EVIDENCE ON IO TECHNOLOGY ASSUMPTIONS FROM THE LONGITUDINAL RESEARCH DATABASE

By

Joe Mattey*

CES 93-8 May 1993

information about the series should contact Sang V. Nguyen, Editor, Discussion Papers, Center for Economic Studies, Room 1587, FB 3, U.S. Bureau of the Census, Washington, D.C. 20233-6300, (301) 763-2065.

Abstract

This paper investigates whether a popular IO technology assumption, the commodity technology model, is appropriate for specific United States manufacturing industries, using data on product composition and use of intermediates by individual plants from the Census Longitudinal Research Database. Extant empirical research has suggested the rejection of this model, owing to the implication of aggregate data that negative inputs are required to make particular goods. The plant-level data explored here suggest that much of the reject of the commodity technology model from aggregate data was spurious; problematic entries in industry-level IO tables generally have a very low Census content. However, among the other industries for which census data on specified materials use is available, there is a sound statistical basis for rejecting the commodity technology model in about one-third of the cases: a novel econometric test demonstrates a fundamental heterogeneity of materials use among plants that only produce the primary products of the industry.

Keyword: input-output, materials use, industry classification

*Board of Governors of Federal Reserve Systems and research associate at the Center for Economic Studies (CES), U.S. Bureau of the Census. This paper presents the author's own views, not those of the Federal Reserve System or the Census Bureau. Many individuals gave helpful advice on data, econometric, and more general issues, including Joe Baal, Robert McGuckin, Mike Mohr, Mark Planting, Paula Young and several CES staff members. The research assistance of Mark Rodini is gratefully acknowledged.

1 Introduction

In the input-output literature, a "technology assumptions" is a means for disentangling the requirements for material inputs to meet a given final demand vector from observed aggregate data on the make and use of commodities by industries. The presence of secondary production can render understanding the relation between material inputs and the product of specific commodities more difficult. In a given industry, many establishments are likely to produce more than one type of commodity (Streitwieser (1991)), including commodities classified as primary to other industries[1]. In the benchmark input-output accounts for the United States of BEA (1991), the use table gives only information on the composition of inputs for the industry as a whole and one cannot infer from this data the composition of inputs to the product of the specific commodities made in the industry. A technology assumption is a means for disentangling the input

---

[1] In addition to describing the extent of secondary production, Streitwieser investigates product composition in a context where the ownership of the plant is known; she finds that grouping by common ownership is important in discerning patterns in product mix.

structure for both primary and secondary products from the convolution of data in the use table.

The BEA (1991) presents a direct requirements matrix for the United States that embodies the industry technology assumption, as if the plants in a industry use a fixed "recipe" to make the bundle of commodities they produce; the material input requirements for a given commodity are assumed to depend only on the industry affiliation of the plants making that commodity, no on the nature of the commodity per se. In a series of papers, ten Raa and coauthors (1984, 1988, 1989) have argued that the industry technology assumption is fundamentally unsound and explored the viability of the commodity technology assumption, which proceeds as if the "recipe" for production of a commodity depends only on the nature of the commodity, not on the industry affiliation of the plant in which it is made.

The empirical work in this research on IO technology assumptions generally has been limited to data pertaining to totals for all establishments within an industry. One cannot discern any variation in the product or input mix from such data, so competing technology assumptions were evaluated on grounds other than the correlations between product and input composition. For example, ten Raa, Chakraborty and Small (1984) proceed as if the presence of negatives in the Leontief inverse is sufficient for rejecting a technology Leontief inverse by defining a mixed technology model that also allows for the presence of byproducts. However, not all of the negatives can be eliminated this way, so ten RAA (1988) and ten Raa and van de Ploeg (1989) consider a statistical approach that allows for the possibility of measurement error in the published make and use tables. Their work assumes that the published tables were unbiased but imprecise estimators of the true make and use of commodities by industries. Using subjective estimates of the imprecision of specific coefficients, ten Raa (1988) and ten Raa and van de Ploeg (1989) found that the kind of reallocation of entries in make and use tables needed to eliminate negatives were implausible. Implicitly, plant-level technologies were treated as identical. Here, I take a more direct statistical approach to examining the problem of negatives, suing the variation across plants in reported product and input mixes to test the commodity technology assumption. Measurement error is modelled as the result of nonreporting of specified materials use by particular plants[2]. Where reports on specified materials are available, the commodity technology model is estimated from the distributed of material input intensities among plants that are 'pure' in the sense that they make only the characteristics commodities of the

---

[2] If this study pertained to non-Census years, incorrect identification of industry affiliation would be another possible source of large measurement errors in the industry-level data. As emphasized by McGuckin and Peck (1992), the reclassification of establishments that occurs in Census years-when detailed product composition is more accurately reported-significantly affects the time-series properties of industry-level data.

industry[3]. The commodity technology model is tested by seeing whether the material use patterns of pure plants are relatively homogeneous.

To preview the empirical results, I find that rejections of the commodity technology model from aggregate data generally are for the wrong reasons; the problematic entries in industry-level IO tables generally have a very low Census content and do not provide a sound statistical basis for the rejection of the commodity technology model. Second, I find that, when available, sometimes the micro-data do provide a sound reason for rejection the commodity technology model; there is substantial heterogeneity of materials use among plants that only produces the primary products of an industry.

The next section of the paper reviews how industry-level aggregate data can generate the problem if negatives. Using plant-level data, the third section examines possible explanations for the problem of negatives. One of the major findings is that a lot of heterogeneity underlies the aggregate in published input-output use tables, and the concluding remarks point to further work that would be useful for understanding and coping with the implications of this heterogeneity.

---

[3] In constructing an index of diversification, Gollop and Monahan (1991) also infer product-specific input requirements from the input requirements from the input structure of pure plants.

2 The Problem of Negatives

In the 'use' table U of the U.S. IO accounts, entry _______ is the amount of commodity i used as an intermediate input in industry j. In the 'make' table v, entry ____ is the amount of commodity k produced by industry j. When the industry is the unit of observation, the commodity technology assumption is specified as the constraint that the use of a particular intermediate i in the production of a given commodity k is fixed proportion ____ of the make of that commodity k, no matter the extent of other production in the industry:

INSERT FROM PAGE 3

This aggregate commodity technology model implies __________. The latter formula indicates that to infer direct requirements coefficients _____ for a given commodity ____ from industry-level use and makes tables, one computes a weighted average _____ of the use of the intermediate in question ___ by all industries ___, with weights _____ from the inverse of the transposed make table that depend on the extent to which the commodity ___ is made in each industry. Generally, many of the weights will be negative, reflecting the need to purge the total input of the intermediate in given industry of the requirements for that input in producing the secondary commodities of that industry. IO researchers have found in producing the secondary commodities of that industry. IO researchers have found that in practice, some of the implied direct requirements coefficients are negative; for some intermediates, more use of an intermediate is purged than actually consumed.

For example, calculations from the benchmark U.S. input-output accounts for 1982 show that 363 commodity technology direct requirements coefficients for manufactured goods are negative and large in absolute value (line 1, column 3 of table 1). Only about 5000 of the elements of this matrix are large in absolute value (to be exact, there are 5131 large elements, 4668 positive and 363 negative). Thus, more than 3 out of 50, or 6 percent, of the large elements of the direct requirements matrix are negative.

3 Possible Explanations of Negatives

I consider two explanations for the negatives that appear to signal the failure of the commodity technology model: **measurement error** and **heterogeneity among pure plants.** The **measurement error** hypothesis is the possibility that the model (1) is literally true at the plant level, but the entries in the aggregative make and use tables are imprecise and possibly biased. Some researchers have argued that the problem of negatives often is derived from **heterogeneity among pure plants** in materials use intensities. Because the definition of an industry can be quite broad, there can be significant variation in materials use intensities among the products that are primary, and aggregation of plants that make distinct primary products with very different materials use patterns can lead to the apparent anomaly of negative requirements. In this section of paper, I consider these two possible explanations for the failure of the commodity technology model on aggregate data.

3.1 Measurement Error and the Census Content of Use Tables

Ten Raa (1988) and ten Raa and van der Ploeg (1989) emphasized measurement error in their statistical approach to re-estimating the published aggregate input-output tables for the U.K.. They assumed that the observed industry-level data _____ and _____ included measurement errors ____ and ____, so that the relation between true values and observed values is:

(2) INSERT FROM PAGE 4 HERE

(3) INSERT TOP OF PAGE 5 HERE

In principal, there also can be more disaggregative use and make tables with entries ____ and ____ that record corresponding statistics for each plant ____ within an industry. I offer an interpretation of equation (2) that makes the relation of the industry-level and plant-level statistics explicit.

Specifically, I assume that the plant-level statistics on the use of materials also might be measured with error:

(4) INSERT SECOND PART OF PAGE 5 HERE.

but regard the degree of measurement error in the make statistic as negligible for the purposes of this study (i.e., ___). This emphasis on uncertainty about use table entries is based on the nature of the underlying source data from the U.S. Census of Manufactures[4]. Relatively complete data on the primary/secondary product split is available, but almost all (98 percent) IO use table entries for manufacturing industries have a low Census content in the sense that they are not based on reports of specified materials use (last column of line 2, table 1). In order to keep reporting burdens down, the Census questions on specified materials use are narrow in the sense that they cover few materials.

It is not clear whether this reporting constraint harms the quality of the input-output use table in a significant way. In

---

[4] The plant-level data on specified materials use us drawn form the Longitudinal Research Database, Which is described in McGuckin and Pascoe (1988). To convert the data from the Census SIC classification system to the input-output classifications used by BEA, I have applied the BEA concordance for 1982, which was kindly provided by Mark Planting.

terms of the percentage of the dollar value of materials use that is specified, this reporting constraint is not very important because in any given industry a few particular materials comprise the bulk of total materials use. The effect of this narrowness of actual use is shown in table 1 (line 1, column 2), which notes that 97 percent of direct requirement coefficients are estimated to be near zero. Most of the near-zero entries do not have direct Census content (line 2, column 2). Although there are a large number of non-zero entries that are not based on Census reports of specified materials use (line 2, columns 1 and 3), these entries still can incorporate Census data indirectly; for example, total materials use is reported in the Census, and the sum of these other non-zero use table entries is known to be the difference between total materials use and the sum of specified materials use.

Another constraint on the Census content of the use tables is that not all plants report specified materials use (table 2). For example, in the 1982 Census, 72 percent of U.S. manufacturing plants were nonreporters of specified materials. However, the nonreporters tend to be the smallest plants, and this 72 percent of plants only accounted for 15 percent of the dollar value of total materials use by the manufacturing sector. Most of the nonreporters are not required to respond to questionnaires on specified materials because they are so small that the Census Bureau gathers information on their activities from the administrative records of other federal government agencies, most often tax records. Another numerous group of establishments (34 percent) fails to comply with the Census requests to specify materials use, but the damaging effects of this noncompliance on the quality of the use tables is held down by the fact that noncomplying establishments also tend to be small. Taken together, the lack of coverage of some materials and nonreporting by some plants results in a loss of information on about one-third of the dollar value of materials consumed by manufacturers; 67 percent of materials use is specified by kind and explicitly reported (line 7).

Despite the moderate dollar values involved in coverage problems, it is possible that the pattern of coverage of specified materials use in the Census of Manufactures biases the use of table toward finding a problem of negatives under the commodity technology model. The lower rows of table 1 provide informal evidence for this latter idea that commodity technology negative requirements attend to be associated with a lack of coverage of specified materials. Remember that in the calculation of the commodity technology coefficients, _____, the weights in the averaging of use table values, ____ are from the inverse of the transposed make table. Generally, the where make of the item is primary. Among the 363 large negative direct requirements coefficients, 328 are for materials that are not specified by kind in the Census reports for the primary industry (line 2, column 3 of table 1). In other words, in 90 percent of the cases of negative direct requirements, specified materials use is not available in the Census reports. This percentage is much higher than the roughly two-thirds of the large positive coefficients for which materials use is not specified by kind in the primary industry.

Because there is some indirect Census content in use table cells not covered by reports on specified materials use, the percentages given in line 2 of table 1 understate the Census content of the use tables. Using a more appropriate, broader definition of Census content, I have computed bounds on the effects of nonreporting on the accuracy of the aggregate use table. To explain the bounds, further notation is needed. Let ____ be an indicator variable that is one if plant __ is a nonreporter of specified use of material __ and is zero otherwise. For the purposes of calculating the bound, I assume that if specified materials use is reported at all, it is reported at all, it is reported exactly[5]; i.e., if _________. Estimate of total materials use are available for all plants, so an upper bound on the unknown value of a nonreporting plant's actual use ____ is the difference between the plant's total

---

[5] This assumption ignores a censoring constraint that is discussed further in the following section. The censoring constraint likely biases the bound on ____ downward slightly, leading to an understatement of the importance of measurement error in explaining the problem of negative.

materials use ____ and the sum of reported specified use of other materials:

(5) INSERT FROM PAGE 7

For a particular industry ___, the upper bound on the unknown use table cell is the sum across the upper bounds on the planet-level entries:

(6) INSERT FROM PAGE 7

The lower bound on actual use of a nonreported material at the plant level is always zero, so at the industry level that lower bound on an unknown use table cell is the sum of reported plant-level use of the specified material:

(7) INSERT FROM PAGE 7

Let us focus on the issue of whether inaccuracy of use table cells has created negatives in the direct requirements matrix of the commodity technology model. The scalar expression for the ___ element of the direct requirements matrix ____, is

(8) INSERT FROM PAGE 7

where the weights in the averaging of use table values, ____, are from the inverse of he transposed make table. Let ____ be an indicator variable that is one if the weight ___ is positive and zero if the weight negative. Then for given weight ___ is a positive and zero if the weight is negative. Then, for given weights ____, the Census records on specified materials use place the following upper bound on the ____ element of the direct requirements matrix:

(9) INSERT FROM PAGE 8

The sign of this upper bound (9) for direct requirements entries that are negative and large in absolute value is summarized in the last row of table 1. Out of the 363 large negative entries, the Census upper bound permits reversal in 140 cases. This finding suggests that measurement error in the aggregate use tables is an important part of the problem of negatives; the amount of unknown use of materials is large enough for 39 percent of the large negative direct requirements entries to possibly be due to measurement error alone. However, measurement error likely is not the only source of the problems of negatives. In 71 percent of the cases of large negatives, the unknown amounts in cells of the cases table are not large enough to eliminate the large negatives in the direct requirements matrix.

3.2 Heterogeneity of Use among Pure Plants

Any large degree of heterogeneity in use intensities among pure plants signals a failure of the commodity technology assumption, whether or not such heterogeneity leads to negative values in commodity technology coefficients derived from aggregate make and use tables[6]. Moreover, it is well-known in the IO literature that such heterogeneity does contribute to the problem of negatives. For example, Rainer and Richter (1992) discuss how in the Austrian IO system a failure to distinguish between electric utilities that produce power and those that distribute power can lead to the problem of negatives. The problem is that electricity distribution facilities have a very high own input, the purchase of electricity from other establishments, which boosts the average own input for the electric utility sector as a whole. Electricity generating plants do not purchase much electricity. When electricity is made as a secondary product in other industries, the compositions of intermediate inputs for electricity generating plants than that of the distribution facilities. Thus, the commodity technology solution on aggregate data seems to imply that there

---

[6] This statement is conditional on the implemented classification system. It is conceivable that there are alternative classification systems which re-define primary products in such a way that the commodity technology assumption is valid. Triplett (1992) makes a similar point; he argues that the current SIC systems often groups establishments with quite different production functions into a single SIC category and notes that this might be the causer of failures in attempts to estimates establishment-level production functions.

is a negative electricity requirement for the primary products of industries with large secondary production of electricity.

Let us try to discover the extent to which such heterogeneity of use intensities among plants producing only primary products is a problem in the U.S. manufacturing sector, looking beyond the electricity example. To investigate this we need technology coefficient estimates that are not contaminated by the measurement error from nonreporting discussed above. I derive technology coefficient estimates from the records for the plants that reported specified materials use, focusing on the 62,757 reporting pure plants (table 2). Notationally, a P superscript denotes requirements ____ from a plant ___ in this sub-sample of pure plants. The commodity technology model, equation (1), represents the production process of multi-product plants as a liner combination of single-product technologies,

(10) INSERT FROM PAGE 9

and one can take the intensity of use of a material ___ at a pure plant __ in industry ___ as representative of the materials requirement for the corresponding commodity, whenever produced, ____. Almost all industries have some pure plants, so I am able to calculate pure-plant materials use intensities for 3754 of the 3904 materials-industry combinations where use is specified by kind (last column of line 3, table 1).

Table 3 presents four statistics describing the distributions of pure-plant materials use intensities, ___. First consider the weighted average of the actual use intensities of all reporting plants in the industry, where the total product output of each plant is used as the weight. These averages, ___ are equivalent to the ratio of the total use of a given material by all pure plants in an industry to the total output of these pure plants, so the averages are analogous to the calculated commodity technology model coefficients, ___ from ____, that would be derived from the standard aggregate use and make table entries, ____ , if there were no secondary production of the commodity and only pure plants in the industry in question. In more than half of the cases, average consumption of the dated material among pure plants is less than one percent of the 3754 cases, the average requirement for the material is more than 14.21 percent of output.

The distribution (across material-industry combinations) of the median (across plants) materials use intensities, ___, is very shrewd towards zero. This happens because for any given specified material, the Census records record that many plants in the reporting industry do not use any of the material at all. This tendency for there to be a mode at zero in specified materials use is s pronounced that more than 75 percent of the medians are zero.

Table 3 also presents two dispersion measures. The first is computed by scaling the standard deviation of the pure-plant materials use intensities ___ by 1.34, which is the interquartile range of a standard normal distribution. The second is the empirical interquartile range, the distance between the 25th and 75th quantiles of ___. Many of he interquartile ranges are zero, reflecting the fact that in more than half of the material-industry reporting combinations, more than 75 percent of the plants record zero use of the specified material. Inspection of the empirical distributions of the raw data on specified materials use revealed that may were bi-modal, with a first peak at zero and a second positive peak that sometimes was quite far from zero.

Next, let us investigate the extent to which the bi-modal distributions of reported materials use actually reflect heterogeneity of use among pure plants. This is a non-trivial effort because there is an additional reason why recorded use of specified materials might be zero: plants are told to omit listing of specified materials use if the amount used falls below a given censoring threshold, usually 10,000 dollars. Thus, the statistics in table 3 need to be interpreted with caution because they fail to distinguish between plants that actually use none of the specified material and plants that record use of zero because the amount is rounded down to zero by censoring.

To provide a formal analysis of the effects of censoring, first consider the possible that censoring is the only source of reporting of zero use of a specified material and that pure plants are essential homogeneous in their materials use. I parameterize this null hypothesis by interpreting the commodity technology assumption (10), which is written as is a single coefficient ___ applies to all plants, a s statement about the central tendency of a (unimodal) distribution of requirements. Specifically, I assume that the logic of the actual commodity technology coefficient that applies to a particular plant ___, log (______), is drawn from a normal distribution ____ with mean log (____) and variance____.

(11)

A pure plant's actual use of the material is

(12) _____

The censoring constraint is that the recorded use of the material ___ equals zero if the actual use ____ is less than 10,000 dollars:

(13) INSERT FROM PAGE 11 HERE

Equivalently, the observation is censored of the logic of the pure plant's actual use intensity, log ____, is below a known threshold[7]:

(14) INSERT FROM PAGE 11 HERE

The system of equations (11)-(14) constitute a standard tobit model. To estimate the unknown parameters ___ and ___ under this null hypothesis of homogeneity, I use the maximum-likelihood method described by Amemiya (1973). Under this null hypothesis, the maximum likelihood estimates of the tobit model provide consistent and asymptotically efficient estimates of the unknown parameters ___ and ____.

To formalize the alternative idea that heterogeneity of materials use among pure plants creates a problem of negatives, supposes that among the primary products of each industry there are two types of commodities, which I call *low* and *high* to denote their relative intensity of use of a specified material. The commodity technology assumption as given in (11) and (12) is assumed to be appropriate for the commodities taken individually, but the mean materials requirements for these two commodities ___

---

[7] Any plant with total output less than 10,000 dollars is assumed to have all detailed materials use censored, so the censoring threshold in (14) is well-defined

and _____ are assumed to be quite different[8]. Specifically, I assume that the mean and dispersion ___ parameters for the low materials use commodity are so small that it is unlikely for the censoring threshold (13) to be exceeded in samples of the size used here.

To help us relate the parameters of the null and alternative hypotheses, characterized the location of the mixture distribution by its median $2._5$, which is the value of "^{C} which exactly half of the plants are expected to fall below. Under the null hypothesis-a single type of plants with materials use requirements characterized by a censored normal distribution. The interquartile range ____ describes the dispersion of the mixture distribution. Under the null hypothesis of a non-mixed normal distribution, the scaled standard deviation 1.34____ equals the interquartile range. To estimate ____ and ____ under the alternative with heterogeneity, I employ Powell's (1984, 1986) Censored Least Absolute Deviations (CLAD) and Censored Regression Quantile (CRQ) estimators.

Table 4 presents a summary of the results. The estimated central location of most direct requirements distributions is quite small with the tobit correction for censoring. The tobit results imply that for 95 percent of the material-industry

---

[8] To estimate this alternative, I do not actually rely on the normality or symmetry assumptions of (11). The test statistic has the appropriate size and power to find heterogeneity under a wide class of bimodal distributions.

combinations, the implied mean of the (uncensored) distribution is less than 2.96 percent of output. The tobit estimated of the mean (table 4, column 1) tend to be much smaller than the uncorrected means in which weights are proportional to the output size of the plant (table 3, column 1). At first glance, this finding might seen anomalous because the rounding of small positive numbers down to zero biased an uncorrected mean downward, not upward. But, in this case removing the effects of censoring has resulted in a further lowering of the estimated of the mean.

Inspection of the distribution for particular material-industry combinations revealed the nature of the apparent anomaly in the tobit estimates. At a given materials use intensity, the likelihood that the censoring constraint will be binding decreases with the output size of the plant. But in many cases, some very large plants report zero use of a specified materials. The tobit models tend to locate the uncensoring distribution close to zero because censoring is such an unlikely explanation for the large plants' failure to report use if the specified material.

Chart 1 illustrates this effect with hypothetical data that resembles the distributions of specified materials use in industries with significant heterogeneity among plants[9]. The upper panel shows that in this case the average (across the 100 plants) materials consumption as a share of output is a bit less than 50 percent, but use of this specific materials would be recorded at zero for the slightly more than one-fifth of the plants represented by slashed bars. Tobit estimates are formed by analyzing the distribution shown in the middle panel; for plants not reporting zero materials use, the variable of interest is the observed logit of materials use, this variable is replaced with the logit of the censoring threshold, which depends on the level of the plant's output, as in equation (14). If the censoring threshold were a constant fraction of a plant's output, say 20 percent, then this mass of the logit of censoring thresholds would appear in a single bar in the middle panel, the bar spanning the -1 to -2 range. But with a variable censoring threshold, any large plants reporting zero materials use show up as even lower negative observations. For example, the largest plant reporting zero materials use-shown toward the right of the lower panel-made about 1.4 million dollars of goods, so this plant's actual material consumption would be censored to zero only if the actual share was well below 1 percent of output; this is a logit in the -5 to 04 range. the tobit estimates of the

---

[9] This chart does not use actual data in order to safeguard the confidentiality of individual plant records. The hypothetical data was created by Monte Carlo and is representative in the relevant respects discussed in the text.

mean (of the logit distribution ar pulled well below zero by the relatively large plants that report no use of the specified material.

The alternative estimates of location, shown in the table 4 column labelled CLAD median, are robust to this kind of heterogeneity. If the uncensored distribution actually has a mode near zero and a second mode that is quite a bit larger, the CLAD median will tend to find the actual mid-point of the uncensored, mixed distribution. As shown in the table, the CLAD medians generally remain well above the estimates from the tobit model, a finding that is consistent with bi-modality in the uncensored distribution, the tobit estimates of location appear to be biased down by heterogeneity of materials use among pure plants. Also, the robust estimates of dispersion tend to be quite a bit larger than the tobit estimates of dispersion, likely reflecting the ability of the robust method to capture the wider spread of the mixture of distributions.

Of course, the tobit and robust estimates of location and dispersion are imprecise, so it is important to check whether the deviations of the parameter estimates are beyond the range of estimation error. The final column of table 4 presents the results of such a formal test of whether the differences between the tobit and CLAD-CRQ estimates are statistically significant, using the specification testing technique of Hausman (1978) to infer the asymptotic distribution of the difference between the estimators[10]. In 25 percent of the material-industry combinations where such a test was calculable, the null of homogeneity can be rejected very strongly-at the .55 percent marginal significance level. At the more conventional significance level of 5 percent, this test shows evidence of heterogeneity among plants for about one-third of the material-industry combinations.

Table 5 repeats the analysis of table 4 for a subsample of material-industry combinations in which the specification test for the heterogeneity is calculable, the null of no heterogeneity tends to increase with the material cost share. For these larger material requirements, the specification that rejects the null of no heterogeneity in 70 percent of the cases; tobit means tend to be below CLAD medians, and the robust estimates of dispersion are wider than the tobit estimates of dispersion.

Specific examples are helpful for culling possible explanations for the heterogeneity of materials use among pure plants. Towards this end, table 6 presents the tobit and robust location estimates for the largest fifteen material requirements, where the CLAD median estimates range from 50 percent to 89 percent. From the inspection of specific results, it was apparent that the specification test often rejected the absence

---

[10] Newey (1987) suggests applying the Hausman test in a context like this where tobit maximum likelihood estimates are being compared with CLAD estimates. Estimates of the density at the quantiles are needed to calculate a variance-covariance matrix for the CRQ/CLAD estimates, and I used an Epanechnikov kernel for such density estimates.

of heterogeneity because the robust estimates of dispersion were wider than the tobit estimates of dispersion. Focussing more narrowly on the locations of the distributions, table 6 show that (at the 5 percent level) a test that the tobit and robust estimates of location are equal still rejects in 60 percent of the cases (9 out of 15). However, there is no significant evidence of heterogeneity in plants' use of rice in milling, barely in making malt, cottonseed in making cottonseed oil, nonferrous metals in nonferrous rolling, yarns in fabricating tire cord, or fishery products in canning fish. for those cases where evidence of heterogeneity is significant, a number of possible explanations were culled from the plant-specific product and material records of these (usually large) plants that contributed the most to the lowering of the tobit means. For example, the petroleum refining industry includes some pure plants that make specialty products from own-industry inputs, and such plants report little or no use of crude oil and natural gas. Some special products sawmills do not use the dominant material input-logs-but rather rely on other wood products as raw material itself is defined rather narrowly, so the use of close substitutes-corn, barely, oats or sorghum grain-appears as a form of heterogeneity. Inspection of additional specific examples both revealed repetition of these basic patterns and added a bit to the list of possible explanations for heterogeneity.

4 Conclusion

This paper has done something different than other attempts to investigate the failure of the commodity technology model on aggregate data, such as those by ten Raa (1984), ten Raa and van der Ploeg (1989), and Ranier and Richter (1992). Measurement error in aggregate use tables was quantified not by asking compilers about the precision of the estimates, but rather by going directly to census records from individual manufacturing plants and tabulating the effects of nonreporting. Heterogeneity of materials use among pure plants was investigated by actually looking at the empirical distributions across pure plants of materials use and by summarizing the evidence of heterogeneity in test statistics with known properties.

Ten Raa (1984) and ten Raa and van der Ploeg (1989) concluded that the degrees of measurement error in aggregate make and use tables cannot fully account for the problem of negatives. Although I have not overturned this conclusion by taking an extensive look at the plant-level records, what I find striking is not that some negatives cannot be explained away on these grounds, but rather that many negatives are likely due to measurement error.

With regard to those material-industry combinations where measurement error does not appear to be a serious problem, there is broad-based evidence of heterogeneity of materials use among pure plants. Further work is needed to understand this heterogeneity and to develop ways of proceeding with IO analysis in the presence of such heterogeneity. Most such research can proceed along one of the following three lines.

First, it is possible that all variant of the Leontif technology assumption are poor approximations to the true relation between the intensity of factors of production and the level of attained output. Neoclassical economic theory certainly suggests that substitutability of capital, labor, and other materials for the specified materials could induce heterogeneity in plant-level use, and no-one has demonstrated whether or not such variations in the factor input mixes can explain the heterogeneity among pure plants that I have documented here. Data on prices and capital and labor input for individual plants also is available in the LRD, so further research in this direction is feasible[11].

Second, it possible that much of the apparent heterogeneity among plants within a given industry is due to inadequate industry classification. The Standard Industrial Classification (SIC) system of the United States does not always give similarity of input structures primacy, partly because the system has never

---

[11] Whether or not the heterogeneity described here is to be taken as evidence against the Leontief technology or the neoclassical Cobb-Douglas technology depends on whether one assumed that all pure plants in an industry face the same prices for the input and materials. By working in terms of current dollar cost shares and implicitly assuming that all plant face the same prices, I have been able to describe the results as a test of the Leontief technology. Alternatively, under the assumption of heterogeneous prices, this paper presents evidence that the share elasticity of substitution is not unity; i.e, the technology is not Cobb-Douglas.

had a coherent, model-based rationale. Triplett (1992) and others are beginning to lay the conceptual foundation for model-based economic classification. I suspect that much of the apparent heterogeneity of materials use could be eliminated if similarity of material input structure were a more universally-applied criterion in the design of the industrial classification system. In some empirical work along these lines, Abbott and Andrews (1990) were able to devise to devise some interesting alternative classifications.

Last, I think it is important to investigate a related problem, whether the heterogeneity among pure plants reflects what Rainer and Richter (1992) call inhomogeneity due to vertical integration. As an example, Rainer and Richter point out that among plants in the iron and steel industry group, many will engage in both the production of finished steel products and the smelting of iron ore. If the latter commodity (smelted ore) is produced and consumed within the same establishment, the vertically integrated plant still appears to be pure in the vertically-integrated plant is going to be very different from that in non-integrated plants. It is clear that vertical integration within plants exists. Further research is needed to document the extent of vertical integration within plants and to explain why some plants choose to integrate and others do not. In his recent Nobel lecture, R.H. Coase (1992) singled out the work at the Center for Economic Studies (CES) as especially important to further understanding the activities of firms.

This paper uses the data available at the CES to provide clear evidence of heterogeneity among plants that appear to similar by conventional classification measures. It remains to be seen what theories can explain this heterogeneity or whether alternative classification systems can remove the appearance of heterogeneity.

TABLE 1

TABLE 2

TABLE 3

TABLE 4

TABLE 5

TABLE 6

CHART 1

References

Abbott, Thomas A. and Stephen H. Andrews (1990). "The classification of manufacturing industries: an input-based clustering of activity," Proceedings, Census Bureau Annual Conference.

Amemiya, T. (1973). "Regression analysis when the dependent variable is truncated normal," Econometrica, 42, 999-1012.

Bureau of Economic Analysis (1991). "Benchmark input-output accounts for the U.S. Economy, 1982" Survey of Current Business, July.

Coase, R.H. (1992). "The Institutional Structure of Production" American Economic Review, September, 713-719.

Gollop, Frank M. and James L. Monahan (1991). "A Generalized Index of Diversification: Trends in U.S. Manufacturing," Review of Economic and Statistics, Volume LXXIII, No. 2, May, 318-30.

Hausman, J. (1978). "Specification tests in econometrics," Econometrica, 46, 1251-72.

McGuckin, Robert H. and SuZanne Peck (1992). "Manufacturing Establishments Reclassified into New into New Industries: the Effect of Survey Design Rules," Center for Economic Studies Discussion Paper 92-14, Bureau of the Census.

Newey, W.K. (1987). "Specification tests for distributional assumptions in the tobit model,' Journal of Econometrics, 34, 125-145.

Powell, J.L. (1984) "Least absolute deviations estimation for the censored regression model," Journal of Econometrics, 25, 303-325.

Powell, J.L. (1986) "Censored Regression Quantiles." Journal of Econometrics, 32, 143-55.

Rainer, Norbert and Josef Richter (1992). "Some aspects of the analytical use of descriptive make and absorption tables,' Economic Systems Research, 4(2), 159-72.

Streitwieser, Mary L. (1991). "The extent and nature of establishment level diversification in sixteen U.S. manufacturing industries," Journal of Law and Economics, Fall, 496-551.

ten Raa, T. (1988). "An alternative treatment of secondary products in input-output analysis: frustration," Review of Economics and Statistics, 70(3), 535-38.

Triplett, Jack E. (1992). "Perspectives on the SIC: Conceptual Issues in Economic Classification," manuscript, Bureau of Economic Analysis, U.S. Department of Commerce.

CENTER FOR ECONOMIC STUDIES DISCUSSION PAPERS

88-1 "Long-Run Expectations and Capacity," by Robert H. McGuckin & Peter A. Zadrozny, 4/88. Forthcoming in Proceedings, Conference on Capacity Utilization, (1988). (29 pages, $7.25)

88-2 "The Longitudinal Research Database (LRD): Status and Research Possibilities," by Robert H. McGuckin & George A. Pascoe, Jr., 7/88. Survey of Current Business, Volume 68, Number 11, pp. 30-37 (1988). (29 pages, $7.25)

88-3 "Public Use Microdata: Disclosure and Usefulness," by Robert H. McGuckin & Sang V. Nguyen, 9/88. Journal of Economic and Social Measurement, Volume 16, Number 1, pp. 19-40, (1990). (29 pages, $7.25)

88-4 "Modelling Technical Progress and Total Factor Productivity: A Plant-Level Example," by Edward C. Kokkelenberg & Sang V. Nguyen, 10/88. Journal of Productivity Analysis, Volume 1, Number 1, pp. 69-90, (1989). (37 pages, $9.25)

88-5 "Analytic Derivatives For Estimation of Linear Dynamic Models," by Peter A. Zadrozny, 11/88. Journal of Computers and Mathematics with Applications, Volume 18, pp. 539-553, (1989). (41 pages, $10.25)

89-1 "Measuring the Trade Balance in Advanced Technology Products," by Thomas A. Abbott, III, Robert H. McGuckin, Paul Herrick, and Leroy Norfolk, 1/89. Proceedings, 1989 American Statistical Association. (33 pages, $8.25)

89-2 "Using Linked Census R&D-LRD Data to Analyze the Effect of R&D Investment on Total Factor Productivity Growth," by Frank R. Lichtenberg and Donald Siegel, 1/89. Economic Inquiry under title "The Impact of R&D Investment on Productivity - New Evidence Using Linked R&D-LRD Data," Volume 29, pp. 2-13, (1990). (46 pages, $11.50)

89-3 "The Effect of Takeovers on The Employment and Wages of Central Office and Other Personnel," by Frank R. Lichtenberg and Donald Siegel, 6/89. Journal of Law and Economics, Volume 33, pp. 383-408, (1990). (36 pages, $9.00)

89-4 "Entrepreneur Factor Inputs and Small Business Longevity," by Timothy Bates, 6/89. Review of Economics and Statistics under the title "Entrepreneur Human Capital Inputs and Small Business Longevity," Volume LXII, Number IV, (1990). (40 pages, $10.00)

89-5 "The Effects of Leveraged Buyouts on Productivity and Related Aspects of Firm Behavior," by Frank R. Lichtenberg and Donald Siegel, 7/89. Presented at the Conference on Innovation in Corporate Organization, Control, and Incentives, 1990. #3022, NBER Working Paper Series. Journal of Finance and Economics, Volume 27, pp. 165-194, (1990). (53 pages, $13.25)

89-6 "Firm Entry and Post-Entry Performance in the U.S. Chemical Industries," by Timothy Dunne, Mark J. Roberts, and Larry Samuelson, 7/89. Journal of Law and Economics, (1989). (49 pages, $12.25)

89-7 "Price Dispersion in U.S. Manufacturing," by Thomas A. Abbott, III, 10/89. (16 pages, $4.00)

89-8 "Multifactor Productivity and Sources of Growth in Chinese Industry: 1980-1985," by Robert H. McGuckin, Sang V. Nguyen, Jeffrey R. Taylor, and Charles A. Waite, 10/89. Presented at the winter meetings of the American Statistical Association, 1/90. An Industrial Census of China, Xu Gang, ed., Beijing, People's Republic of China: The Secretariat of the International Seminar of Analysis on China's 1985 Industrial Census Results, pp. 113-138, (1990). (35 pages, $8.75)

89-9 "The Characteristics of Business Owners (CBO) Database," by Alfred Nucci, 10/89. (28 pages, $7.00)

90-1 "Longitudinal Economic Data at the Census Bureau: A New Database Yields Fresh Insights on Some Old Issues," by Robert H. McGuckin, 1/90. Proceedings of the Statistics Canada Symposium on Analysis of Data in Time, pp. 243-253, (1990). (33 pages, $8.25)

90-2 "An Analysis of Small Business Size and Rate of Discontinuance," by Timothy Bates and Alfred Nucci, 1/90. Journal of Small Business Management, pp. 1-7, (1990). (14 pages, $3.50)

90-3 "Export Performance and State Industrial Growth," by Rodney A. Erickson, 1/90. Economic Geography, Volume 65, Number 4, pp. 280-292, (October 1989). (31 pages, $7.75)

90-4 "Gross Job Creation, Gross Job Destruction, and Employment Reallocation," by Steven J. Davis and John Haltiwanger, 2/90. Forthcoming, Quarterly Journal of Economics. (52 pages, $13.00)

90-5 "Estimating a Multivariate Arma Model with Mixed-Frequency Data: An Application to Forecasting U.S. GNP at Monthly Intervals," by Peter A. Zadrozny, 6/90. (58 pages, $14.50)

90-6 "Wages and the Risk of Plant Closings," by Timothy Dunne and Mark J. Roberts, 7/90. American Statistical Association Proceedings of the Business and Economics Statistics Section, pp. 275-280, (1989). (31 pages, $7.75)

90-7 "The Classification of Manufacturing Industries: An Input-Based Clustering of Activity," by Thomas A. Abbott, III and Stephen H. Andrews, 8/90. Proceedings, Census Bureau Annual Research Conference, pp. 26-45, (1990). (47 pages, $11.75)

90-8 "The Extent and Nature of Establishment Level Diversification in Sixteen U.S. Manufacturing Industries," by Mary L. Streitwieser, 8/90. Journal of Law and Economics, Part 2, pp. 503-534 (1991). (39 pages, $9.75)

90-9 "Self-Employment Trends Among Mexican Americans," by Timothy Bates, 8/90. Quarterly Review of Economics and Business, (1990). (26 pages, $6.50)

90-10 "Gross Job Creation and Destruction: Microeconomic Evidence and Macroeconomic Implications," by Steven J. Davis and John Haltiwanger, 9/90. NBER Macroeconomics Annual, MIT Press: Cambridge, MA, (1990). (51 pages, $12.75)

90-11 "Returns to Scale in Small and Large U.S. Manufacturing Establishments," by Sang V. Nguyen and Arnold P. Reznek, 9/90. Small Business Economics, Number 3, pp. 197-214, (1991). (35 pages, $8.75)

90-12 "The Relationships Among Acquiring and Acquired Firms' Product Lines," by Robert H. McGuckin, Sang V. Nguyen and Stephen H. Andrews, 9/90. Journal of Law and Economics, Part 2, pp. 477-502 (October 1991). (45 pages, $11.25)

90-13 "The Determinants of U.S. Intra-Industry Trade," by F.M. Scherer and Keun Huh, 12/90. International High-Technology Competition, Cambridge, Massachusetts: Harvard University Press (1992). (19 pages, $4.75)

91-1 "Published Versus Sample Statistics from the ASM: Implications for the LRD," by Steven J. Davis, John Haltiwanger and Scott Schuh, 1/91. Proceedings of the American Statistical Association Business and Economics Section, (1990). (38 pages, $9.50)

91-2 "R&D Reactions to High-Technology Import Competition," by F.M. Scherer and Keun Huh, 3/91. Forthcoming in Review of Economics and Statistics. (37 pages, $9.25)

91-3 "Measuring Total Factor Productivity, Technical Change and the Rate of Returns to Research and Development," by Sang V. Nguyen and Edward C. Kokkelenberg, 5/91. Journal of Productivity Analysis, Number 2, pp. 269-282, (1992). (25 pages, $6.25)

91-4 "Decomposing Technical Change," by Michael Gort, Byong-Hyong Bahk, and Richard A. Wall, 5/91. Southern Economic Journal, Volume 60, Number 1, pp. 220-235. (40 pages, $10.00)

91-5 "The Structure of Production Technology, Productivity and Aggregation Effects," by Phoebus J. Dhrymes, 8/91. Proceedings, 1990 Annual Research Conference as "The Structure of Production Technology: Evidence from the LED Sample 1." (62 pages, $15.50)

91-6 "Technical Inefficiency and Productive Decline in the U.S. Interstate Natural Gas Pipeline Industry Under the Natural Gas Policy Act," by Robin C. Sickles and Mary L. Streitwieser, 10/91. Journal of Productivity Analysis, Volume 3, pp. 119-133, June 1992. (27 pages, $6.75)

91-7 "Technology Usage in U.S. Manufacturing Industries: New Evidence from the Survey of Manufacturing Technology," by Timothy Dunne, 11/91. RAND Journal of Economics. (31 pages, $7.75)

91-8 "Multiple Classification Systems for Economic Data: Can a Thousand Flowers Bloom? And Should They?" by Robert H. McGuckin, 12/91. Proceedings of the 1991 Conference on the Classification of Economic Activity, 1991. (36 pages, $9.00)

91-9 "Commercial Bank Lending Practices and the Development of Black-Owned Construction Companies," by Caren Grown and Timothy Bates, 12/91. Journal of Urban Affairs, pp. 25-41, (1992). (31 pages, $7.75)

91-10 "The Influence of Location on Productivity: Manufacturing Technology in Rural and Urban Areas," by Sheila A. Martin, Richard McHugh, and Stanley R. Johnson, 12/91. (33 pages, $8.25)

92-1 "Productivity Dynamics: U.S. Manufacturing Plants, 1972-1986," by Eric J. Bartelsman and Phoebus J. Dhrymes, 2/92. (51 pages, $12.75)

92-2 "The Dynamics of Productivity in the Telecommunications Equipment Industry," by G. Steven Olley and Ariel Pakes, 2/92. (56 pages, $14.00)

92-3 "Price Dispersion in U.S. Manufacturing: Implications for the Aggregation of Products and Firms," by Thomas A. Abbott, III, 3/92. (36 pages, $9.00)

92-4 "Estimating Capital Efficiency Schedules Within Production Functions," by Mark E. Doms, 5/92. Forthcoming in _Economic Inquiry_. (26 pages, $6.50)

92-5 "Costs, Demand, and Imperfect Competition as Determinants of Plant-Level Output Prices," by Timothy Dunne and Mark J. Roberts, 6/92. _Empirical Studies in Industrial Organization_, pp. 13-24, (1992). B.B. Audretsch and J.J. Siegfried, eds., Clover Publishers. (29 pages, $7.25)

92-6 "On the Turnover of Business Firms and Business Managers," by Thomas J. Holmes and James A. Schmitz, Jr., 7/92. (49 pages, $12.25)

92-7 "The Characteristics of Business Owners Database," by Alfred Nucci, 8/92. (82 pages, $20.50)

92-8 "Analytic Use of Economic Microdata: A Model for Researcher Access with Confidentiality Protection," by Robert H. McGuckin, 8/92. _Proceedings, International Seminar on Statistical Confidentiality_, 1992. (23 pages, $5.75)

92-9 "The Structure of Technology, Substitution, and Productivity in the Interstate Natural Gas Transmission Industry under the NGPA of 1978," by Robin C. Sickles and Mary L. Streitwieser, 8/92. (39 pages, $9.75)

92-10 "The Time-Series Pattern of Firm Growth in Two Industries," by Kenneth R. Troske, 9/92. (37 pages, $9.25)

92-11 "Managerial Tenure, Business Age and Small Business Dynamics," by Thomas J. Holmes and James A. Schmitz, Jr., 9/92. (47 pages, $11.75)

92-12 "The Aggregate Implications of Machine Replacement: Theory and Evidence," by Russell Cooper and John Haltiwanger, 10/92. _The American Economic Review_, Volume 83, Number 3, pp. 360-382, (June 1993). (41 pages, $10.25)

92-13 "Gender Segregation in Small Firms," by William J. Carrington and Kenneth R. Troske, 10/92. Revised, May 1993. Forthcoming in _Journal of Human Resources_. (45 pages, $11.25)

92-14 "Manufacturing Establishments Reclassified Into New Industries: The Effect of Survey Design Rules," by Robert H. McGuckin and SuZanne Peck, 11/92. _Journal of Economic and Social Measurement_, Volume 1, Number 2 (1993). (25 pages, $6.25)

92-15 "Wages, Employer Size-Wage Premia and Employment Structure: Their Relationship to Advanced-Technology Usage at U.S. Manufacturing Establishments," by Timothy Dunne and James A. Schmitz, Jr., 12/92. Forthcoming in _Economica_. (36 pages, $9.00)

92-16 "Decomposing Learning by Doing in New Plants," by Byong-Hyong Bahk and Michael Gort, 12/92. Journal of Political Economy, Volume 101, Number 4, August 1993. (33 pages, $8.25)

93-1 "Academic Science, Industrial R&D, and the Growth of Inputs," by James D. Adams and Leo Sveikauskas, 1/93. (48 pages, $12.00)

93-2 "Science, R&D, and Invention Potential Recharge: U.S. Evidence," by James D. Adams, 1/93. Forthcoming in American Economic Review Papers and Proceedings. (13 pages, $3.25)

93-3 "LBOs, Debt and R&D Intensity," by William F. Long and David J. Ravenscraft, 2/93. The Deal Decade, 1993, Margaret M. Blair, ed., Washington, D.C.: The Brookings Institution. (38 pages, $9.50)

93-4 "Learning By Doing and Competition in the Early Rayon Industry," Ronald S. Jarmin, 2/93. Forthcoming in RAND Journal of Economics. (32 pages, $8.00)

93-5 "Inter Fuel Substitution and Energy Technology Heterogeneity in U.S. Manufacturing," by Mark E. Doms, 3/93. (48 pages, $12.00)

93-6 "Environmental Regulation and Manufacturing Productivity at the Plant Level," by Wayne B. Gray and Ronald J. Shadbegian, 3/93. (29 pages, $7.25)

93-7 "Asymmetric Learning Spillovers," by Ronald S. Jarmin, 4/93. (22 pages, $5.50)

93-8 "Evidence on IO Technology Assumptions from the Longitudinal Research Database," by Joe Mattey, 5/93. (26 pages, $6.50)

CES Discusion Papers are available in a variety of subscription formats: (INTERNET, WordPerfect 5.1 Diskette, and Yearly Series single paper copy). Ordering information is found on the attached CES Discussion Paper Subscription & Request Form. In additon, papers are sold individually for the amount specified after each title.

CES Discussion Paper Subscription and Request Form

☐ **Yes, I'd like to subscribe to the following:**

- ☐ Regular announcements of the **CES Discussion Paper** Series on the INTERNET, free of charge.
- ☐ Yearly subscription to the **CES Discussion Paper** Series in single paper copy for a fee of $125.00 per year.
- ☐ Yearly subscription to the **CES Discussion Paper** Series on computer diskette in WordPerfect 5.1 for a fee of $60.00 per year. Diskette Size ☐ 3½" ☐ 5¼"

The CES Annual Report is available (free of charge).
☐ Please add my name to your mailing list.

☐ **No, I don't want to subscribe at this time but, I would like to order the following CES Discussion Paper(s):**

CES# _________ CES# _________ CES# _________ Total $_________

NAME ______________________________________

ADDRESS _____________________________________

____

____

Phone ( ) ___________________ FAX ( ) ____________________

INTERNET ADDRESS ___________________________
☐ Check here if this is an change of address.

METHOD OF PAYMENT (Payable to: Commerce, Census
☐ Check ☐ Money order ☐ Mastercard ☐ Visa

_____________________________ Expiration Date: ______________________
Card Number Month Year
Amount enclosed: $_______

Please send order form and payment to:
Bureau of the Census
Economic Planning and Coordination
Center for Economic Studies

Room 1587-3
Washington D.C. 20233-6101
Attn: Tracy C. Crosby