Technological and Organizational Change and the Employment of Women: Early Twentieth Century Evidence from the Ohio Manufacturing Sector

Marina Adshade and Ian Keay

Summer 2009

Data Appendix

A1: The Ohio Division of Labor Statistics Data

Starting in 1878 the Ohio Division of Labor Statistics conducted annual surveys which were used to derive employment and average weekly wage distributions by industry and sector, where industries were identified at roughly the four digit SIC code level of aggregation. Starting in 1886 wage and employment figures for men and women (over and under the age of 18) were reported separately. By 1912 all firms with five or more workers (revised to three or more workers in 1924) were required by law to report monthly employment figures and weekly wage distributions by occupation. Three types of occupations were included in the distributions: waged workers; bookkeepers, stenographers and typists; and; salesmen. The results from these surveys were tabulated and published in the Division's annual reports for the years 1914, 1915, 1923-1937. The data for the years 1916-1920 were collected but never published, and the data for 1921 was collected but never compiled. From 1923-1927 only information for industries identified at the more aggregated three digit level were included in the published reports. Data on hours worked per week was also collected, but only published in 1914.

The distribution of industry observations across years is fairly uniform, with a maximum of 211 industries included in 1936, a minimum of 145 industries included in 1914.

The published weekly wage and employment distributions were reported in very narrowly defined intervals. For example, the 1914 annual report lists the number of workers in each industry by gender, age, and occupation type who earned: "Under \$4"; "Over \$4 but Under \$5"; "Over \$5 but Under \$6"; ...; "Over \$35". Using the published wage distributions and a Tobit regression analysis, adjusted to account for the open ended upper interval, we have estimated mean wages for each industry, gender, and occupation for the available years between 1914-1937. The estimation approach we employ requires an assumption of log normality for each wage distribution. For industry-years in which workers' wages were only reported in one interval, the mean of the interval was used. In the rare cases where the workers' wages were only reported in the upper interval, the lower bound of that interval was used.

A2: Industry Group Sub-Samples

We have categorized each of the industries included in both the Division of Labor Statistics and Census of Manufacturing data sets on the basis of their organizational structures. This exercise is not straight forward, but our main qualitative conclusions are robust across most reasonable categorization schemes. There is a wide range of literature that describes the "key" features of the organizational transformation that was occurring in manufacturing establishments during the early 1900s. In our categorization procedure we group all of the four-digit industries based on four of these features: average establishment size (measured by total employment and value added), capital intensity (measured by physical capital per production worker ratios and capital cost shares), market structure (measured by concentration ratios and the extent of horizontal integration), and complexity (measured by occupational diversity). In total, therefore, we consider seven characteristics to assess these four organizational features.

For each available year between 1914-1937 we have constructed a series of dummy variables for each of our seven organizational characteristics. If an industry's measured characteristic - average number of employees per establishment, for example - is greater than the median industry in each year, then the dummy variable for that characteristic takes on a value of one for that industry. If an industry's measured characteristic is less than the median industry in each year, then the dummy variable takes on a value of zero. The industries with four or more dummy variables that take on a value of one in each year are considered big/integrated/complex (BIC) industries. The industries with three or fewer dummy variables that take on a value of one are considered small/unitegrated/simple (SUS) industries.

Average total employment has been derived by simply dividing the total number of employees in each industry by the total number of establishments reported in the manufacturing census for the state of Ohio. Average value added has been derived by dividing total value added in each industry by the total number of establishments. We expect industries with larger establishments to be organizationally unique from industries with smaller establishments. We also expect industries that were relatively capital intensive to be distinct from industries that used little capital. We measure capital-labor ratios by dividing the total fixed capital figures provided in the manufacturing census by the total number of production workers reported for each industry in each year. Capital cost shares are derived by dividing the value of residual gross output, after subtracting payments to labor and intermediate inputs, by the value of gross output. This approach assumes that rev-

Table 1: Simple Correlation Coefficients among Organizational Characteristics

	# Employees	Value Added	K/L Ratio	K Share	CRatio	Chandler
# Employees	1.000					
Value Added	0.816	1.000				
Capital:Labor	-0.086	0.083	1.000			
Capital Cost Shares	-0.124	0.018	0.761	1.000		
Concentration Ratio	0.102	0.157	0.168	0.064	1.000	
Chandler	0.189	0.159	0.028	-0.061	0.281	1.000
Michaels	0.080	0.162	0.222	0.074	0.398	0.310

enues are equal to costs, and capital costs are calculated as a residual. In 1958 the US Senate's Committee on the Judiciary established a subcommittee on Anti-Trust and Monopoly Formation. This subcommittee published information on manufacturing industry concentration ratios for 1947. Industry concentration ratios were defined as the proportion of total output produced by the four largest firms in each industry. We consider concentration ratios in excess of 33 percent to have been relatively integrated industries, and industries with concentration ratios of 33 percent or less to have been relatively unintegrated. Chandler (1969, Chart 1) provides an alternate categorization of integrated and unintegrated two digit manufacturing industries for 1939 based on his own measurement of industry concentration. We use Chandler's assessment of industry integration in addition to the Anti-Trust and Monopoly Formation's assessment because, although it is not as disaggregated, it is chronologically closer to the end of our sample period and it is very widely cited in the literature on organizational transformations. Michaels (2006, Appendix Table A1) provides us with our final measure of organizational structure. He categorizes two digit industries on the basis of their production "complexity" in 1910 and 1940. Complexity is measured as the inverse of the Herfindahl index of occupational diversity within each industry, which simply implies that industries with more occupations identified among their workers are considered more complex.

Over all of the available years 49.4 percent of our industries have been categorized as small/unintegrated/simple and 50.6 percent have been categorized as big/integrated/complex. From Table 2 we can see that industry categorization across our seven organizational characteristics tends to be positively correlated, but with just a few exceptions the correlation coefficients are not high. This suggests that while big industries also tend to be capital intensive, integrated and complex, each of our measures is capturing different information about industry structure.

A3: The Ohio Census of Manufacturing Data

To estimate translog production functions for Ohio's manufacturing industries we require more information about the inputs employed and outputs produced than is available from the Ohio Division of Labor Statistics annual reports alone. We have, therefore, matched our gender, occupation, and industry specific employment and wage data with information from the biennial Census of Manufactures for the state of Ohio. The census data and employment data overlap for seven years during our sample period: 1914, 1923, 1929, 1931, 1933, 1935, and 1937. Again industries are defined at the four digit SIC code level of aggregation.

For each industry-year the census provides us with information on: gross output (the value of gross output deflated by a manufacturing wholesale price index (US Historical Statistics, Series E86)); materials (the value of all materials used deflated by an intermediate materials used in manufacturing wholesale price index (US Historical Statistics, Series E79)); total employment; and; two capital proxies (the fixed capital stock figures reported in the census, and as a sensitivity check we have also used value added less the total payment to all types of labor (assuming a 50 week work year) deflated by a user cost for capital index, which is comprised of a purchase price for capital multiplied by a nominal interest rate, the GDP deflator, and an assumed 10 percent depreciation rate. Our constructed capital proxy assumes that there was a competitive market for capital in Ohio, and purchase prices, depreciation rates and tax treatments were approximately equal across industries. Other capital proxies may also be used, including the horsepower figures reported in the census, but for some years the reduction in the number of observations available is severe).

A4: Matching the Division of Labor Statistics Data to the Census Data

Croxton (1935, p. 4) tells us that the Ohio Division of Labor Statistics employment and wage data accounted for 95.2 percent of wage earners and 96.4 percent of all wage payments reported in the Census of Manufactures for the state of Ohio in 1914. However, because we wish to match individual industries in these two data sets we need to do more than simply compare aggregate coverage. To match the Ohio Division of Labor Statistics wage and employment data to the Ohio Census of Manufacturers input and output data we first identified identical four digit SIC industry titles. Because the census published information aggregated from manufacturing establishments with $500 or more in gross output, but the Ohio Division of Labor Statistics annual reports published information aggregated from establishments with five or more employees (later three or more employees), even the industries with identical titles may have included different establishments in each data set. To improve our matching procedure we then dropped industries that reported total employment figures and total payments to labor figures that differed *jointly* by more than 25 percent. We also constructed a dummy variable that took the value one for all industries for which the total employment, total labor payment *and* number of establishment figures differed jointly by more than 25 percent. All of our results were derived with and without the industries tagged by our matching dummy with no substantive changes in any of the qualitative conclusions. Where there was any difference in the total number of employees reported in the two data sets, we used the gender and occupation distributions implied by the Division of Labor Statistics data to adjust the total number of employees reported in the Census data. Over all available years, among the industries included in our matched data set the average industry in the census data reported just 0.8 percent fewer total employees and 2.07 percent more labor payments than the average industry in the Division of Labor Statistics annual reports. In total the matched data set includes 737 industry-years, with a maximum of 127 matches in 1923, and a minimum of 80 matches in 1933.

Econometric Appendix

The translog production function, described by Equation (1), is particularly desirable for our purposes because it is a second order approximation of any arbitrary, twice differentiable production function for a given input combination, and unlike the more common Cobb-Douglas or CES production functions, it does not impose any restrictions on the elasticities of substitution. Diewert (1976) describes the desirable features of the translog production function in considerable theoretical detail. The translog specification also accommodates an explicit decomposition of the total impact of technological change – broadly defined to include any increases in output that cannot be attributed to increases in measured input employment – into a neutral component that increases the efficiency of all inputs in equal proportions, and input specific "biased" components that increase the efficiency of individual inputs in isolation. A cost function approach would also be feasible given our purposes and the data we have access to, but with our translog specification we need not impose any restrictions on substitution elasticities to guarantee appropriate curvature properties.

Using all 737 industry-year observations organized into an unbalanced panel we estimate a translog production function for the entire Ohio manufacturing sector, and after disaggregating our full sample we have used 364 industry-year observations to estimate a translog production function for the Group SUS industries, and the remaining 373 industry-year observations to estimate a separate translog production function for the Group BIC industries. We use generalized least squares to derive our parameter estimates, with a correction for first order autocorrelation among the errors, and a set of seven fixed effects variables to control for heteroskedasticity within each panel, where panels are identified by industrial sector. A random effects correction for cross panel heteroscedasticity was rejected on the basis of a standard Hausman test. To test the sensitivity of our results, we have estimated systems of cost share equations, and generalized Leontief production functions, rather than the full translog production functions. These tests added virtually no statistical power to our parameter significance and they do not change any of our qualitative conclusions. Table 3 includes the full set of parameter estimates (and their standard errors) for all industry-years, the two industry group sub-samples, and a chronologically truncated sample that drops the depression years.

Table 2: Translog Parameter Estimates by Industry Group (With and Without Depression Years)

	1914-1937			1914-1929		
	All Industries	SUS Group	BIC Group	All Industries	SUS Group	BIC Group
λ_{MPL}	-3.755	-5.979	-4.683	-8.062	-14.490	-6.802
	(1.057)	(2.021)	(1.141)	(1.741)	(3.662)	(1.949)
λ_{FPL}	0.175	0.467	0.530	0.037	1.157	-0.076
	(0.334)	(0.634)	(0.302)	(0.578)	(1.124)	(0.589)
λ_{MCL}	1.484	3.213	0.628	-0.487	6.251	3.577
	(1.152)	(2.388)	(1.040)	(2.020)	(5.413)	(2.341)
λ_{FCL}	-2.337	-1.410	-1.965	-1.150	-5.416	-4.334
	(1.136)	(2.332)	(1.069)	(1.922)	(5.013)	(2.158)
λ_{K}	-0.955	-0.730	0.026	-2.927	-2.906	3.345
	(1.001)	(2.050)	(0.990)	(1.719)	(3.929)	(1.833)
λ_{M}	5.960	5.564	5.432	14.481	18.033	4.010
	(0.936)	(1.719)	(0.902)	(1.503)	(2.816)	(1.766)
β_{Yrs^2}	0.00001	0.00001	0.00001	0.00002	0.00002	0.00001
	(0.000001)	(0.000003)	(0.000002)	(0.000002)	(0.00001)	(0.000003)
β_{MPL^2}	0.066	0.058	0.070	0.068	0.070	0.072
	(0.004)	(0.007)	(0.004)	(0.005)	(0.009)	(0.005)
β_{FPL^2}	0.008	0.013	0.004	0.008	0.010	0.004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
β_{MCL^2}	0.009	0.003	0.019	-0.001	-0.002	0.005
	(0.005)	(0.009)	(0.005)	(0.006)	(0.101)	(0.011)
β_{FCL^2}	0.004	0.011	0.005	0.002	0.016	-0.001
	(0.002)	(0.004)	(0.003)	(0.003)	(0.007)	(0.006)
β_{K^2}	0.067	0.054	0.076	0.065	0.043	0.075
	(0.003)	(0.006)	(0.003)	(0.004)	(0.010)	(0.004)
β_{M^2}	0.091	0.089	0.107	0.088	0.086	0.101
	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)	(0.005)
β_{MPLt}	0.003	0.004	0.003	0.005	0.008	0.004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
β_{FPLt}	-0.0001	-0.0002	-0.0003	0.00003	-0.001	0.0001
	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.001)	(0.0003)
β_{MCLt}	-0.001	0.002	-0.0002	0.0002	-0.003	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)
β_{FCLt}	0.001	0.001	0.001	0.001	0.003	0.002
	(0.0006)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)
β_{Kt}	0.001	0.001	0.0001	0.002	0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
β_{Mt}	-0.003	-0.003	-0.003	-0.008	-0.010	-0.002
	(0.0005)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
$\beta_{MPL-FPL}$	-0.009	-0.008	-0.005	-0.012	-0.014	-0.005
	(0.002)	(0.003)	(0.002)	(0.003)	(0.005)	(0.003)
$\beta_{MPL-MCL}$	-0.013	0.011	-0.007	-0.007	0.006	-0.002
	(0.007)	(0.013)	(0.008)	(0.009)	(0.022)	(0.012)
$\beta_{MPL-FCL}$	0.007	-0.016	0.004	0.005	-0.016	-0.005
	(0.007)	(0.011)	(0.008)	(0.008)	(0.021)	(0.013)
β_{MPL-K}	-0.040	-0.020	-0.042	-0.050	-0.031	-0.062
	(0.006)	(0.009)	(0.005)	(0.007)	(0.011)	(0.009)
β_{MPL-M}	-0.076	-0.078	-0.086	-0.070	-0.071	-0.075
	(0.005)	(0.007)	(0.006)	(0.006)	(0.009)	(0.009)
$\beta_{FPL-MCL}$	0.004	0.015	0.001	0.005	0.007	0.002
	(0.002)	(0.005)	(0.002)	(0.002)	(0.010)	(0.003)
$\beta_{FPL-FCL}$	-0.007	-0.022	-0.001	-0.007	-0.015	-0.002
	(0.002)	(0.005)	(0.002)	(0.003)	(0.010)	(0.003)
β_{FPL-K}	-0.007	-0.011	-0.006	-0.005	-0.002	-0.004
	(0.002)	(0.004)	(0.002)	(0.002)	(0.005)	(0.003)
β_{FPL-M}	0.002	0.003	0.004	0.001	0.003	0.003
	(0.001)	(0.003)	(0.002)	(0.002)	(0.004)	(0.003)
$\beta_{MCL-FCL}$	-0.005	-0.001	-0.020	0.002	-0.004	0.002
	(0.005)	(0.009)	(0.007)	(0.007)	(0.010)	(0.015)
β_{MCL-K}	-0.015	-0.049	-0.014	-0.003	-0.009	-0.011
	(0.007)	(0.020)	(0.007)	(0.010)	(0.037)	(0.010)
β_{MCL-M}	0.015	0.024	0.001	0.008	-0.001	0.009
	(0.005)	(0.015)	(0.007)	(0.007)	(0.021)	(0.010)
β_{FCL-K}	0.010	0.040	0.020	0.006	0.021	0.024
	(0.007)	(0.017)	(0.007)	(0.010)	(0.034)	(0.011)
β_{FCL-M}	-0.017	-0.022	-0.010	-0.009	-0.005	-0.020
	(0.005)	(0.014)	(0.007)	(0.007)	(0.020)	(0.010)
β_{K-M}	-0.096	-0.094	-0.118	-0.094	-0.089	-0.109
	(0.004)	(0.008)	(0.005)	(0.006)	(0.012)	(0.008)
θ	-17.191	-11.871	-16.569	-57.486	-66.751	-30.141
	(5.286)	(9.690)	(5.372)	(9.007)	(19.503)	(9.508)
AR(1) Correction	✓	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓	✓
# Industry-Years	737	364	373	340	159	181

Note: MPL represents male production labor, FPL represents female production labor, MCL represents male clerical labor, FCL represents female clerical labor, K represents real fixed capital, and M represents intermediate materials. Standard errors are provided in parentheses.