls includes log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects. An observation is an establishment and each establishment is weighted by its number of blue collar employees. The excluded region is New England. Standard errors are robust.

Table 3: Regressions of Blue Collar Earnings per Worker: Durable Goods Industries

	Blue collar earnings per worker		
	(1)	(2)	(3)
	1929	1933	1935
Incorporated	0.01	-0.09**	0.07***
	(0.03)	(0.04)	(0.03)
Multiplant firm	0.02	0.02	0.02
	(0.03)	(0.04)	(0.03)
Revenue	0.05***	0.09***	0.05***
	(0.01)	(0.01)	(0.01)
Mid-Atlantic	-0.05	-0.18**	-0.17**
	(0.05)	(0.07)	(0.07)
East North Central	0.02	-0.08	-0.16**
	(0.05)	(0.07)	(0.07)
West North Central	-0.06	-0.28*	-0.22**
	(0.05)	(0.15)	(0.10)
South Atlantic	-0.20***	-0.22***	-0.21***
	(0.06)	(0.08)	(0.07)
East South Central	-0.38***	-0.32***	-0.33***
	(0.07)	(0.09)	(0.09)
West South Central	-0.31***	-0.24***	-0.31***
	(0.07)	(0.09)	(0.09)
Mountain	0.03	-0.11	-0.01
	(0.08)	(0.09)	(0.09)
Pacific	-0.04	0.00	-0.05
	(0.06)	(0.07)	(0.07)
Observations	7269	3606	4747
Year	1929	1933	1935

Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. The set of controls includes log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects. An observation is an establishment and each establishment is weighted by its number of blue collar employees. The excluded region is New England. Standard errors are robust.

Table 4: Summary Statistics of Earnings per Hours and Workweek

	Difference between percentiles...					
	75-25	75-50	50-25	90-10	90-50	50-10
Panel A: Blue collar earnings per hour						
1929	0.77	0.28	0.49	1.43	0.50	0.93
1935	0.68	0.27	0.41	1.10	0.47	0.63
Panel B: Blue collar hours per week						
1929	0.14	0.10	0.04	0.29	0.18	0.11
1935	0.00	0.00	0.00	0.29	0.18	0.11

Notes: The statistics correspond to the difference between the two percentiles listed. Each establishment is weighted by the number of workers in the particular "color" group. All variables are log transformed. We do not have information on blue collar hours per week in 1933 because the variable on the COM schedule for that year is not consistent with the other years. For earnings per hour, we assume that all workers in a given year work the same number of weeks. This constant is differenced out when we take the log transformation.

Table 5: Regressions of Blue Collar Earnings per Hour

	Blue collar earnings per hour	
	(1)	(2)
	1929	1935
Incorporated	-0.01	0.03
	(0.02)	(0.02)
Multiplant firm	0.01	-0.00
	(0.02)	(0.02)
Revenue	0.06***	0.04***
	(0.01)	(0.01)
Mid-Atlantic	-0.11***	-0.05
	(0.03)	(0.04)
East North Central	-0.08**	-0.06
	(0.03)	(0.04)
West North Central	-0.14***	-0.15**
	(0.03)	(0.08)
South Atlantic	-0.44***	-0.15***
	(0.02)	(0.03)
East South Central	-0.54***	-0.25***
	(0.04)	(0.04)
West South Central	-0.40***	-0.25***
	(0.04)	(0.06)
Mountain	0.03	-0.01
	(0.06)	(0.06)
Pacific	-0.02	-0.01
	(0.04)	(0.04)
Observations	18563	14059
Year	1929	1935

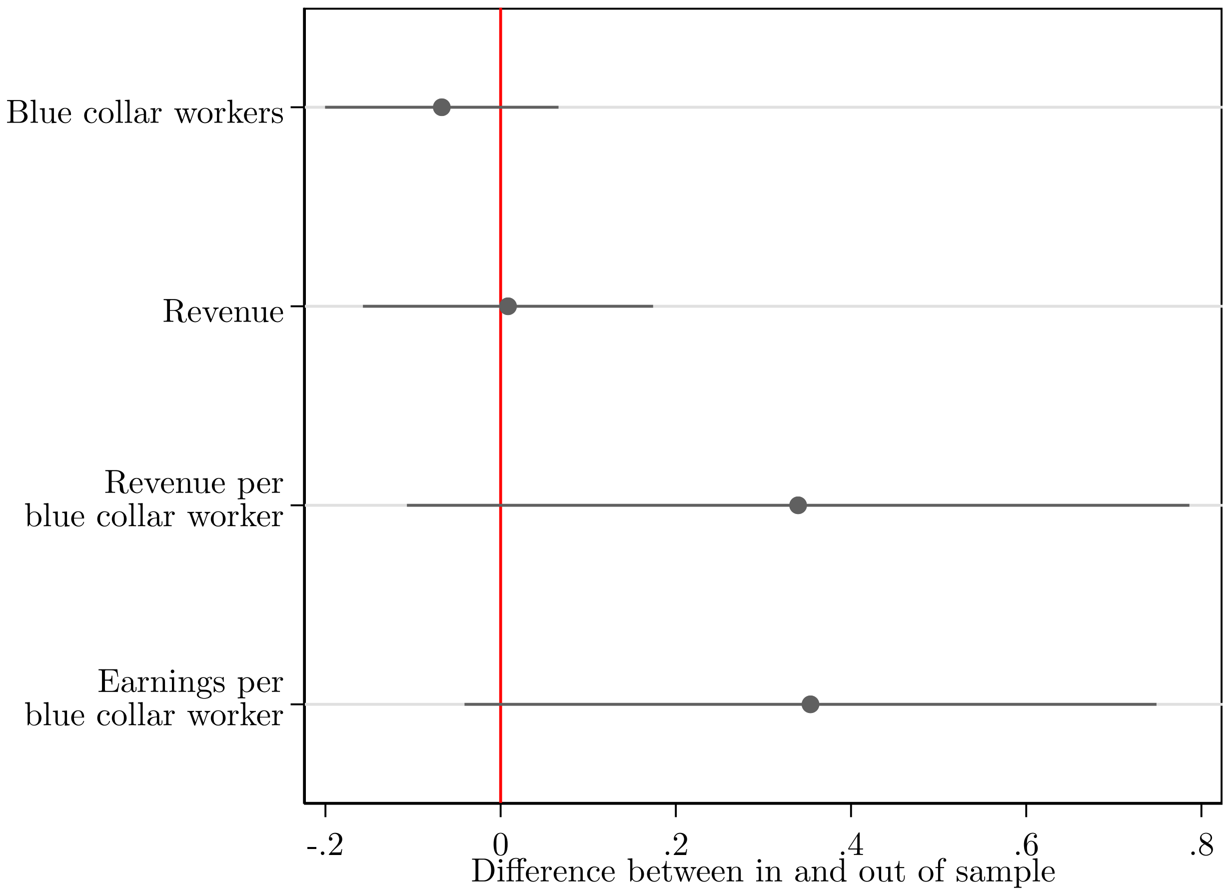
Notes: The earnings per week variable is log transformed and the 1% tails are winsorized. The set of controls includes log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects. An observation is an establishment and each establishment is weighted by its number of blue collar employees. The excluded region is New England. Standard errors are robust. The workweek variable is only reported in 1929 and 1935 so we can only calculate this variable in those two years. We assume that all workers work the same number of weeks in a given year. Once we apply the log transformation, this year-specific constant will be absorbed by the year fixed effects.

Table 6: Regressions of Blue Collar Hours per Week

	Blue collar hours per week	
	(1)	(2)
	1929	1935
Incorporated	0.01	-0.02
	(0.01)	(0.01)
Multiplant firm	-0.00	0.02
	(0.01)	(0.01)
Revenue	-0.00	0.01**
	(0.00)	(0.00)
Mid-Atlantic	0.03***	-0.03
	(0.01)	(0.02)
East North Central	0.03**	-0.03
	(0.01)	(0.02)
West North Central	0.03**	0.01
	(0.02)	(0.02)
South Atlantic	0.08***	-0.04***
	(0.01)	(0.02)
East South Central	0.10***	-0.00
	(0.02)	(0.02)
West South Central	0.06***	0.01
	(0.02)	(0.04)
Mountain	-0.02	0.05
	(0.02)	(0.05)
Pacific	-0.05***	0.03
	(0.01)	(0.02)
Observations	18775	14065
Year	1929	1935

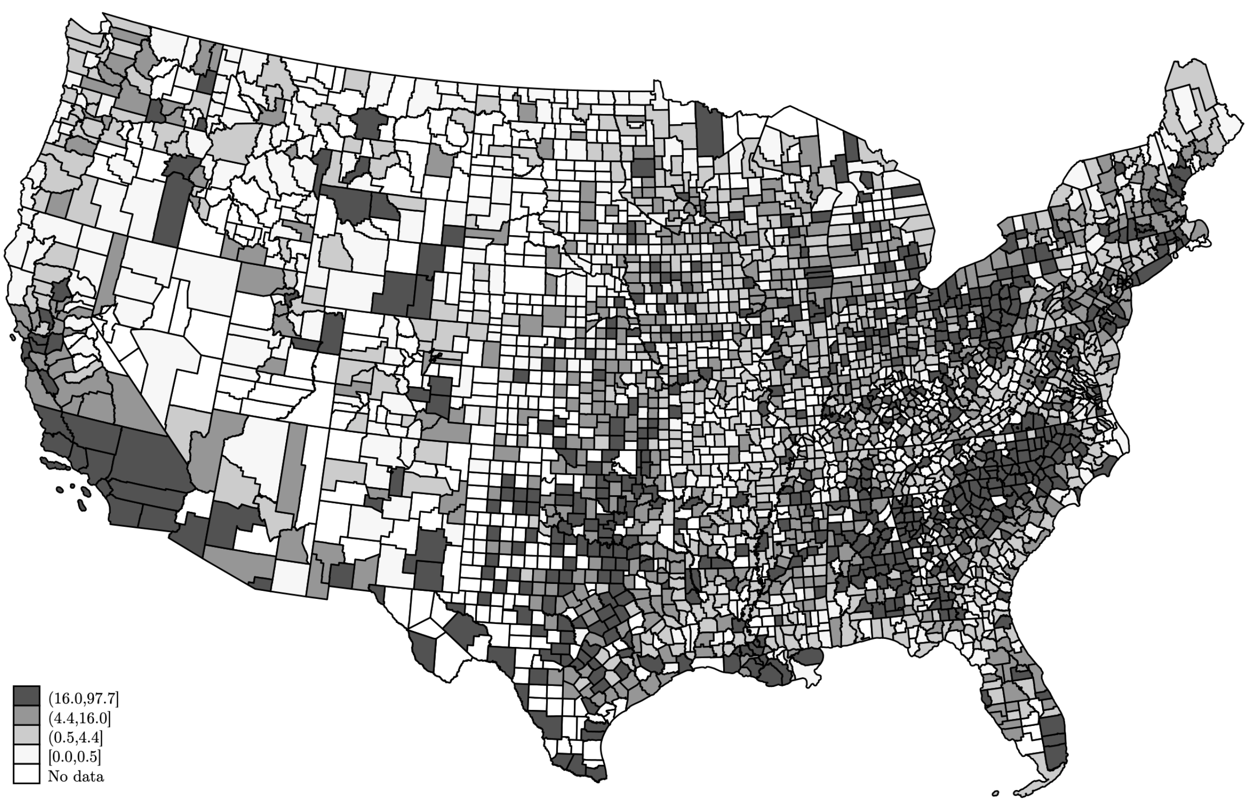
Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. The set of controls includes log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects. An observation is an establishment and each establishment is weighted by its number of blue collar employees. The excluded region is New England. Standard errors are robust. This variable is only reported in 1929 and 1935.

Figure 1: Comparing Industries in Our Sample to Those That Are Not



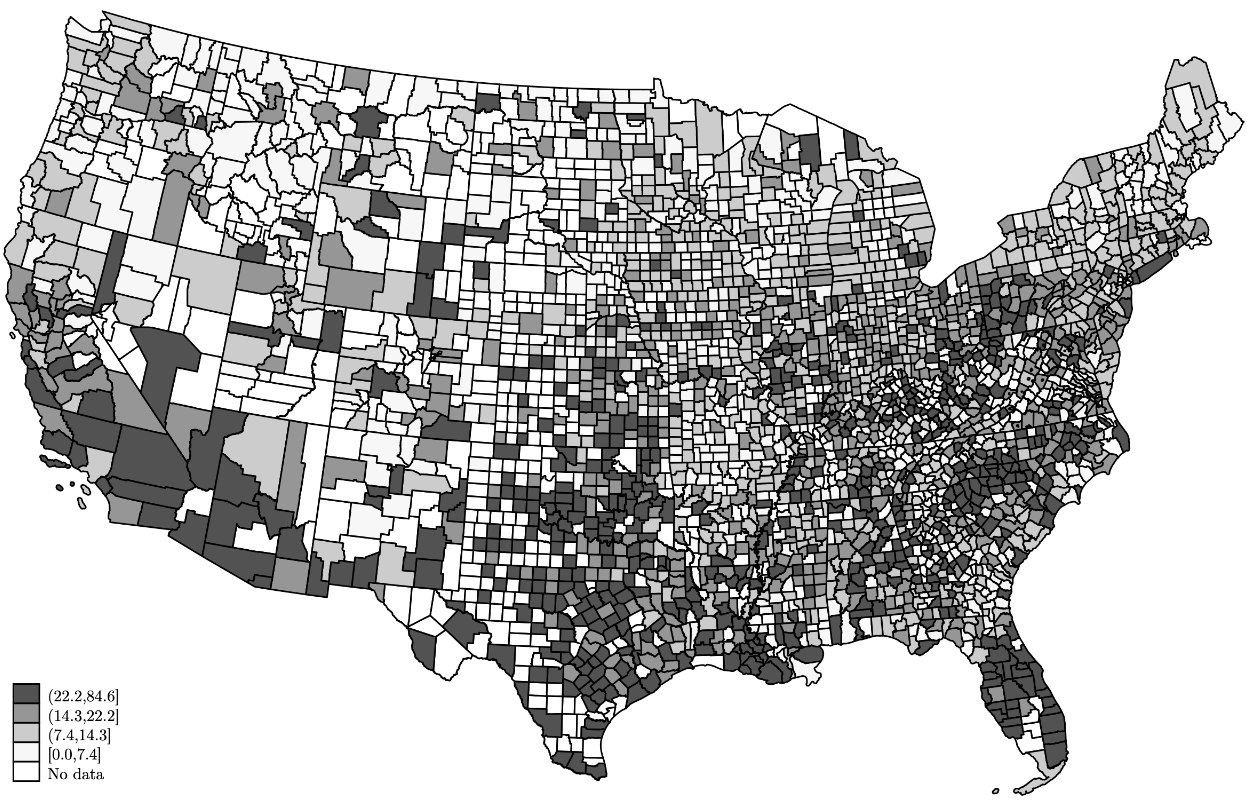
Notes: All variables are log transformed. We report the mean differences in 1929 industry-level averages measured in units of the standard deviations of the dependent variables. 95% confidence intervals of the difference based on robust standard errors are reported.

Figure 2: Geographic Coverage of Our Sample: Wage Earners



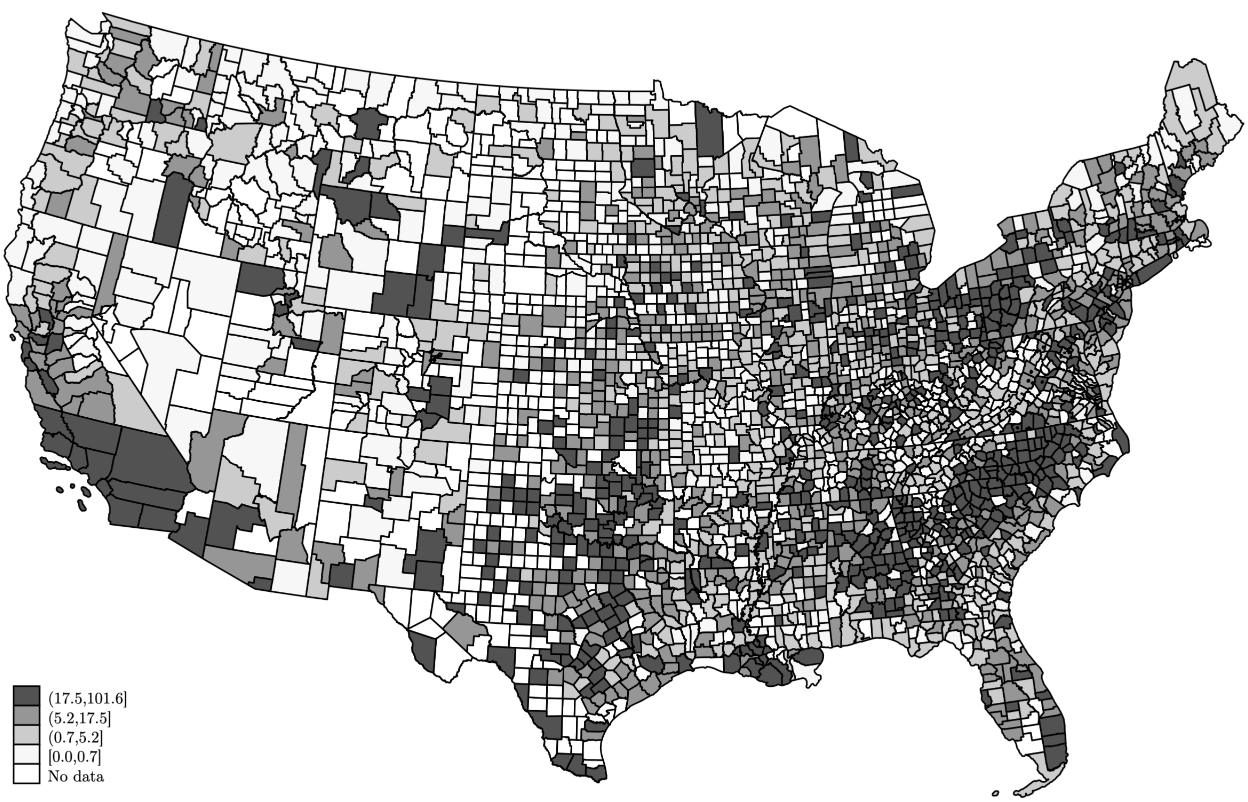
Notes: The percentage of wage earners is the total number of wage earners in our sample in 1929 relative to the total number of manufacturing wage earners reported in the published volume. “No data” means that a county’s value was not reported in the published volume for 1929.

Figure 3: Geographic Coverage of Our Sample: Establishments



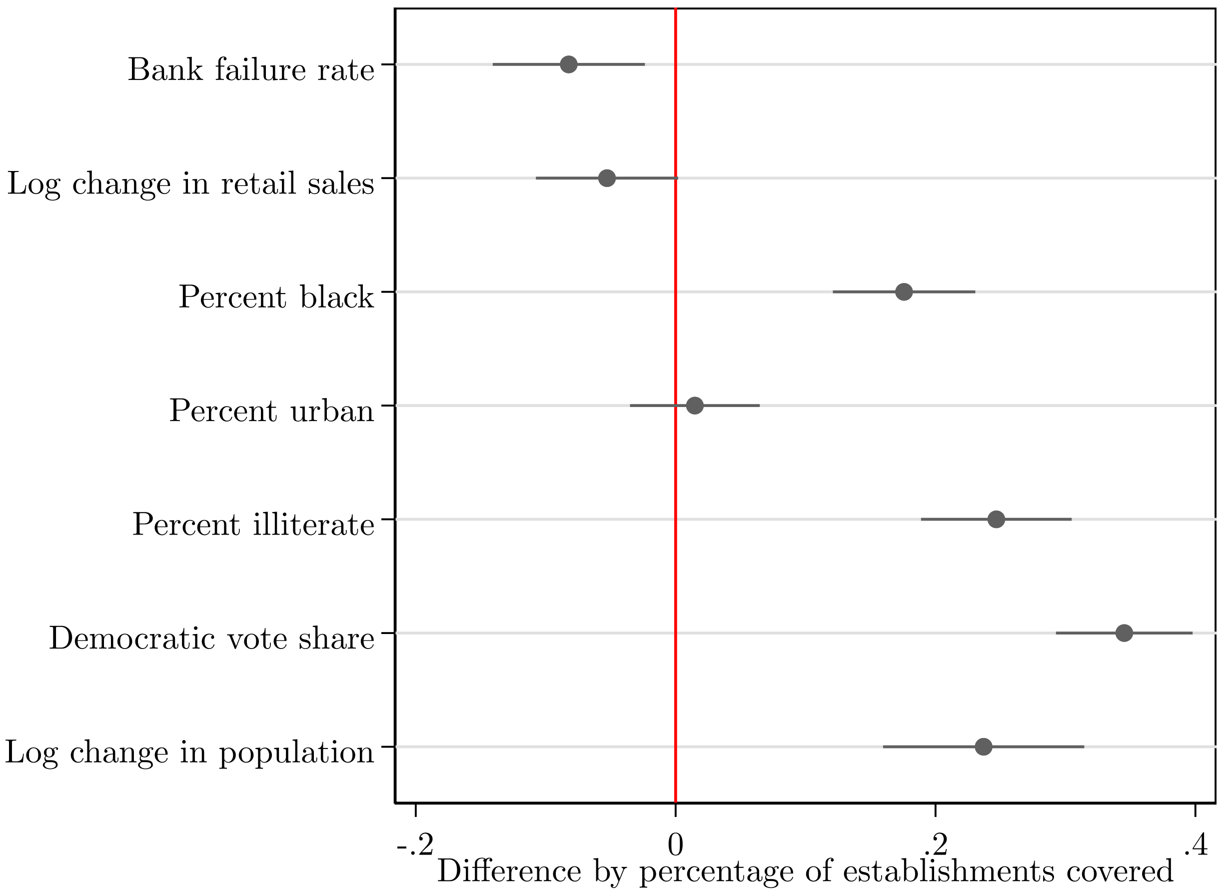
Notes: The percentage of establishments is the total number of establishments in our sample in 1929 relative to the total number of manufacturing establishments reported in the published volume. “No data” means that a county’s value was not reported in the published volume for 1929.

Figure 4: Geographic Coverage of Our Sample: Blue Collar Earnings



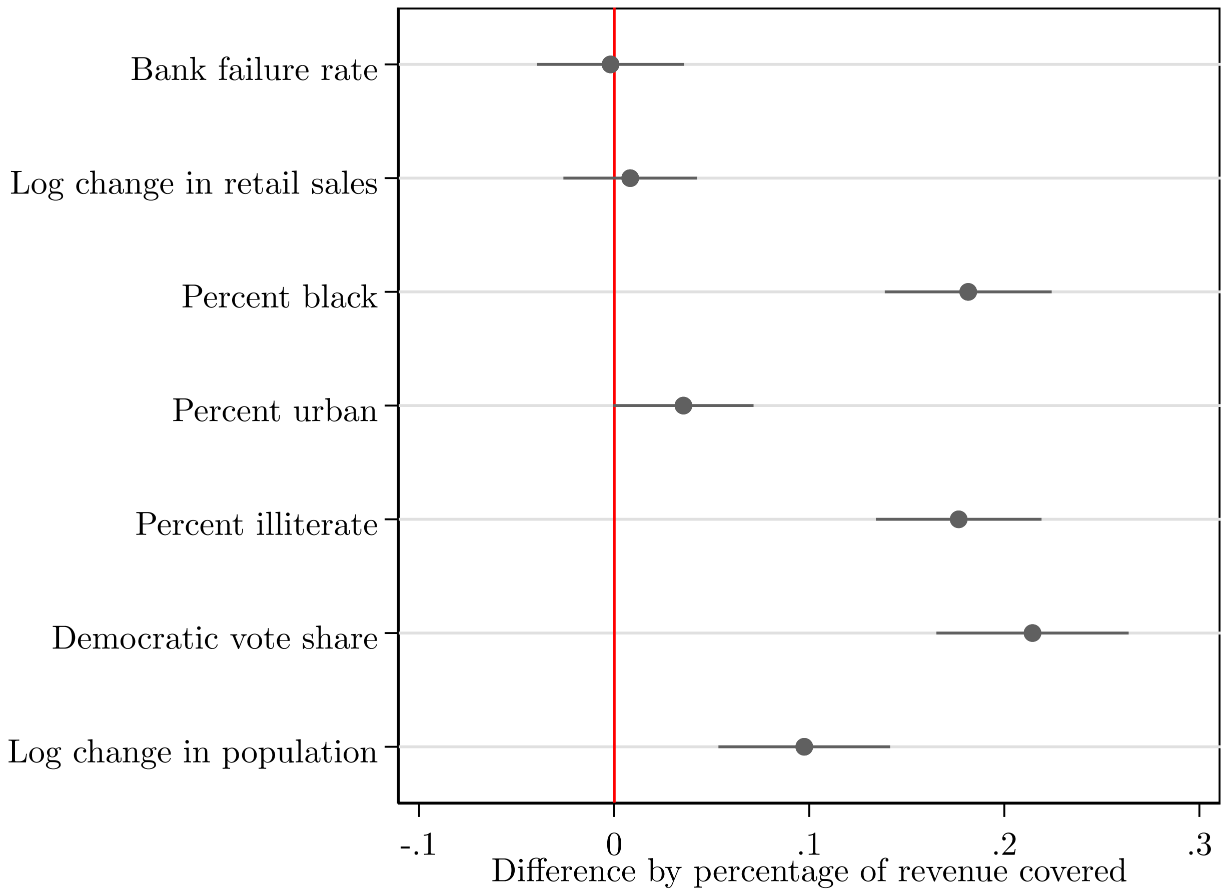
Notes: The percentage of blue collar income is the total blue collar income in our sample in 1929 relative to the total blue collar income reported in the published volume. "No data" means that a county's value was not reported in the published volume for 1929.

Figure 5: Comparing County Characteristics by Sample Coverage



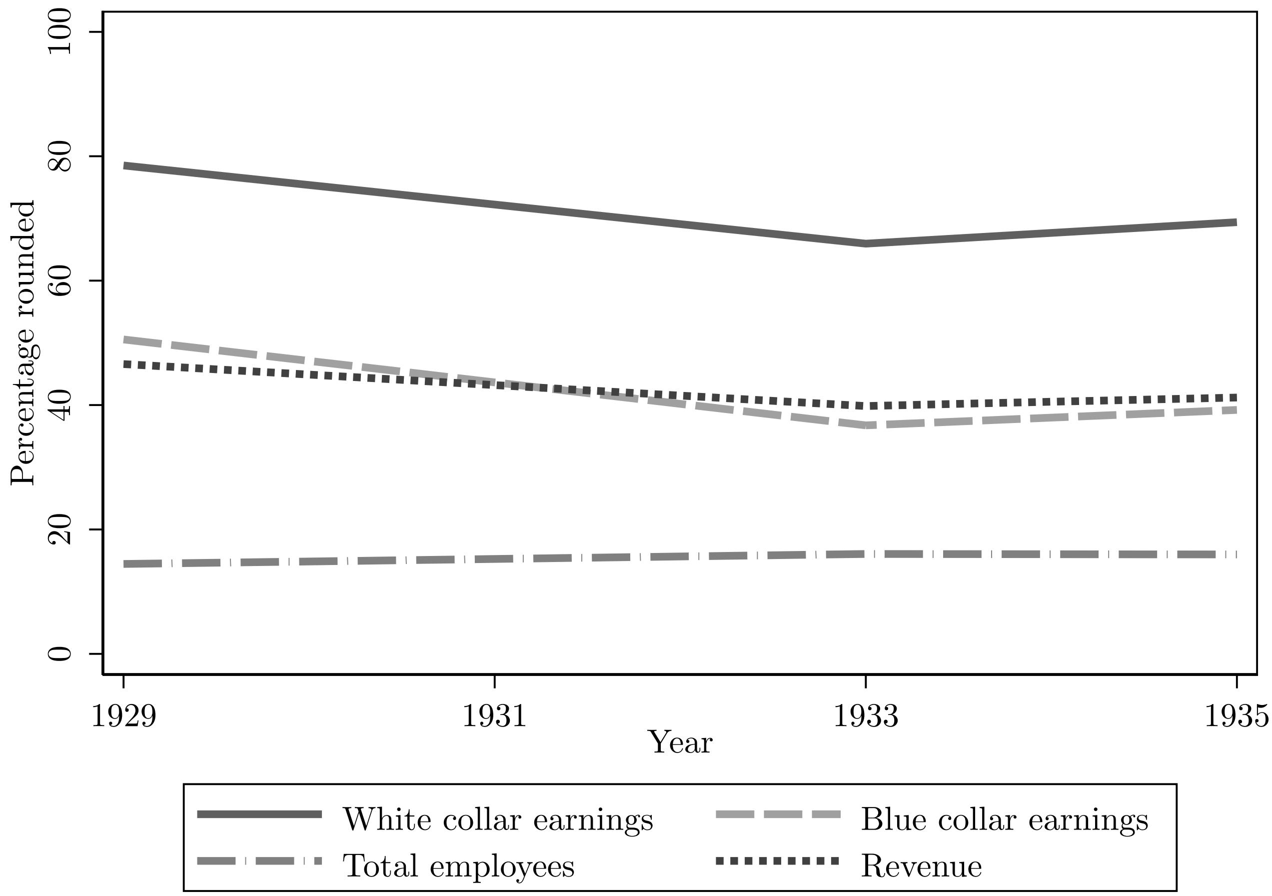
Notes: We report the coefficient on the percentage of establishments covered by our sample in a county scaled by the ratio of the 75th minus 25th percentile of this percentage. We report the effect in units of the standard deviation of the dependent variable. 95% confidence intervals based robust standard errors are reported. The bank failure rate is the share of ratio of the number of banks that fail between 1929 and 1932 relative to the number that exist in 1929. The log change in retail sales is the change between the years of 1929 and 1933 (Neumann, Fishback, and Kantor, 2010). Percent black, urban, and illiterate are all from the 1930 Population Census. The Democratic vote share is the average share for the Democratic candidate in Presidential elections between 1896 and 1928. The log change in population is between the years of 1920 and 1930.

Figure 6: Comparing County Characteristics by Sample Coverage: Revenue



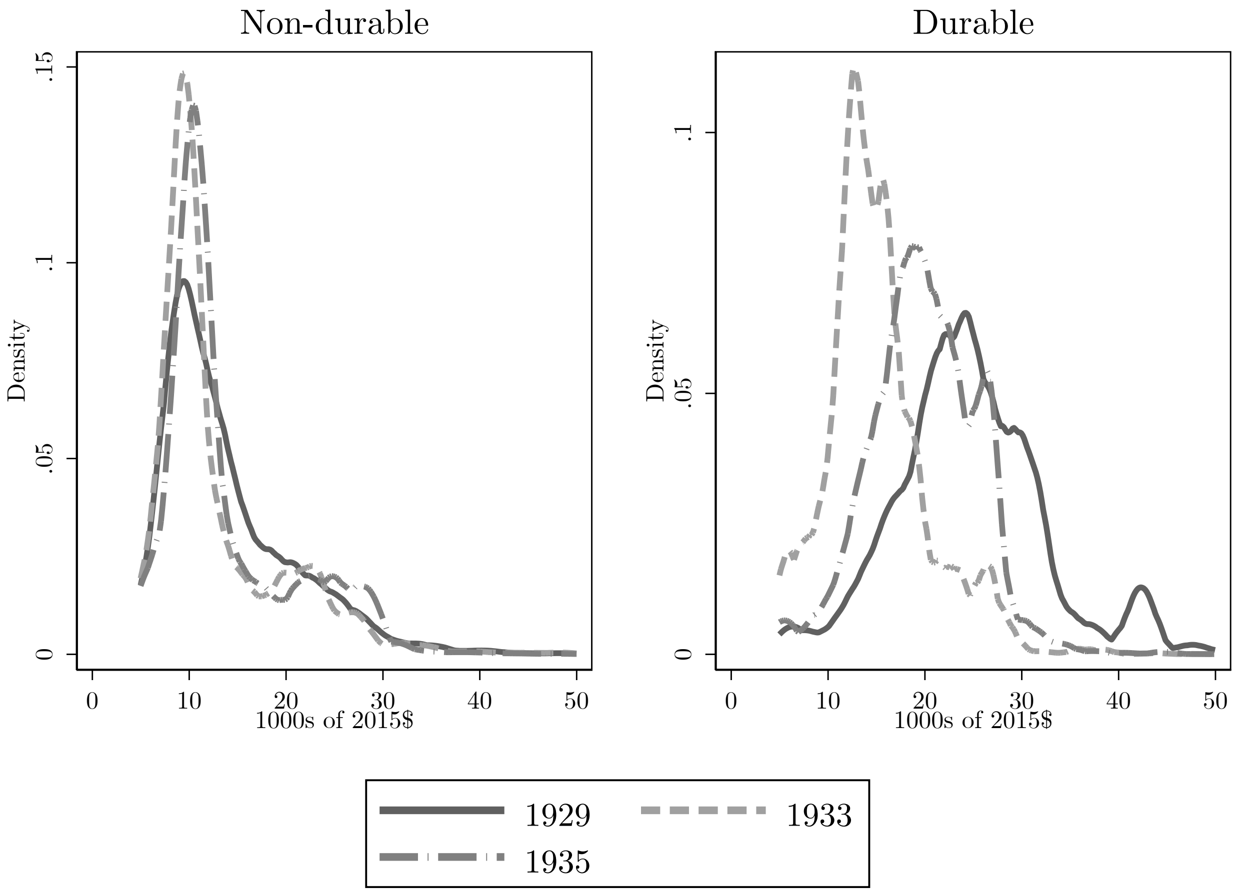
Notes: We report the coefficient on the percentage of total revenue covered by our sample in a county scaled by the ratio of the 75th minus 25th percentile of this percentage. We report the effect in units of the standard deviation of the dependent variable. 95% confidence intervals based robust standard errors are reported. The bank failure rate is the share of ratio of the number of banks that fail between 1929 and 1932 relative to the number that exist in 1929. The log change in retail sales is the change between the years of 1929 and 1933 (Neumann, Fishback, and Kantor, 2010). Percent black, urban, and illiterate are all from the 1930 Population Census. The Democratic vote share is the average share for the Democratic candidate in Presidential elections between 1896 and 1928. The log change in population is between the years of 1920 and 1930.

Figure 7: Percentage of Rounded Values



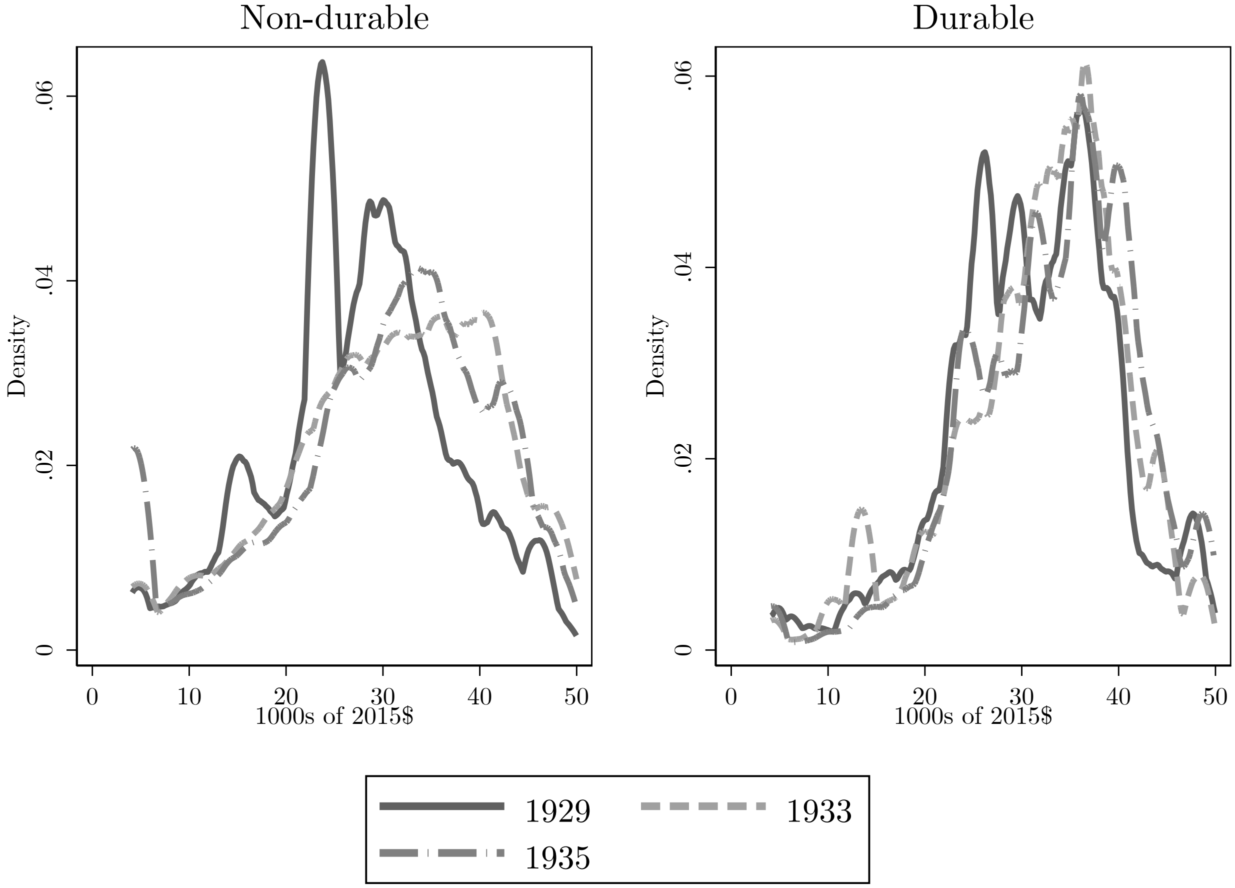
Notes: We define the value of a particular variable (white collar earnings, blue collar earnings, total employees, and revenue) as "rounded" if it ends in a 0 or 5. In calculating the percentage, we equally weight each establishment. Total employees is the sum of blue and white collar workers.

Figure 8: Distribution of Blue Collar Earnings per Worker by Durability of Product



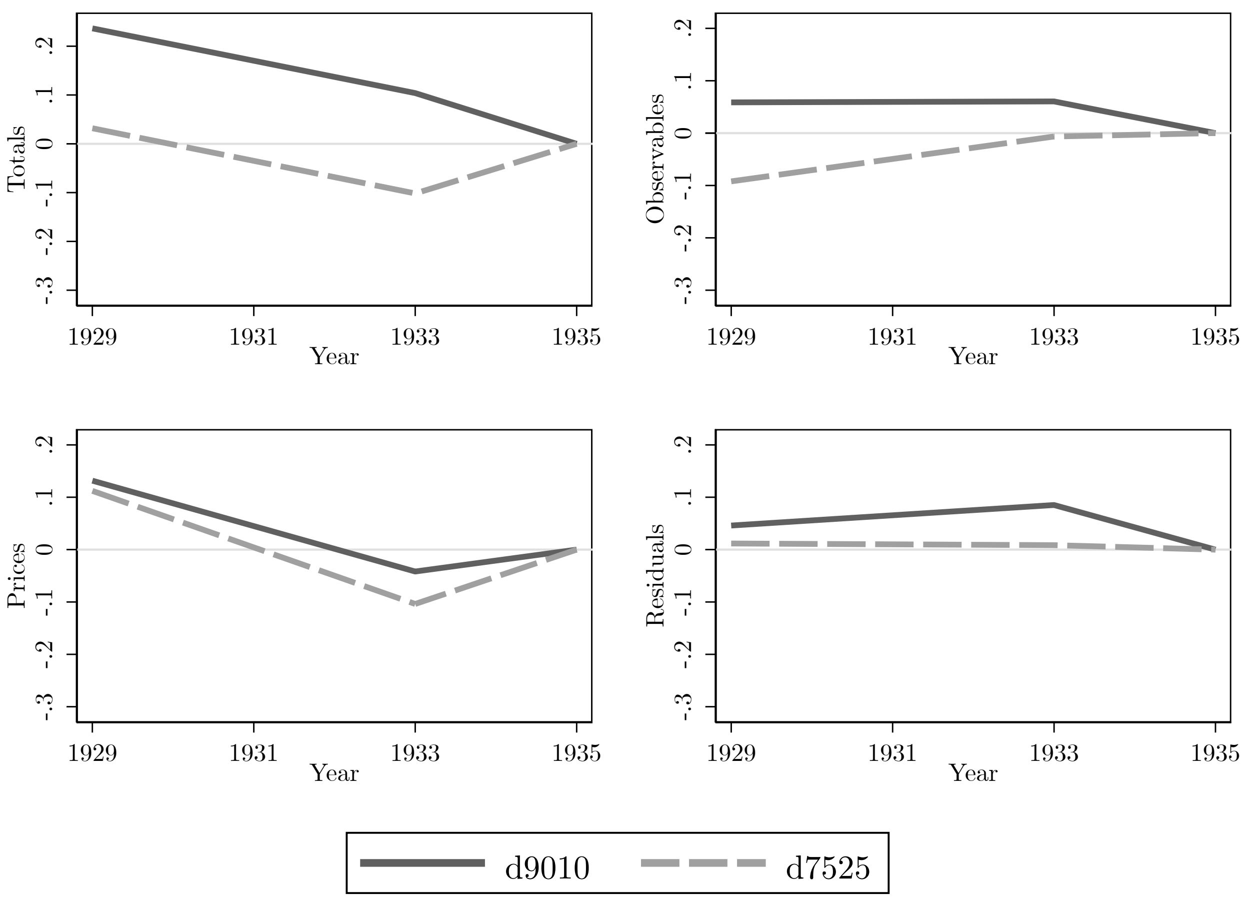
Notes: The earnings per worker variable for a particular group of employees is calculated as total earnings of that group divided by total number of workers in that group. The 1% tails of the distribution are winsorized. We report this variable in thousands of $ 2015. Durability is defined based on the industry's product. Employment weights are based on the number of blue collar employees at a given establishment.

Figure 9: Distribution of White Collar Earnings by Durability



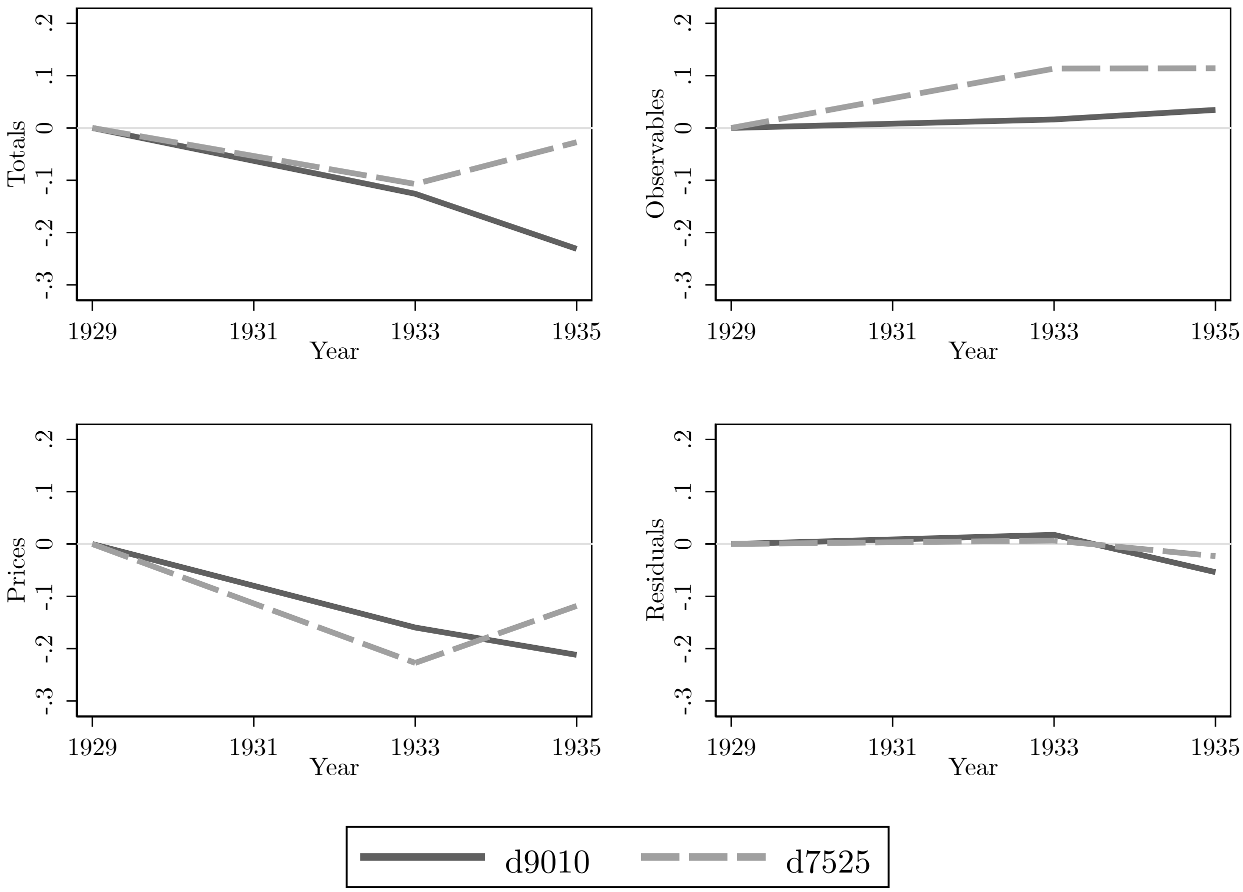
Notes: The earnings per worker variable for a particular group of employees is calculated as total earnings of that group divided by total number of workers in that group. The 1% tails of the distribution are winsorized. We report this variable in thousands of $ 2015. Durability is defined based on the industry's product. Employment weights are used and based on the number of white collar employees at a given establishment.

Figure 10: JMP Decomposition of Earnings per Worker: 1935 as Base Year



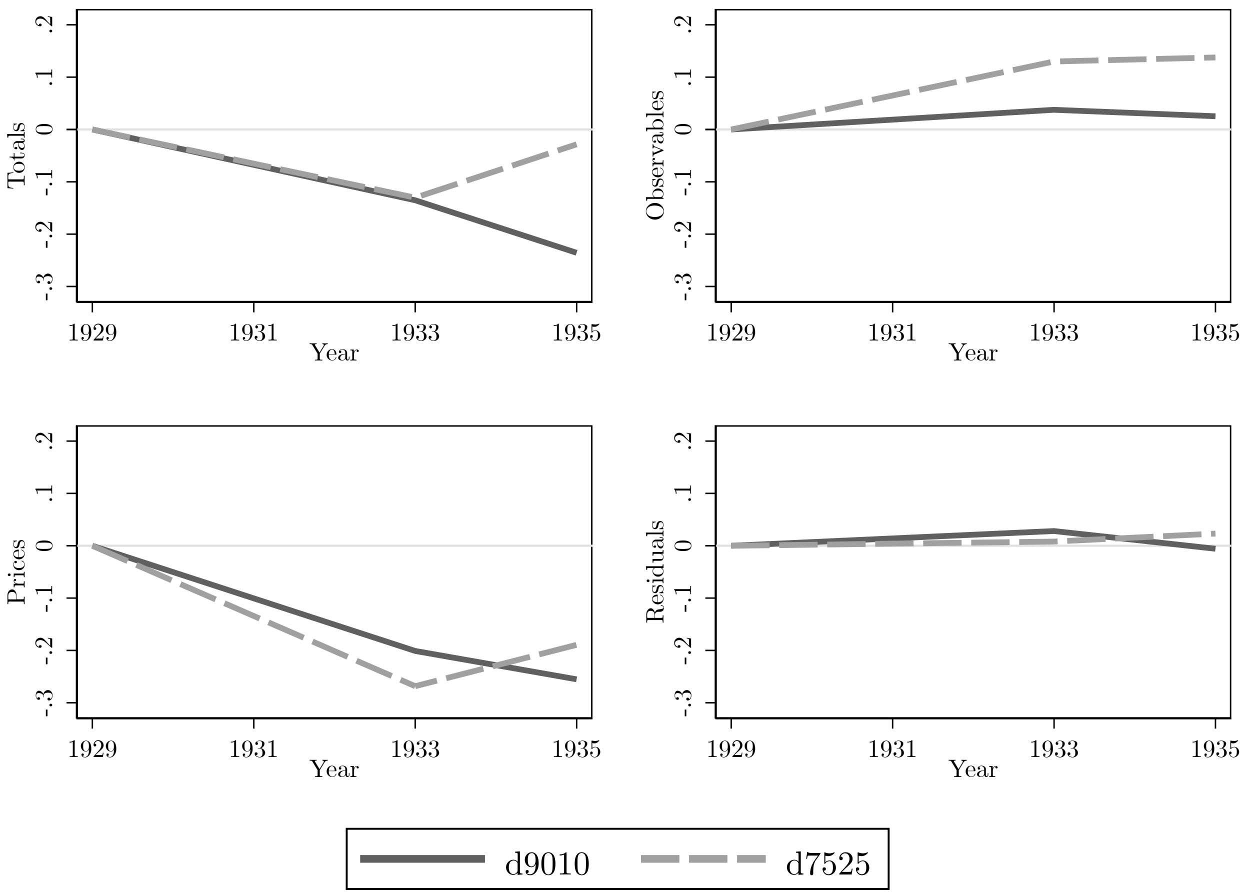
Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. The regressions here include the same set of controls as the JMP regressions in the main body of the text. The statistic "d9010" is the difference between the 90th and 10th percentiles. The statistic "d7525" is the difference between the 75th and 25th percentiles.

Figure 11: JMP Decomposition of Earnings per Worker: Employment as a Measure of Size



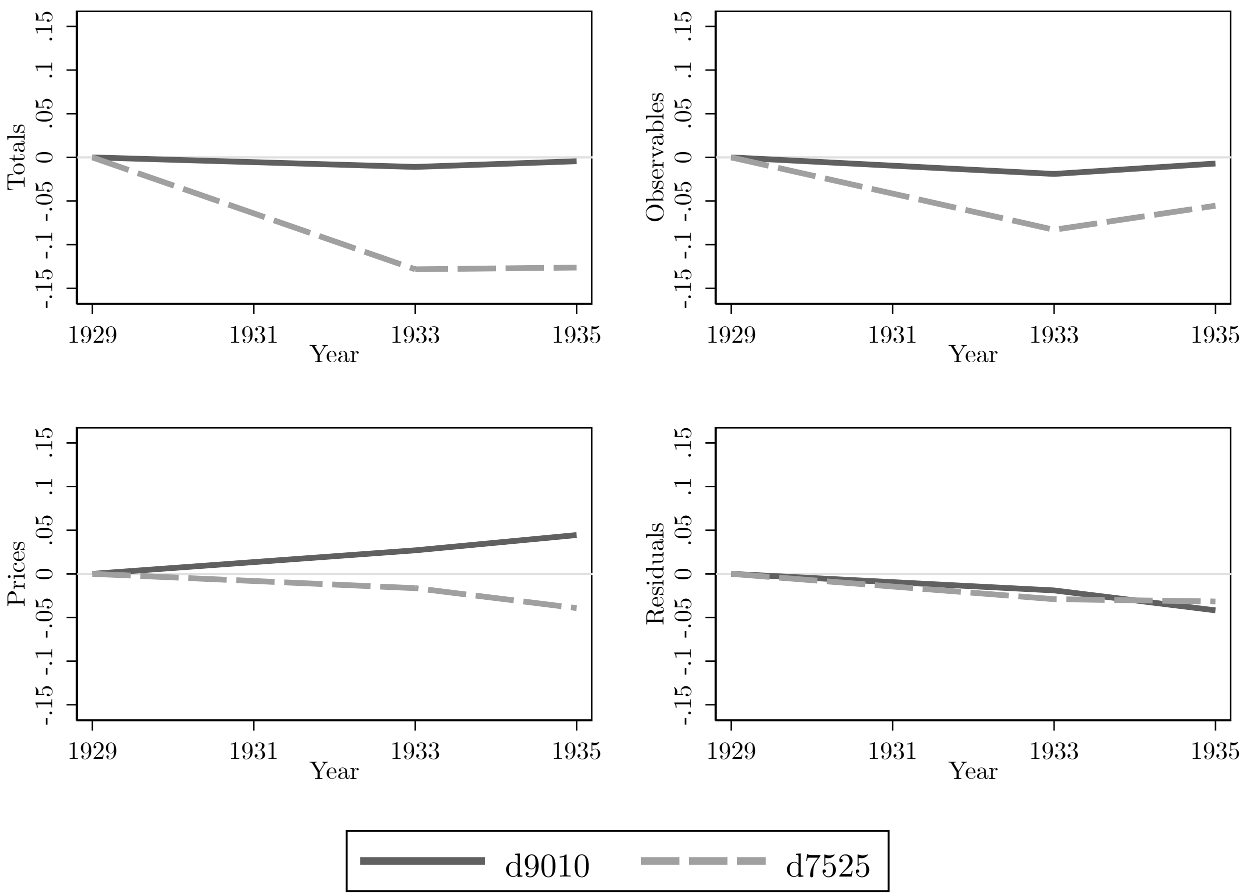
Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. The regressions here include the same set of controls as the JMP regressions in the main body of the text.The only difference is that we have replaced revenue with employment as our measure of establishment size. The statistic "d9010" is the difference between the 90th and 10th percentiles. The statistic "d7525" is the difference between the 75th and 25th percentiles.

Figure 12: JMP Decomposition of Earnings per Worker: Unwinsorized



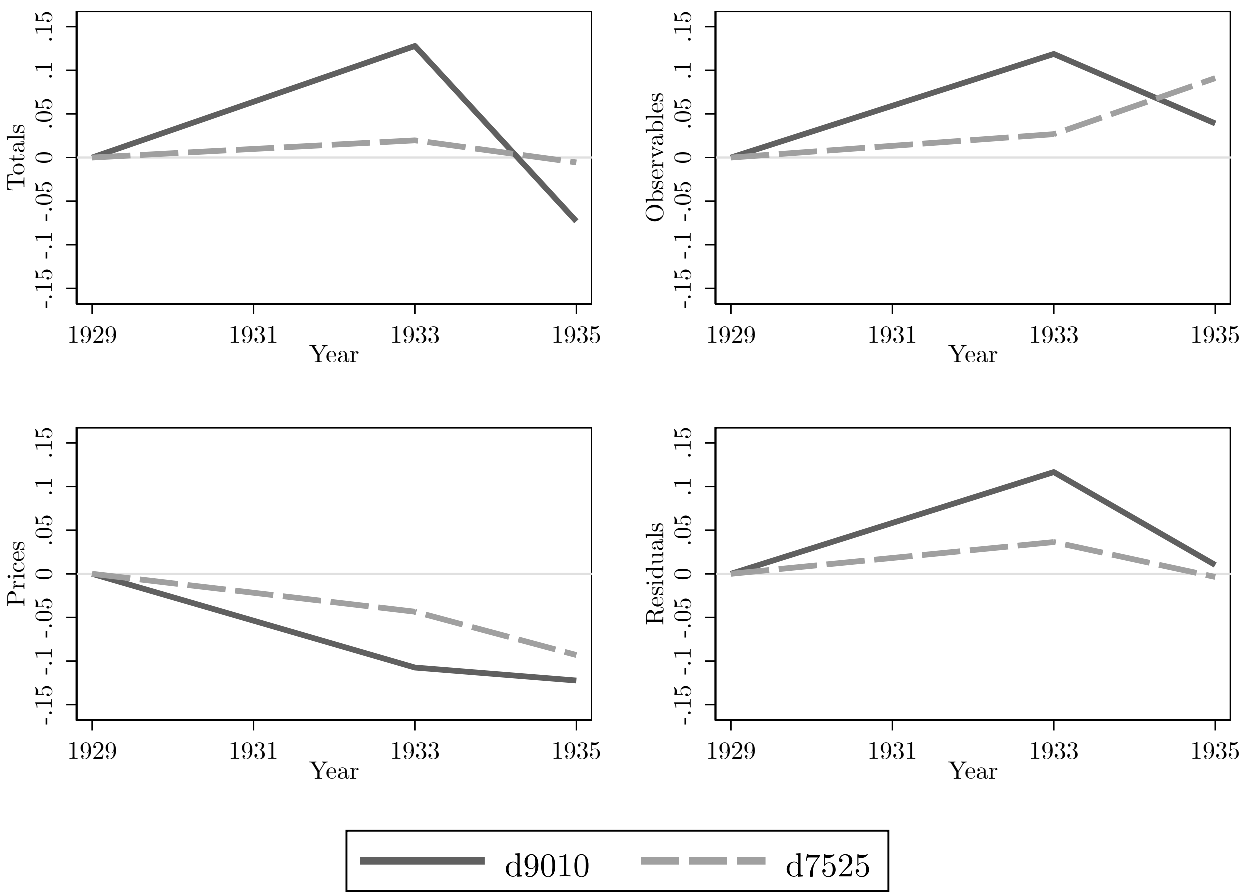
Notes: The earnings variables is log transformed but not winsorized. The regressions here include the same set of controls as the JMP regressions in the main body of the text. The statistic "d9010" is the difference between the 90th and 10th percentiles. The statistic "d7525" is the difference between the 75th and 25th percentiles.

Figure 13: JMP Decomposition of Blue Collar Earnings per Worker: Non-durables



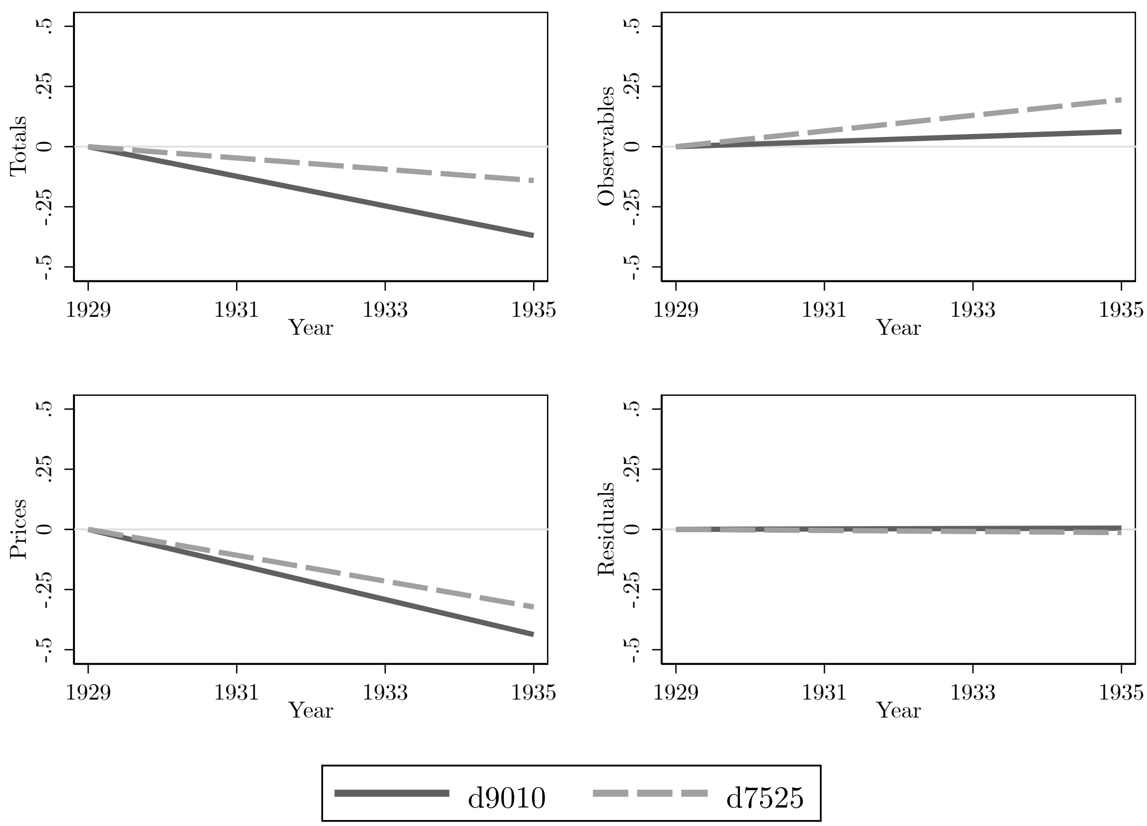
Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. These are based on regressions in Table 2 controlling for log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects and restricted to non-durable goods industries. 1929 is the base year. The statistic “d9010” is the difference between the 90th and 10th percentiles. The statistic “d7525” is the difference between the 75th and 25th percentiles.

Figure 14: JMP Decomposition of Blue Collar Earnings per Worker: Durables



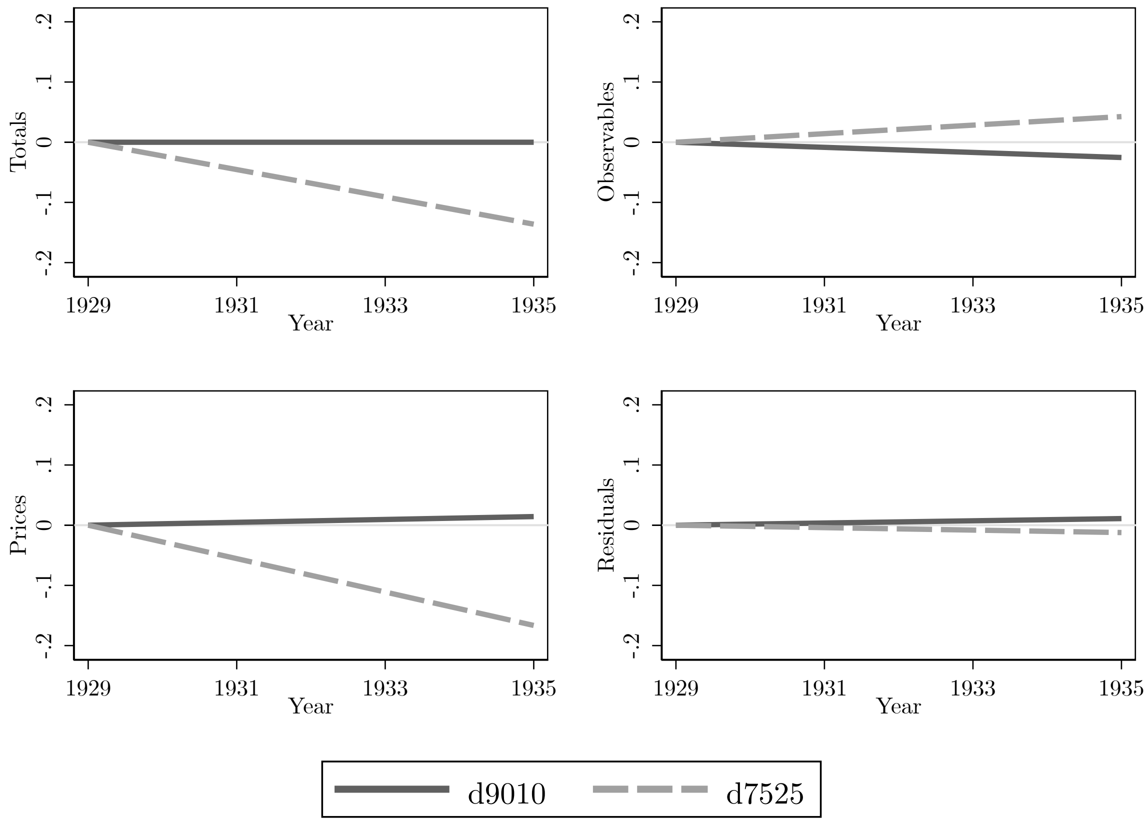
Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. These are based on regressions in Table 3 controlling for log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects and restricted to durable goods industries. 1929 is the base year. The statistic "d9010" is the difference between the 90th and 10th percentiles. The statistic "d7525" is the difference between the 75th and 25th percentiles.

Figure 15: JMP Decomposition of Blue Collar Hourly Earnings



Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. These are based on regressions in Table 5 controlling for log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects. 1929 is the base year. The statistic “d9010” is the difference between the 90th and 10th percentiles. The statistic “d7525” is the difference between the 75th and 25th percentiles.

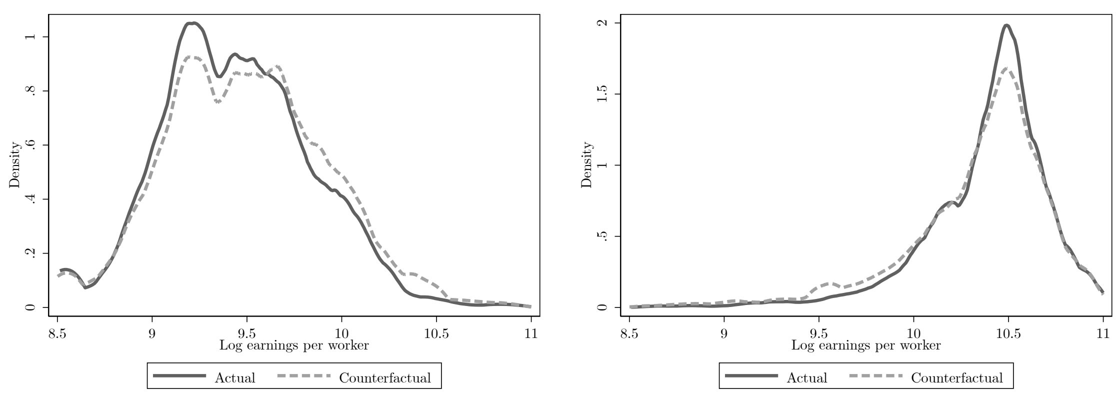
Figure 16: JMP Decomposition of Blue Collar Workweeks



Notes: The workweek variable is log transformed and the 1% tails are winsorized. These are based on regressions in Table 6 controlling for log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects. 1929 is the base year. The statistic "d9010" is the difference between the 90th and 10th percentiles. The statistic "d7525" is the difference between the 75th and 25th percentiles.

Figure 17: Exit Counterfactual Earnings per Worker Distribution: 1933

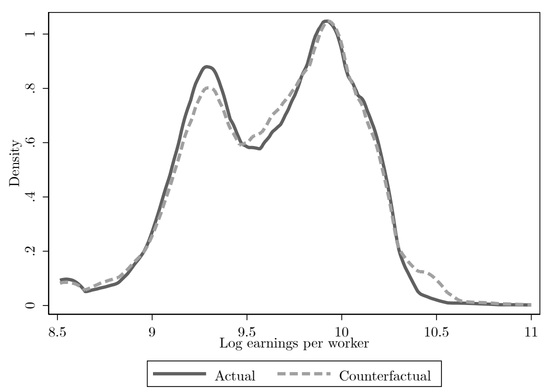
(a) Blue Collar (b) White Collar



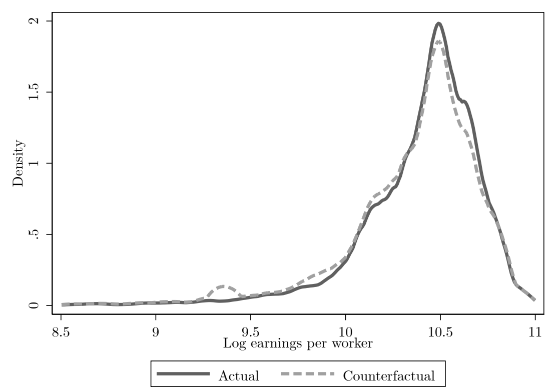
Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. The "Counterfactual" distribution imputes earnings per worker for establishments that exit between 1929 and 1933 based on 1929 observables including establishment size as well as industry and state fixed effects. The distribution and regressions use weights based on the number of employees in a particular occupational group at a given establishment. Employment for exiting establishments is imputed based on predicted values of employment growth rates using the same set of observables used to predict earnings. The "Actual" distribution excludes any establishments that entered in 1933.

Figure 18: Exit Counterfactual Earnings per Worker Distribution: 1935

(a) Blue Collar



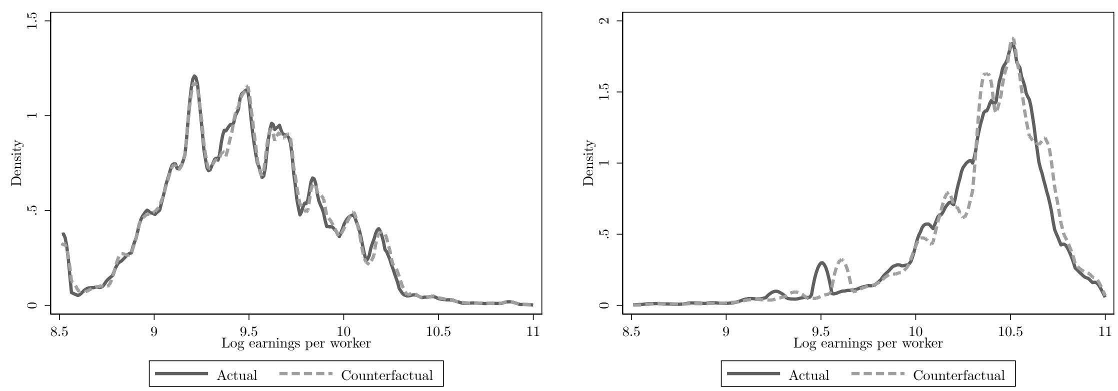
(b) White Collar



Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. The "Counterfactual" distribution imputes earnings per worker for establishments that exit between 1933 and 1935 based on 1933 characteristics that include establishment size as well as industry and state fixed effects. The distribution and regressions use weights based on the number of employees in a particular occupational group at a given establishment. Employment weights for exiting establishments is imputed based on predicted values of employment growth rates using the same set of observables used to predict earnings growth. The "Actual" distribution excludes any establishments that entered in 1935.

Figure 19: Entry Counterfactual Earnings per Worker Distribution: 1933

(a) Blue Collar (b) White Collar

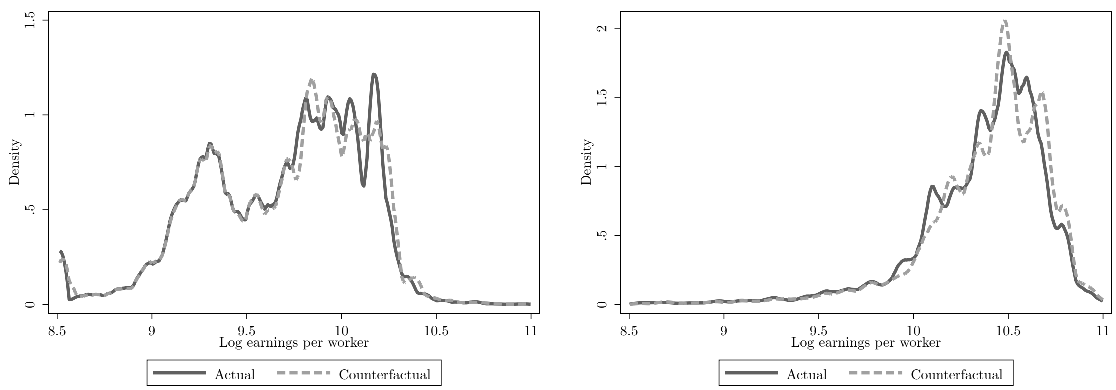


Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. The "Counterfactual" distribution removes the effect associated with entry based on a regression of earnings per worker that includes log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects. The distribution and regressions use weights based on the number of employees in a particular occupational group at a given establishment.

Figure 20: Entry Counterfactual Earnings per Worker Distribution: 1935

(a) Blue Collar

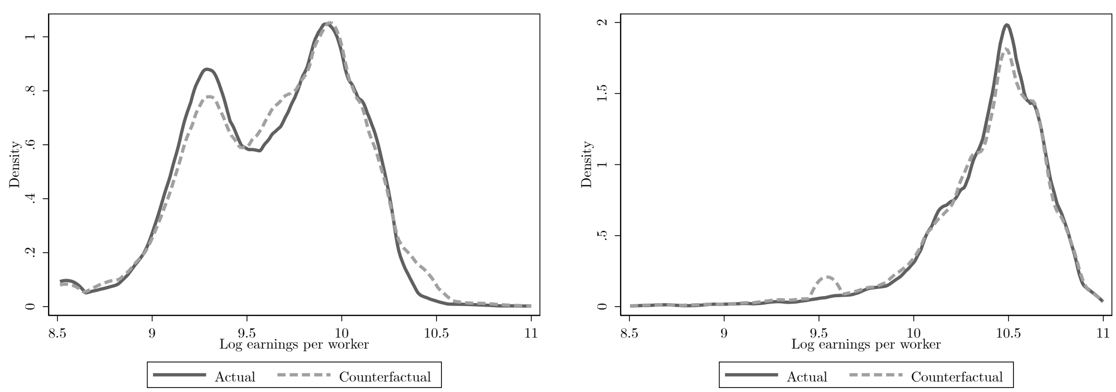
(b) White Collar



Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. The "Counterfactual" distribution removes the penalty associated with entry based on a regression of earnings per worker that includes log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects. Employment weights are based on the number of employees in a particular occupational group at a given establishment.

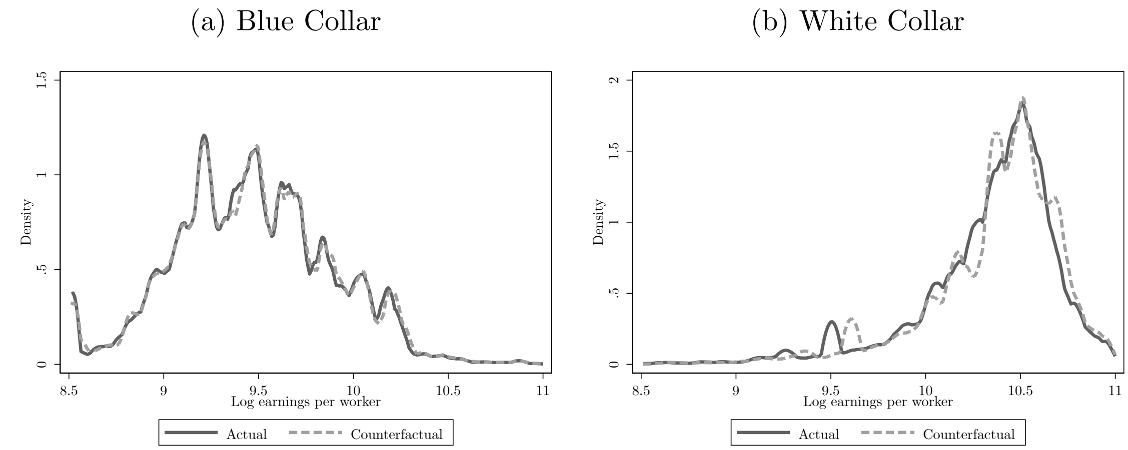
Figure 21: 1933 Exit Counterfactual Earnings per Worker Distribution: Limited Observables

(a) Blue Collar (b) White Collar



Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. The "Counterfactual" distribution imputes earnings and employments for establishments that exit between 1929 and 1933 based on state and industry specific growth rates of earnings. Employment weights are based on the number of employees in a particular occupational group at a given establishment.

Figure 22: 1933 Entry Counterfactual Earnings per Worker Distribution: Employment Adjustment



Notes: The earnings per worker variable is log transformed and the 1% tails are winsorized. The "Counterfactual" distribution removes the effect of entry on earnings based on a regression that includes log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects. Employment weights are based on the number of employees in a particular occupational group at a given establishment. We adjust the employment of entering establishments by removing any employment penalty associated with entry after controlling for log revenue, dummies for incorporation and multiplant status as well as region and industry fixed effects.