4-25-2019

# Factor Biases in the Sawmilling Industry from 1850 to 1870

Nathan Williams
*College of Saint Benedict/Saint John's University*, NJWILLIAMS@CSBSJU.EDU

# Factor Biases in the Sawmilling Industry from 1850 to 1870

By Nathan Williams

A special thanks to my parents, Dr. Johnston, Dr. Wheatley, and Dr. Olson. Without their guidance, help, and support, this would not have been possible.

**D22, D24, N51**


**Abstract**

Lumber milling contributed greatly to the growth of the American economy in the nineteenth century. Scholars have done very little research with regard to the sources of the industry's growth, especially the relative importance of capital inputs versus labor inputs. The aim of this project is to determine whether nineteenth-century lumber milling technology had a bias towards using relatively more capital or labor over time. The data to analyze this industry was drawn from the Census of Manufactures for 1850, 1860, and 1870, a period of rapid technological and demographic change. A transcendental logarithmic production function was used in estimating the factor biases of capital and labor. It was found that there was a significant capital-using bias and a labor-using bias during this period but using both inputs in concert on a large scale was inefficient. These results indicate there was a bifurcation of the sawmilling industry from 1850 to 1870, with some companies choosing capital-intensive processes, whereas others used labor-intensive processes.

**Introduction**

The period of 1850 through 1870 was one of intense change in America and in the lumber milling industry. With the purpose of sawing and milling wood, thousands of lumber mills operated during the Antebellum and Postbellum periods of America. Lumber mills were dependent on location and season, requiring fast-moving water to power their machines. Due to this restriction, the growth of large sawmills across the frontier was greatly constrained. Everything changed with the implementation of the steam engine, which allowed firms to locate in places with the most natural resources. Beyond this technical change, the period of 1850 through 1870 saw immense decline in the working population in America, due to the Civil War. These changes led to shifts in the way inputs were used for the lumber milling industry from 1850 through 1870, especially with respect to capital and labor. Due to the rapid pace of technical advancement as well as a lack of available workers, there is a hypothesized capital-using bias and a labor-saving bias present in the sawmilling industry from 1850 through 1870.

Factor biases were calculated using a transcendental logarithmic function with five independent variables and one dependent variable. The data set used to estimate the relationship between these variables was constructed through the combination of three data sets, a sample of the Census of Manufactures from 1850, 1860, and 1870. Combined through a Chow Test, the data set used to estimate relationships between inputs and outputs is pooled and cross-sectional. The coefficients for the regression were estimated using this dataset, and the functional form of the regression was:

$$lnVAL_{ADD} = \beta_0 + \beta_{lnCAPITAL} \ln CAPITAL + .5\beta_{lnCAPTIAL2} lnCAPITAL2^2 + \beta_{lnWAGE_{BILL}} lnWAGE_{BILL} + .5\beta_{lnWAGE_{BILL}} lnWAGE_{BILL}{}^2 + \beta_{lnLABCAP} lnLABCAP.$$

To test for factor biases, the first order necessary conditions were calculated for both capital and labor, and the results were interpreted. This process garnered the results necessary to draw a conclusion with regard to factor biases in lumber milling from 1850 through 1870.

The following sections will test whether there were capital-using and labor-saving biases present in the lumber milling industry. The Literature Review will analyze the relevant literature with respect to the functional form of the regression, the data used, and the historical context. The Data section defines each variable used in the regression, as well as the construction of the data set. The section, Empirical Methods outlines the functional form of the regression and the constraints necessary for the industry to operate in an environment conducive with economic theory. The Estimated Regression section estimates the coefficients of the regression and estimates the factor biases of the sawmilling industry and discusses the validity of the initial presented hypothesis.

## <u>Literature Review</u>

The translog functional form is a production function, meaning it serves as an estimation tool for determining the effects that inputs have on an output. In the field of economic history, the translog is often used to estimate the factor biases present in industries through the. Using methods consistent with the fields of economic history, industrial organization, and econometrics, a regression will be estimated to measure the factor biases present in sawmilling from 1850 to 1870.

The initial application of production functions in economic history were conducted by Bateman and Weiss (1975). They used a Cobb-Douglas Production Function to estimate the returns to scale in a variety of industries in the United States in 1860. The authors were one of the first to analyze firms using the samples from the full Census of Manufactures and digitized much of the data used in the estimation process. One of the industries they chose to analyze was the lumber milling industry, and they found significant returns to scale for lumber milling in 1860 for the regions defined as the North and the South. The region classified as West did not have increasing returns to scale, due to the smaller sample size. The use of a Cobb-Douglas production function does not fully encapsulate the relationships of capital, labor, and output due to the forced constant returns to scale, but this method was the first step in using a production function to model factor biases.

Beyond calculation of returns to scale, Bateman and Wiess calculated a variety of statistics with respect to the composition of industries across the United States. One of the most important statistics, with regards to industry structure, was the calculation of firm concentration. The authors found that lumber milling was an industry that was one of the least concentrated in America in 1850 and 1860. "Comparatively low [concentration] ratios are common only in boot and shoe making, flour milling, and lumber production." (204). This suggests that lumber milling was an industry that was relatively unconcentrated, but one whose profits were highly dependent on the level of concentration due to the calculated Spearman coefficient.  This coefficient was estimated to be .821 for sawmilling in 1860 (207), indicating a strong level of rank correlation between concentration and profit. This established an environment for increasing returns to scale and greater investment in labor and capital, through consolidation of the industry throughout the next twenty years. Their work was continued by the work of their graduate assistant, Atack.

The work of Bateman and Weiss (1975) with regard to the Census of Manufactures was continued through the work of Atack (1986). Atack analyzed the firm structure of specific industries in the United States and how they changed from 1850 to 1870. The author analyzed the lumber industry extensively, analyzing how the firm size evolved over time. The geometric means and medians of the firms increased by a factor of 4 during the period 1820 to 1870. The geometric mean increased from 633 dollars to 2,460 dollars, and the median increased from 590 dollars to 2,160 dollars.  The increase of both the mean and the median indicates that the industry was rapidly increasing in size across all levels of firm output. Using the methodology of Bateman and Weiss (1975), the economic concentrations for the lumber milling industry were calculated and used to calculate a Gini Index. The lumber industry became more concentrated over time, despite the number of firms increasing, indicating a fundamental shift in the industry and the way lumber was milled. This information further supports the hypothesis presented; significant industrial change fomented through technological advancement and labor shortages led to a capital-using bias and a labor-saving bias.

Sawmilling, as an industry, was regressed due to its rapid technological change during the 19th Century, as well as its growth as an industry. Atack and Passell (1994) documented the shift of lumber milling from water power to steam engines. “But high interest rates in the West, where many mills operated, raised the cost of capital-intensive waterpower above that of steam. […] why the lumber industry used more steam power than any other in 1840.” (198, 1994). A rapid technological shift, one from waterpower to steam power, made the sawmilling industry more capitally intensive. This change fomented the use of a capital-intensive process present in during the 19th Century, which led to a capital-using and labor-saving bias.

The transcendental logarithmic production function, created by Christenson et al. (1972), serves as the basis for the research conducted to estimate factor biases in the sawmilling industry. The translog production function was used to estimate the non-linear parameterized relationship between the inputs, in this case capital and labor, and the output, the value added of an individual firm. The general form of the function was:

$$\ln Q = a_0 + \sum_i a_i \ln(X_i) + \frac{1}{2}\sum_i \sum_j \gamma_{ij} \ln(X_i) \ln(X_j)$$

(37), with $X_i$ serving as an input, and $X_j$ serving as a separate input. The translog production function was derived as a second order Taylor approximation of the production frontier. Unlike many production functions, the translog function has relatively few restrictions, making it a versatile production function for estimation purposes. Two relevant constraints imperative to estimation of the production function to ensure proper isoquant behavior were $\beta_{lnCAPITAL} + 2 * 5\beta_{lnCAPTIAL2} + \beta_{lnLABCAP} = 1$ and $\beta_{lnWAGE_{BILL}} + 2 * .5\beta_{lnWAGE_{BILL}} + \beta_{lnLABCAP} = 1$ (39). Due to the rapid consolidation of the lumber milling industry during the period 1850 to 1870 and the possibility of non-constant returns to scale across the period, the translog production function was ideal for the purposes of discussing the proffered hypothesis.

The transcendental logarithmic production function was initially created by Christensen et. al. (1972) and was applied to factor biases through the work of Binswanger (1974). Binswanger proposes two distinct transcendental logarithmic production functions to calculate factor bias, Model A and Model B. Model A does not contain a time trend variable, and Model B does. "Model A assumes that the rate of biases is not constant over time. For shorter time periods, however, it is possible to assume that the biases are constant." (967). To calculate the factor biases of industries, the author used the first order necessary condition of the estimated translog cost function and used the sign present on the independent variable being derived. The estimated regression for the estimation of factor biases in the lumber milling industry does not use a time-trend independent variable, because it is assumed that the rate of the biases is not constant over time.

Cain and Paterson (1981) used the work of Binswanger (1974) to calculate the factor biases present in 19 separate industries from the period of 1850 to 1919. Four variables were used in their estimated transcendental logarithmic cost function: capital services, labor inputs, material inputs, and other inputs. Labor inputs were constructed using both census data and sample data, and was defined as, "the sum of all wage payments multiplied by the number of hourly paid employees." (358). This construction was then weighted by the number of hours that were worked on average by employees during the period, which decreased over time. Capital was also constructed through the use of capital depreciation variables estimated through the bond appreciation rate of the New England Bank. This number was used to form an index to estimate the relationship between the use of capital and the total cost incurred by a firm. These two variables were constructed as part of a larger regression to calculate factor biases in four industries. The regression equation used closely resembles Model B in Binswanger, due to the hypothesized lack of biased change present in the chosen industries during the selected period. "[…] as expected given relative factor price movements, labor -saving biases were experienced in a large number of industries as well as capital-using." (348). The implicit assumption, that factor biases remained constant throughout the estimated period was found to be accurate, and the biases were interpreted. This paper concludes that there were capital-using and labor-saving biases present between 1850 and 1919 and served as a basis for the hypothesis presented in this paper.

Building on all previous work, James (1983) used a transcendental logarithmic production function to estimate the relationship between inputs and two inputs, capital and labor. The functional form of the equation follows closely the functional form of the final estimated regression in this paper. The functional form used by James is as follows:

$$\ln Q = a_0 + \sum_i a_i \ln(X_i) + \frac{1}{2}\sum_i \sum_j \gamma_{ij} \ln(X_i) \ln\left(X_j\right) + a_t t + \sum_i \gamma_{it} \ln(X_i) * t + \frac{1}{2}\gamma_{tt} t^2$$

This functional form has two inputs, capital and labor. These two inputs estimate their respective relationships with the value added with respect to output. The functional form of this regression served as a secondary basis for the analysis of factor biases of lumber milling from 1850 through 1870.

The variables used in the James (1983) will be used to estimate the factor biases in sawmilling from 1850 to 1870. The construction of the first variable, capital, follows the work of James closely. Capital, as defined by James, is "[…], the reported capital stock figures, adjusted for price level changes by the Warren-Pearson wholesale price index, are used as a measure of capital." (457). The capital input used in the Census of Manufactures was used, adjusted for inflation in price. The second variable, adjusted by James (1983), is the input for labor. "Wage earner totals were computed by the Census by summing twelve monthly figures, each referring to the number of full and part time workers and then dividing by twelve." (457). This variable cannot be calculated in the regression ran in this paper, due to the proprietary level at which the information is held. In order to calculate the labor input, the aggregate wage of all workers at a given firm will be used to capture the labor inputs present during the period of 1850 to 1870.

The differences between a translog cost and translog production functions are also detailed in the work of James (1983). The reason to use production functions to capture factor substitutes comes from the availability of data. “The choice in this case is clear, because of the lack of local factor prices, interest rates in particular, for most years precludes estimations of cost functions.” (453). The lack of input prices present in the data indicates that there are difficulties calculating cost functions. Cain and Paterson (1981) calculated the interest rate using a variety of indexes, but there is little need to extend the data further than necessary given the availability of information with regard to translog production functions. Therefore, the translog production function will be used to estimate the factor biases present in the lumber milling industry from 1850 to 1870.

## Data

The data used to estimate the factor biases of the sawmilling industry from 1850 through 1870 originates from the Census of Manufacturers. Started in 1850, this census measured a variety of inputs and outputs, delineated by both industry and state. Originally sampled and compiled by Atack, following the work of Bateman, Weiss, and Foust, three variables are used to estimate the relationship between value added output, capital, and labor for lumber milling during the Antebellum and Postbellum periods. The data compiled by Atack et al. originates in the Census of Manufactures.

The Census of Manufacturers was initially created in the through the Census Act of 1850, which was set to enumerate “all the products of industry of each producer or establishment” (Atack and Bateman 1999, 178). Every ten years, per the Census Act, collectors sampled every industry in the United States by region. This data is viewed with some skepticism by some, due to technological and scientific constraints of data collection at this time, specifically with respect to the ability to collect representative data. While this trepidation is noted, there was no evidence of systemic bias present in the collection methods, maintaining its lack of bias (James 1983, 456). Therefore, the data can reasonably be used to draw conclusions about the factor biases of the lumber milling industry.

The Business Census served as the cornerstone for estimating the relationships between capital, labor, and other materials to costs. This source was the subject of the paper, “Nineteenth-Century U.S.

Industrial Development through the Eyes of the Business Census", by Atack and Bateman (1999). This paper discussed much of the sampling methodology, as well as discussed problems associated with using this dataset, which is vital for understand the consistency of variables included in the regression and consistent with other scholarly works.

The sampling methodology for 1850, 1860, and 1870 were similar, and can be compared in a way that reasonably limits the amount of exogenous variation present in the model. Samples from each state were drawn at random from industries, with a goal of between 200 and 300 firms per state. In 1850, there were 8,776 firms in the initial sample, and they were only taken from states, territories were not collected. In 1860 and 1870, respectively, there were 8,828 and 9,964 firms (184). The state level data was then randomly sampled, and a national dataset was achieved, which is what is used to estimate the transcendental logarithmic cost function. Specifications of variables is important in the understanding of the coefficients estimated from the translog cost function for the lumber milling industry. Each industry was delineated by a specific 3-digit code, one that made it easy to identify the industry to which the data belonged. For lumber milling, the specific industry code from 1850 to 1870 is 242.

The dependent variable, value added, is a constructed variable to capture the output of a given firm. It is constructed from subtracting the Input variable by the Output variable and transforming it by the natural logarithm. Output is defined as, "Aggregate values of products, including jobbing and repairing" (Atack 1999, 182). The variable, Input, is described as the, "Aggregate values of raw materials including mill supplies and fuel" (182). The input variable is then subtracted from the output variable, to form the value-added dependent variable. It is then transformed by the natural logarithm, to make it unitless and consistent with a translog function, where all major variables are transformed by the natural logarithm.  This variable is consistent with work done in the field of translog production function modeling. Given that this variable approximates the value added of a given firm, as well as remains consistent with other work done in the field, this variable is consistent for a dependent variable for a translog production function.

Atack and Bateman define each variable present in the dataset, indicating the nature of what was measured. The first pertinent variable defined is capital, defined in the Business Census as, "Capital invested, in real and personal estate, in business." (182). This definition of capital is broad, and leaves room for interpretation that could potentially be detrimental to the validity of the estimated coefficients of the translog cost function, but the specificity cannot be controlled, so it is an issue that must be accepted as a reality of working with the data. James (1983) argues that the amount of error present in the data is often overstated. "There is no evidence of systemic biases in the published reports, such as the omission of smaller firms, so estimated coefficients should be unaffected […]" (456). The estimated coefficients are not affected in a large way, because there is no evidence of systemic bias present the sampling.

Labor, in the Business Census, is broken down to a number of fields, but the variable of interest is "Wage Bill". In the context of the Census of Manufactures, "Wage Bill" is defined as, "Total amount of wages paid in a year" (182). This variable serves as an adequate measure of labor inputs, being the same input measured in Cain and Paterson (1981).

Descriptive statistics were calculated for three unmanipulated variables of note, the total wage bill, capital invested, and the value-added output of a firm for the sawmilling industry. The statistics are listed in the table below:

Table 1: Descriptive Statistics

| | CAPITAL | LABOR | VALUE ADDED |
|---|---|---|---|
| Mean | 3,672.0877 | 1,199.6786 | 3,381.4026 |
| Standard Error | 246.4190 | 62.8956 | 319.1370 |
| Median | 1,500.0000 | 480.0000 | 950.0000 |
| Mode | 1,000.0000 | 240.0000 | 600.0000 |
| Standard Deviation | 11,735.3401 | 2,970.7764 | 15,215.1712 |
| Sample Variance | 137,718,207.9248 | 8,825,512.6621 | 231,501,433.4607 |
| Kurtosis | 172.9147 | 83.0299 | 462.5895 |
| Skewness | 11.5197 | 7.8645 | 18.8797 |
| Range | 250,000.0000 | 46,000.0000 | 465,000.0000 |
| Minimum | 0.0000 | 0.0000 | -5,000.0000 |
| Maximum | 250,000.0000 | 46,000.0000 | 460,000.0000 |
| Sum | 8,328,295.0000 | 2,676,483.0000 | 7,685,928.0000 |
| Count | 2,268.0000 | 2,231.0000 | 2,273.0000 |

The average value-added output in each firm from 1850 through 1870 was $3,381.40, with a median value of $950. This indicated a high-degree of bias, specifically a right-tail bias. This wide diversity in lumber milling firms is further corroborated by the standard deviation, $15,215.17. The kurtosis indicated that there are infrequent, extreme outliers present in the data specifically with respect to the value added of a given firm. The range of the data is $465,000, indicating a wide variety of firms present in the lumber milling industry.

Capital inputs and labor inputs have means of $3,672.08 and $1,119.69, respectively. The kurtosis for both of these indicates that there were infrequent, extreme outliers. The kurtosis, as well as the means, indicated that there were large variations present in the data set. The data indicated that there was a disparity present in the lumber milling industry from 1850 to 1870. With some extremely large firms and others being very small, there was variation present in the data set from 1850 through 1870.

One of the two major independent variables is capital. "lnCAP", as it is labeled within the regression, is defined as, "Capital invested, in real and personal estate, in business" (182). This variable is transformed by the natural logarithm, to maintain consistency with the form of the translog production function. In addition to the base capital variable, there was also a quadratic term, "lnCAP2", to capture non-linear changes in the elasticity of production for capital. The construction of both of these variables are consistent with those in the field and serve as an adequate measure of capital inputs for a firm from 1850 to 1870.

The relationship between labor and the value added of a given firm will be estimated through the two variables, "WAGE_BILL" and "WAGE_BILL2". The former variable is used to calculate the constant returns to labor, and latter variable is used to calculate the non-linear returns to labor. The definition of this variable is, "Total amount paid in wages for a year" (Atack and Bateman 1999, 182). The variable is transformed by the natural logarithm to maintain consistency with the translog production function, and the quadratic term is manipulated in a way consistent with the translog production function.

"WAGE_BILL" and "WAGE_BILL2" serve as adequate estimates of labor inputs in the sawmilling industry from 1850 to 1870.

The final independent variable used to estimate the relationship between capital, labor, and value-added output is the interaction term between capital and labor, "lnLABCAP". Constructed from multiplying the natural logarithm of capital and the natural logarithm of labor together and was used to calculate the interrelated nature of capital and labor. Consistent with the specifications of the translog production function, this variable is the final independent variable present in a two-input translog production function.

Every variable in the data set was adjusted to account for inflation and fluctuation of capital prices. The entire data set was adjusted by the estimated Consumer Price Index and Domestic Capital Price Index, to account for rapid inflation present between the Civil War and the Reconstruction Period (Gallman 1992, 88). 1860 serves as the base year for the indexes calculated (88). The adjusting of each Census, from 1850 to 1870, allows the regression to be estimated using 1860 as a base year. Appendix One contains a more complete explanation of the construction of variables and statistical methods used for compiling the data.

**Empirical Methodology**

Estimation of a translog production function requires a variety of restrictions to maintain statistical saliency in terms of both the data set and the estimation process. This section will discuss the merging of specific data sets as they exist in their current functional form, as well as the constraints necessary to use a translog production function in this compacity.

The relationship between inputs and outputs was estimated using a transcendental logarithmic production function, one that is similar to the initially derived function in Christenson et al. (1976) and employed by James (1983). A translog production function was used to estimate the relationship between the dependent variable, the logarithm of value-added output, and the independent variables, the logarithm of dollars spent on capital and the logarithm of number of employees at a given firm. The functional form of the regression is as follows:

$$lnVAL_{ADD} = \beta_0 + \beta_{lnCAPITAL} \ln CAPITAL + .5\beta_{lnCAPTIAL2} lnCAPITAL2^2 + \beta_{lnWAGE_{BILL}} lnWAGE_{BILL} + .5\beta_{lnWAGE_{BILL}} lnWAGE_{BILL}{}^2 + \beta_{lnLABCAP} lnLABCAP$$

This functional form was ideal for measuring the factor biases of the lumber milling industry from 1850 through 1870.

The translog production function was chosen due to its lack of constraints. This type of function is a second-order Taylor polynomial, so it avoids returns to scale restrictions. With functional forms such as Cobb-Douglas, returns to scale are constrained to be constant, but with a translog production function, the returns to scale can be increasing, decreasing, or constant. Because factor biases are measured through the first order necessary conditions of each factor of production, this lack of restriction is important in correct estimation of factor biases of the lumber milling industry during this time period.

Due to the lack of restrictions present in the model, it was imperative that some constraints be employed to ensure the isoquants of the market behave in a way that is conducive to economic theory. Three constraints were imposed on the function:

$$\beta_{lnCAPITAL} + 2 * 5\beta_{lnCAPTIAL2} + \beta_{lnLABCAP} = 1$$

$$\beta_{lnWAGE_{BILL}} + 2 * .5\beta_{lnWAGE_{BILL}} + \beta_{lnLABCAP} = 1$$

$$\beta_{lnCAPITAL} + \beta_{lnWAGE_{BILL}} = 1$$

These three constraints were used to overcome specific difficulties with using the translog production function and were suggested by Christensen et al. (1972) and Boisvert (1982). These restrictions imposed perfect competition on the market, as well as forced isoquant curvature to be consistent with economic theory. With these restrictions in mind, the production function can be estimated for the sawmilling industry from 1850 to 1870.

**Estimated Regression**

The factor biases present in the sawmilling industry were estimated using a translog production function and a pooled, cross-sectional data set from 1850 to 1870. This section will estimate the coefficients with respect to each independent variable and will calculate the factor biases present in the sawmilling industry.

The results for the regression with 939 observations are in Table Two, listed below:

| Table 2: Regression Results | | | | | | |
|---|---|---|---|---|---|---|
| | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. Interval] | |
| lnCAP | 0.7532 | 0.0115 | 65.38 | 0 | 0.7306 | 0.7758 |
| lnCAP2 | 0.3659 | 0.0062 | 59.29 | 0 | 0.3538 | 0.3780 |
| lnLAB | 0.2468 | 0.0115 | 21.42 | 0 | 0.2242 | 0.2694 |
| lnLAB2 | 0.6191 | 0.0058 | 105.99 | 0 | 0.6077 | 0.6306 |
| lnLABCAP | -0.4851 | 0.0034 | -141.69 | 0 | -0.4918 | -0.4783 |
| Constant | 0.6934 | 0.0853 | 8.13 | 0 | 0.5260 | 0.8608 |

Contextualized within the regression, the results were:

$$\widehat{lnVAL_{ADD}} = .6934 + .7532 \ln CAPITAL + .5(.3659) lnCAPITAL2^2 + .2467 lnLAB$$
$$+ .5(.6191) lnLAB^2 - .4850 lnLABCAP$$

The two variables that capture solely capital inputs, "lnCAP" and "lnCAP2", had hypotheses of $H_0\colon\ \beta_{lnCAP} = 0, H_A\colon\ \beta_{lnCAP} > 0$ and $H_0\colon\ \beta_{lnCAP2} = 0, H_A\colon\ \beta_{lnCAP2} > 0$. The alternative hypotheses for both "lnCAP" and "lnCAP2" are greater than zero because increasing returns to capital are expected due to the advent of the steam engine. The p-values for both of these variables is ostensibly zero, indicating that these two coefficients are statistically significant at all conventional levels. As capital inputs increase by one percent, value added production increased by percentages dictated by the equation:

$$.7532 + .3659 lnCAPITAL.$$

Labor inputs were represented through two independent variables, "lnLAB" and "lnLAB2". The hypotheses for these variables are $H_0\colon\ \beta_{lnLAB} = 0, H_A\colon\ \beta_{lnLAB} > 0$ and $H_0\colon\ \beta_{lnLAB2} = 0, H_A\colon\ \beta_{lnLAB} <$

$0$. A negative sign is expected from the quadratic term of labor because a labor-saving bias is hypothesized. The p-value for both of these terms are ostensibly zero, indicating that the coefficients on the natural logarithm of the total wage and its quadratic manipulation are statistically significant at all conventional levels. This runs contrary to previous work done in calculation of factor biases for the lumber milling industry but has statistical significance and has a historical backing. For every one percent that the wages paid in each firm increased, value-added production increased by the function:

$$.2467 + .6191lnLAB.$$

The final independent variable, “lnLABCAP”, captures the relationship between capital inputs and labor inputs present in the model. The hypotheses for the estimated coefficient are $H_0$: $\beta_{lnLABCAP} = 0, H_A$: $\beta_{lnLABCAP} < 0$. There is a hypothesized negative relationship between labor and capital due to the hypothesized factor biases present during this time period. The p-value for the coefficient of the interaction term is functionally zero, indicating that the coefficient is statistically significantly different than zero at all conventional levels. The estimated relationship between capital and labor is -.4850; for every one percent that capital outputs increase, the output with respect to labor decreases by .4850 percent, and vise-versa. With all the variable coefficients analyzed, the factor biases can be discussed.

The factor biases were calculated using the first order necessary condition with respect to both capital and labor. The results are listed below:

$$\frac{\partial ln\widehat{VAL}_{ADD}}{\partial lnCAPITAL} = .7532 + .3659lnCAPITAL2 - .4850lnLAB$$

$$\frac{\partial ln\widehat{VAL}_{ADD}}{\partial lnLAB} = .2467 + .6191lnWAGE2 - .4850lnCAP$$

The partial derivative with respect to capital inputs indicates a positive relationship with respect to the slope of the production function with respect to capital and the amount of capital inputs present in a given firm. This indicates that there is a capital using bias present in the sawmilling industry from 1850 to 1870. The negative coefficient on the interaction term indicates that there was a productive penalty for increasing both capital and labor simultaneously. Using the means calculated for the sawmilling industry from 1850 through 1870, the increase of a dollar in the amount of capital at a firm led to an increase of $27.71 in value added output. This divergence between capital and labor had an impact on the way that firms would have chosen to allocate resources.

The partial derivative with respect to labor indicates a positive relationship between labor and the slope of the value-added output of a sawmill from 1850 to 1870. The coefficients with respect to labor and the quadratic term for labor are both positive, and the interaction term is negative. This leads to the conclusion that there were factor biases present in the sawmilling industry from 1850 to 1870, that being a capital-using and labor-using bias, but both did not coexist in the same firm. Using the means of the calculated independent variables, for one dollar invested in labor inputs, there was a $155.15 increase in the value added output for a given firm. Appendix Two contains a visual explanation of the results found in this regression. From these results, firms either chose to specialize in capital-intensive processes or labor-intensive processes, but not in both.

With respect to functional form specifications, two types of error are specifically of note and necessary to discuss in the post-estimation phase of the regression, those issues being heteroskedasticity and multicollinearity. Starting with heteroskedasticity, the test used is the Breusch-Pagan, and the error terms will be found to be heteroskedastic if they exhibit a non-constant variance. This test uses a Chi-squared distribution, and the hypotheses that will be tested are $H_0\text{: } Homoskedastic\ Error\ Term;\ H_A\text{: } Heteroskedastic\ Error\ Term.$ A Breusch-Pagan heteroskedasticity test was used, as opposed to a White test for heteroskedasticity, because it is more general. A White test for heteroskedasticity would only capture a linear relationship between the error terms and their estimations. The first step to testing heteroskedasticity is to plot the residuals calculated in the regression as a function of the estimated output. This graph is included in Appendix Three and serves as a visual aid to indicate that there may have been some level of bias, but this apparent bias was estimated through the Breusch-Pagan test.

The process for using a Breusch-Pagan test for heteroskedasticity begins with estimating the residuals present for each observation. After the individual errors were calculated, each of the residuals was squared and used as the dependent variable in a new regression. The regression used to test the hypothesis offered by the Breusch-Pagan test was:

$$e_i^2 = \beta_0 + \beta_{lnCAPITAL} \ln CAPITAL + .5\beta_{lnCAPTIAL2} lnCAPITAL2^2 + \beta_{lnWAGE_{BILL}} lnWAGE_{BILL} + .5\beta_{lnWAGE_{BILL}} lnWAGE_{BILL}{}^2 + \beta_{lnLABCAP} lnLABCAP$$

The R-squared statistic from the regression estimated is .3879. This is multiplied by the number of observations in the data set, 939, to yield the Chi-squared distribution statistic, 364.23. The p-value for this calculated statistic is very close to one, indicating a failure to reject the null hypothesis at all conventional levels. Therefore, the error term is homoscedastic at all conventional statistical levels, and the error terms are normally distributed across the sample.

As for multicollinearity, there is a certain amount of multicollinearity present in any transcendental logarithmic production function, in its unconstrained form. The multicollinearity of the independent variables in the model means that there is a high degree of correlation between independent variables. These problems often manifest when working with regressions with many constructed quadratic and interaction terms, so these problems are expected in the context of the regression estimated.

The results from the correlation coefficient calculation are listed below:

| | Table 3: Correlation Coeffiecent Table | | | | | |
|---|---|---|---|---|---|---|
| e(V) | lnCAP | lnCAP2 | lnLAB | lnLAB2 | lnLABCAP | _cons |
| lnCAP | 1 | | | | | |
| lnCAP2 | -0.9612 | 1 | | | | |
| lnLAB | -1 | 0.9612 | 1 | | | |
| lnLAB2 | 0.9566 | -0.839 | -0.9566 | 1 | | |
| lnLABCAP | 0.1006 | -0.3712 | -0.1006 | -0.1939 | 1 | |
| _cons | 0.2689 | -0.0048 | -0.2689 | 0.5253 | -0.8878 | 1 |

These results indicate an expectedly strong correlation between each of the independent variables, with most of the correlations existing close to one for quadratic terms. These results are expected, given the presence of quadratic and interaction term variables.

Multicollinearity is a functional form specification problem, meaning that it can be traced back to the types of independent variables present in the regression. The functional form specification that was used in the translog production function was one that has been agreed upon by the field of Cliometrics. While these results are not ideal, one of the translog production function weaknesses is the presence of multicollinearity, so it is expected. Appendix Four contains the Variance Inflation Statistics for further discussion.

Despite the presence of multicollinearity in the model, the statistical findings still are valid, that being that there was a statistically significant capital-using and labor-using bias present in the sawmilling industry from 1850 through 1870. These results are not only significant to the hypothesis presented, but also to economic history as a whole.

**Conclusion**

Factor biases in the sawmilling industry existed from 1850 to 1870. Using a transcendental logarithmic production function, there were clear increasing returns with respect to both capital and labor. Calculation of the first-order necessary condition with respect to both capital and labor yielded positive terms for all variables present in the regression, with the exception of the interaction term for capital and labor inputs. These results ran contrary to much of the research on the calculation of factor biases, but the results indicate a far more compelling and historically sound narrative. It appears there was a divergence in the lumber milling industry, with some firms producing goods using many people, where other firms in the industry produced using mostly capital. However, neither used both in large amounts. The sign on the interaction term indicates a divergence, that firms either chose to use capital-intensive processes or labor-intensive processes to produce lumber from 1850 to 1870.

This is an important finding in the field of economic history. The work done in this field has not been revisited to a large degree since the works of James (1983), and the results with respect to lumber milling differ drastically from the results found in the regression estimated in this paper. While a different technical method was used in James, employing a time-series variable, the work incorporated by James had very little statistical saliency. Employing the Census of Manufacturers data from 1880 and 1890 in conjunction with 1850 through 1870 holds very little statistical merit, due to the lack of congruency with respect to definitions of variables (Atack, 1999). The regression estimated in this paper also did not use a time-series, so biased technical change could be captured by the coefficients present in the regression, as opposed to by a time-trend variable.

With more time, a greater sample of the data set would have behooved the research and conclusions presented through the estimated regression. Resampling the data from the initial Censuses of Manufacturers may allow conclusions that are more impactful, and overcome the statistical imperfections present in other estimations of the factor biases of the sawmilling industry from 1850 to 1890. Increasing the period analyzed by twenty years may create a more complete conclusion with respect to factor biases in the sawmilling industry.

**Bibliography**


Atack, Jeremy. "Firm Size and Industrial Structure in the United States During the Nineteenth Century." *The Journal of Economic History* 46, no. 2 (1986): 463-75. http://www.jstor.org/stable/2122177.

Atack, Jeremy, and Fred Bateman. "Nineteenth-Century U.S. Industrial Development through the Eyes of the Census of Manufactures a New Resource for Historical Research." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 32, no. 4 (1999): 177-88. http://www.jstor.org/stable/2119163.

Atack, Jeremy, and Peter Passell. *A New Economic View of American History: From Colonial Times to 1940*. 2nd ed. New York: W.W. Norton, 1994.

Bateman, Fred, and Thomas Weiss. "Comparative Regional Development in Antebellum Manufacturing." *The Journal of Economic History* 35, no. 1 (1975): 182-208. http://www.jstor.org/stable/2119163.

Binswanger, Hans P. "The Measurement of Technical Change Biases with Many Factors of Production." *The American Economic Review* 64, no. 6 (1974): 964-76. http://www.jstor.org/stable/1815245.

Boisvert, Richard N. "The Translog Production Function: Its Properties, Its Several Interpretations and Estimation Problems." N.P. (1982): 1-27

Cain, Louis P., and Donald G. Paterson. "Factor Biases and Technical Change in Manufacturing: The American System, 1850-1919." *The Journal of Economic History* 41, no. 2 (1981): 341-60. http://www.jstor.org/stable/2120979.

Christensen, Laurits R., Dale W. Jorgenson, and Lawrence J. Lau. "Transcendental Logarithmic Production Frontiers." *The Review of Economics and Statistics* 55, no. 1 (1973): 28-45. doi:10.2307/1927992.

Gallman, Robert E., and John Joseph Wallis. "American Economic Growth and Standards of Living before the Civil War." *National Bureau of Economic Research*, 1992, 79-120.

James, John A. "Structural Change in American Manufacturing, 1850-1890." *The Journal of Economic History* 43, no. 2 (1983): 433-59. http://www.jstor.org/stable/2120831.

## Appendix One: Statistical Methods

The data source used in the estimation of the translog production function for the lumber milling industry from 1850 to 1870 was the Census of Manufacturers. This data set was sampled by Bateman and Wiess for 1850, and Atack for 1860 and 1870. Each individual Census was combined into one pooled, cross-section data set, which is consistent with work done in the field of cliometrics.

The pooling of the cross-sectional data sets for regression purposes was confirmed through the use of a Chow Test for Structural Change. This test had to be run two times, one time to verify the pooling between 1850 and 1860, and the other to verify the pooling between the combined dataset for 1850 and 1860 alongside 1870. The formula for the Chow Test that was used is:

$$\frac{\frac{RSS_{Combined\ Regression} - RSS_{Regresion\ 1} + RSS_{Regresion\ 2}}{k}}{\frac{RSS_{Regresion\ 1} + RSS_{Regresion\ 2}}{n_{Regresion\ 1} + n_{Regresion\ 2} - 2k}} = Chow\ Statistic$$

The general hypotheses used in a Chow Test are $H_0 = No\ Structural\ Change\ Present;\ H_A = Structural\ Change\ Present$. The Chow statistic uses an F-distribution, and the rejection threshold is at the five-percent level. The calculated Chow Statistics are listed in the table below:

Table 4: Chow Statistics

| | Chow Statistic |
|---|---|
| 1850-1860 | 1.301383774 |
| 1850-1860-1870 | 0.360811852 |

The calculated Chow Statistics for both the data sets indicates a failure to reject the null hypotheses at all conventional levels. At all conventional levels, there is no structural difference between the data from the 1850, 1860, and 1870 Censuses of Manufactures. Therefore, each of the data sets can be pooled and regressed together, as opposed to separately.

The data was manipulated from the general format of the descriptive statistics in order to encapsulate the specifications of the translog production function and the inherent variations present in the three Censuses of Manufacturers. The first data manipulation used was an adjusting of the data from 1850 through 1870 with respect to two indexes, to account for inflation and variations in capital prices. The Consumer Price Index and the Material Goods Price Index from 1850 through 1870 were calculated by Gallman (1992). These two measures are listed below, in Table 3. The data from 1850 was divided by .94, the data from 1860 was divided by 1, and the data from 1870 was divided by 1.57 and 1.27, for value added, wage bill, and capital inputs respectively.

Table 5: Adjusted Indexes

| | Consumer Price Index | Material Price Index |
|---|---|---|
| 1850 | 94 | 94 |
| 1860 | 100 | 100 |
| 1870 | 157 | 127 |

After each data set was adjusted to 1860 dollars, with respect to both capital, wage, input, and output, the variables were transformed to meet the specifications of the translog production function. The first variable constructed was the dependent variable, the value-added output of a firm. This variable was constructed by taking the adjusted dollar value of output and subtracting it by the adjusted dollar value of inputs. All three of the variables, value added, capital, and wage, are then transformed by the natural logarithm to maintain consistency with the translog production function.

Beyond the transformation by the natural logarithm, the quadratic terms and the interaction terms were constructed to encapsulate the non-zero returns with respect to capital and labor, as well as their interaction. To construct the quadratic terms “lnCAP2” and “lnLAB2”, the variables “lnCAP” and “lnLAB” were multiplied by themselves, to yield a quadratic term. For ease of interpretation with respect to the first order necessary condition, each variable was multiplied by .5. The interaction term was generated through multiplying the two variables “lnLAB” and “lnCAP” together. These variables are adequate constructions to use to estimate the relationships present in the translog production function.

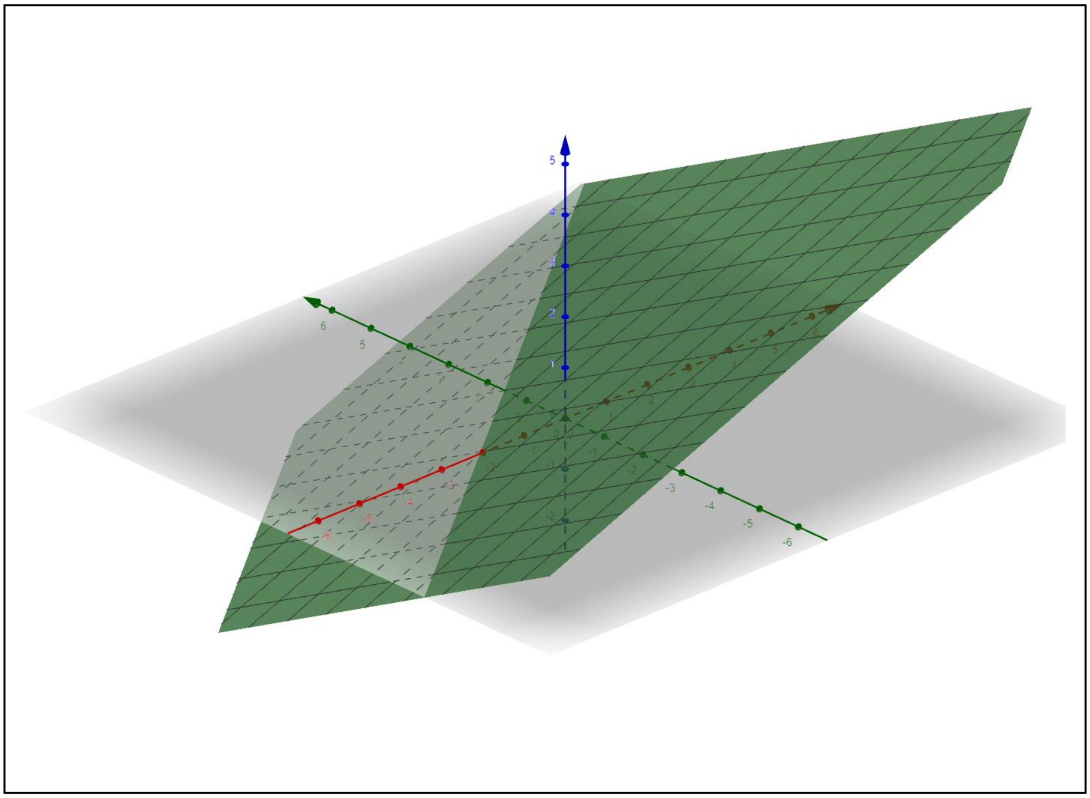
The primary method for calculating the regressions themselves was STATA, a statistical software common in the field of economics. Regressions were calculated using program files, otherwise known as .do files. The .do format used on the final regression is included below, for ease of replicability.

```
**GENERATE**
generate VAL_ADD=OUTPUT_VAL2 - INPUT_VAL2
generate lnCAP=log(CAP2)
generate lnLAB=log(WAGE2)
generate lnVAL_ADD=log(VAL_ADD)
generate lnCAP2=.5*lnCAP^2
generate lnLAB2=.5*lnLAB^2
generate lnLABCAP=lnCAP*lnLAB

**REGRESS**
constraint 1 lnCAP + 2*lnCAP2 + lnLABCAP = 1
constraint 2 lnLAB + 2*lnLAB2 + lnLABCAP = 1
constraint 3 lnLAB + lnCAP = 1
cnsreg lnVAL_ADD lnCAP lnCAP2 lnLAB lnLAB2 lnLABCAP, c(1-3)
```

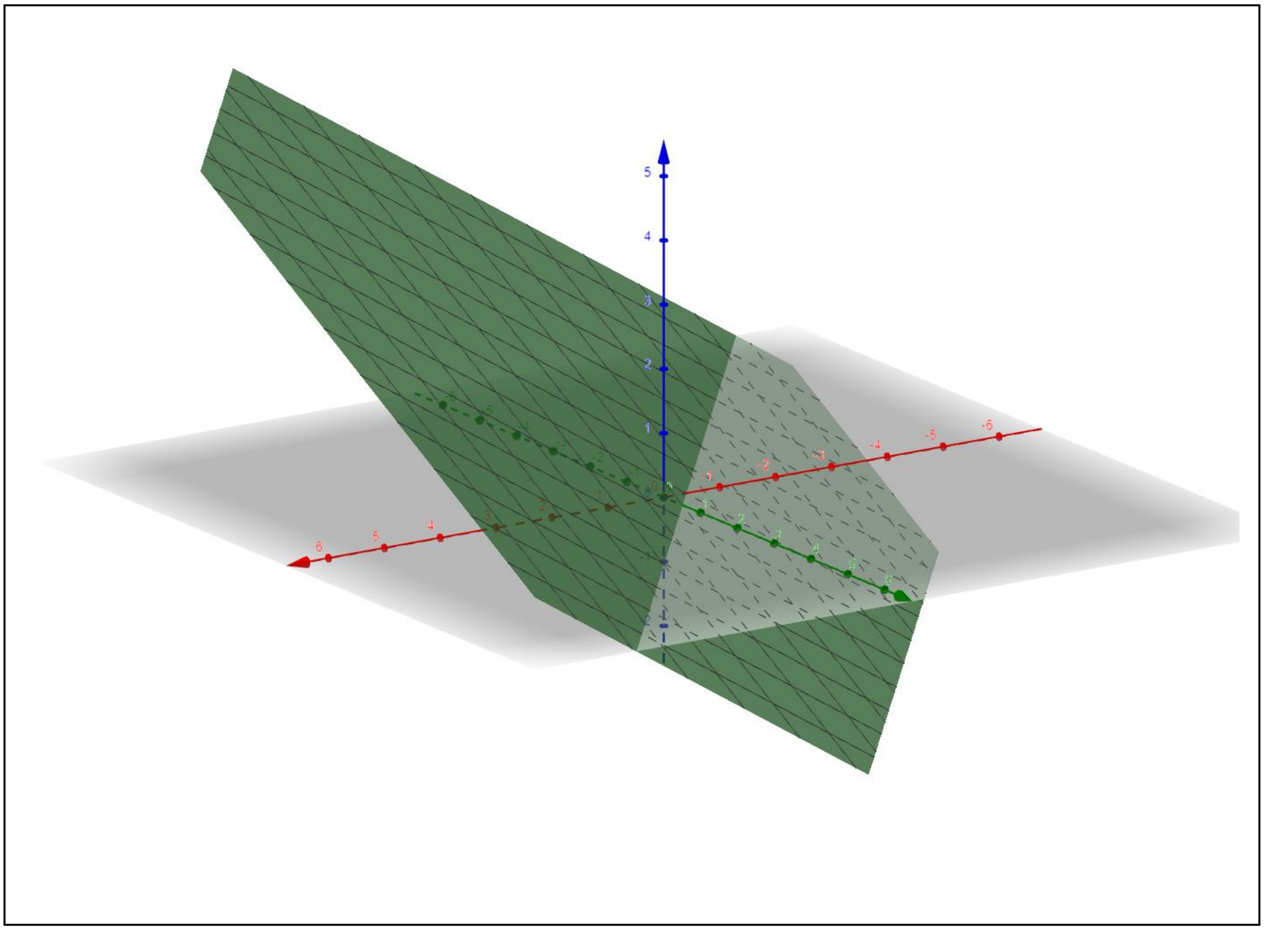
**Appendix Two: Factor Bias Graphs**

Graph of First Order Necessary Condition with Respect to Capital



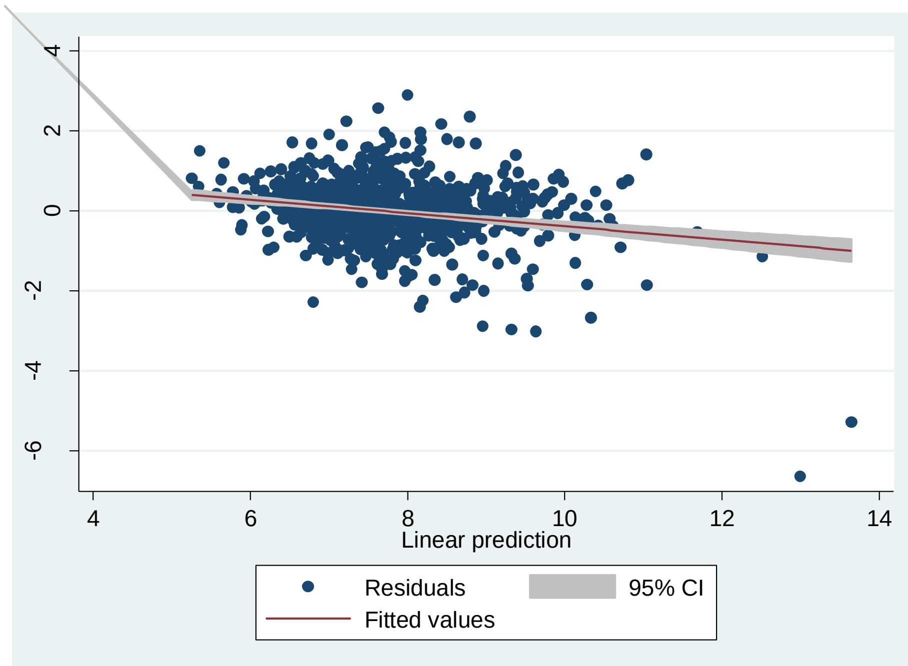
Capital is expressed on the x-axis and labor inputs are expressed on the y axis, with the z-axis being the slope of the production function. The factor biases are evident within the graph, as the amount of capital increases, the slope of the production function increases with respect to capital. With respect to labor, the slope of the production function with respect to capital decreases.

Graph of the First Order Necessary Condition with Respect to Labor



The x-axis for this graph is labor inputs, the y-axis is the capital inputs, and the z-axis is the slope of the production function. The factor bias here is evident, with respect to labor. As the percentage of labor inputs increased, the slope of the production function increases, but, because symmetry is imposed on the production function, the returns to capital with respect to labor decrease the slope of the returns. These graphs reinforce the divergence of capital-intensive production processes and labor-intensive processes.

**Appendix 3: Heteroskedasticity Graph**




**Appendix 4: Variance Inflation Factors**

These variance inflation factors were not possible to calculate given the constraints present on my regression. The entire regression must be run simultaneously in order for the constraints to work, it cannot be run in a piecewise fashion. Therefore, the more powerful test for the effects of multicollinearity, the variance inflation factor, cannot be discussed. Instead, I tested the VIF statistics for the unconstrained regression. I felt that this would be appropriate to show that I could run the proper test, but it makes very little sense in the context of my full, constrained regression. The table below encompasses all relevant VIF statistics for the unconstrained model:

| Table 6: Variance Inflation Factors | | |
|---|---|---|
| | VIF | 1/VIF |
| lnLABCAP | 414.86 | 0.00241 |
| lnLAB2 | 171.1 | 0.005844 |
| lnCAP2 | 167.41 | 0.005973 |
| lnCAP | 94.98 | 0.010528 |
| lnLAB | 79.96 | 0.012506 |
| | | |
| Mean VIF | 185.66 | |

The multicollinearity present in the discussion in the main body of the paper is extremely present here. A VIF statistic of 10 is considered to have extreme multicollinearity, and all calculated statistics achieve far above this level. Therefore, in the unconstrained regression, there will be a lack of statistical significance due to multicollinearity-induced inflation. However, given that all the independent variables in the estimated constrained regression have a high degree of statistical significance, it is safe to infer that multicollinearity has not taken a notable effect on the estimated coefficients.