Technical Memorandum
Project MC-286
Matthew D. Shapiro
January 11, 2009

Using the Survey of Plant Capacity to Measure Capital Utilization: Technical Memorandum

Yuriy Gorodnichenko
UC Berkeley

Matthew D. Shapiro
University of Michigan

Abstract

Most capital in the United States is idle much of the time. By some measures, the average workweek of capital in U.S. manufacturing is as low as 55 hours per 168 hour week. The level and variability of capital utilization has important implications for understanding both the level of production and its cyclical fluctuations. This report investigates a number of issues relating to aggregation of capital utilization measures from the Survey of Plant Capacity and makes recommendations on expanding and improving the published statistics deriving from the Survey of Plant Capacity. The paper documents a number of facts about properties of capital utilization. First, after growing for decades, capital utilization started to fall in mid 1990s. Second, capital utilization is a useful predictor of changes in capacity utilization and other factors of production. Third, adjustment of productivity measures for variable capital utilization improves statistical and economic properties of these measures. Fourth, the paper constructs weights to aggregate firm level capital utilization rates to industry and economy level, which is the major enhancement to available data.

Using the Survey of Plant Capacity to Measure Capital Utilization

Yuriy Gorodnichenko and Matthew D. Shapiro

Executive Summary

This documentation discusses the benefits to the Census Bureau based on the results of the project 286 conducted by Matthew D. Shapiro. The project used Title 13, Chapter 5 data taken from Annual Survey of Manufactures (ASM), Census of Manufacturers (CM), and Survey of Plant Capacity (SPC). The memorandum also presents details involved in constructing consistent series of capital utilization at different level of aggregation. Detailed analysis of the constructed series is in the accompanying working paper.

I. Benefits

In this section we briefly discuss how the project fulfilled the goals outlined in the research proposal.

This project should improve the Census Bureau's understanding of the quality of the Survey of Plant Capacity (SPC). This project investigated a number of issues relating to the SPC measures, such as the framework for measuring capacity utilization and the means of aggregating capacity utilization from the plant to industry level. It aimed to better understand the behavior of these measures in the panel of plants and in the aggregate. It made recommendations on expanding and improving the published statistics deriving from the Survey of Plant Capacity. Expanding and improving the published statistics in the Survey of Plant Capacity could increase the value of this Survey at low incremental cost. It could greatly improve information about the utilization margin for economic and business analysis.

The SPC collects data on number of shifts, hours per shift, weeks of operation, production, and employment. These data are unique in providing time series information on capital hours and capital utilization. Yet, none of these data are publicly available. The published data from the SPC relate solely to subjective measures of capacity utilization reported by survey respondents, which was combined with aggregates from the ASM-CM. This project carried out research that could lead to greatly expanding the published aggregates (manufacturing and industries) of the data already collected by the SPC. The expansion of published aggregates similar to those constructed in this project should be facilitated through our thorough investigation and benefits write-up of alternative means of measuring capacity utilization, in addition to the comparison of our metrics to the Federal Reserve Board's capacity utilization and the published SPC tabulations.

This project addressed a number of conceptual and statistical issues related to measuring and aggregating data on capital utilization and its components (number of shifts, capital hours) using the SPC data. This research examined the properties of different schemes for weighting these data and make recommendations about what weighing scheme would work best for public release of aggregates.

This issue of aggregation depends critically on economic forces determining variation in capital utilization across plants and across time. Hence, the research's modeling of the determinants, such as organizational structure, industry, capital intensity, and productivity, of plants' rates of capital utilization is a critical input into addressing these statistical issues.

This research also addressed several measurement issues relating to combining the data from the SPC and the Annual Survey of Manufacturing (ASM) and Census of Manufacturing (CM). For example, testing if the ASM-CM capital measure is reliable for use in a SPC public-use capacity utilization measure. We found that historical cost of capital is the best practical measure of capital stock available for constructing sampling frames and weights for aggregation. The reason why historical cost of capital dominates other measures of capital stock (most notably, the replacement value of capital based on perpetual inventory) is because *a*) the data requirement is minimal (e.g., no need to construct time series of investment and disinvestment for any given plant, no need to construct disaggregated price indices for investment goods), and *b*) resulting series are highly correlated with other, more rigorous and economical justified measures of capital stock.

Capital stock, employment, and output measures from the ASM and CM are necessary for the analysis and aggregation of the SPC data. The SPC collects information that overlaps with the ASM/CM, such as production workers, so conflicts between the measurements need to be reconciled. We found that series reported in different surveys are high consistent with each other. Only in a handful of cases, we found clear inconsistencies in reported sales and employment. These inconsistencies appear to be related to typos in reported figures. When reported data covers different periods (e.g., reference periods of the SPC and the ASM/CM are different), we still find strong correlation between measures reported in different surveys.

There are a number of issues related to measurements in the ASM-CM that are important for this research on the SPC. Addressing the following is therefore important to producing valid measures of utilization.

i. The ASM-CM provides estimates of the capital stock that are critical for using the SPC to measure and aggregate capital utilization. This research evaluated whether these data are reliable enough to be used in public-use aggregates of SPC data. It was found that these data are highly reliable and should be used in the analysis of SPC data.

ii. The SPC has data on changes in capacity. These were validated against the ASM-CM measures of capital, other inputs, and production. It was found that capacity is strongly correlated with changes in inputs as well as changes in capital utilization.

iii. The SPC collects data on production workers that are also collected in the ASM-CM. This research investigated the consistency between these measures to assess their reliability and found that these measures are highly consistent. It was found that these measures are highly consistent and one may exploit these multiple reporting to verify the quality of the series.

iv. The SPC refers to the fourth quarter. The ASM-CM refer to the calendar year. The ASM-CM also asks for a quarterly breakdown of production-worker employment.

These data are not ideally matched, because the quarterly ASM-CM employment refers to a mid-quarter snapshot while the SPC totals employment over the entire quarter. It was found that these measures are highly correlated. Although it is possible to combine these measures and reduce the measurement issues, the gains appear to be small.

This project increased the Census Bureau's knowledge base regarding capital utilization. Several measures of capital utilization were estimated and provided in a technical memorandum to the Census Bureau. The various measures include different definitions of capacity utilization (i.e. plant hours, labor hours, shift number, employment, inputs) and various means of aggregating the definitions to the industry level. The research paper includes capital utilization as measured by plant hours, labor hours, number of shifts and workers, and by input materials and energy. These measures are aggregated by various weighting methods, such as output, total employment, capital, and largest shift. A series of horserace regressions were used to determine the quality of the various measures. It is shown that capital utilization plays an important role for measuring productivity and capacity utilization, predicting future changes in inputs.

There were two issues which were not delivered as promised. First, in the process of cleaning and analysis data, the author identified several data issues with the data available in the CES RDCs. Specifically, weights and reported values for 1997 and 1998 were distorted and clearly inconsistent with the data in adjacent years. Thought the efforts of CES staff and the authors, the data were fixed and estimates of capital utilization for 1997 were compiled. Final estimates for capital utilization in 1998 have not been compiled yet. The data for 1998 requires further cleaning and analysis to provide a final estimate of capital utilization. Second, the proposal planned to produce historical series for NAICS industries. However, the classification of plants from SIC to NAICS in 1970s and 1980s was particularly complicated and, due to time and other resource constraints, reliable industry classification has not been finished. However, the project produced industry estimates based on SIC classification, which was indicated in the proposal as a fallback option in case conversion to NAICS turns out to be complex.

II. Sample

The sampling frame is the Census of Manufacturing (CM). A fraction of firms (mainly, publishing) is sample from outside of CM. Every year, the sample is replenished with about 200-

300 firms. Survey of Plant Capacity (SPC) has sample weights (that are tailored to compute mean values of the plant capacity).

Before 1997, SPC largely overlaps with the Annual Survey of Manufacturers (ASM). After 1997, about 5,000 firms in SPC cannot be matched to ASM but can be matched to the most recent CM. In the CM years, almost all firms in the SPC can be matched to CM.

After 1997, SPC sample has two sub-samples. In the first subsample (about 6,000-7,000 firms), the survey collects basic demographics about the firms (e.g., idle, seasonality) and three levels of production (actual, potential, emergency). In the second sub-sample (8,000-10,000) firms, the survey additionally collects information about plant hours, workers, etc. (Item 3 in the survey form).

III. Capital stock

The basic measure of capital stock used in our analysis is the historical cost of structures (BAE in ASM/CM coding) and machines/equipment (MAE in ASM/CM coding) reported by establishments. The advantage of this measure is that it is readily available for reporting by firms and it is strongly correlated with measures of real capital stock constructed using perpetual inventory methods. In addition, this variable has been cleaned and, hence, is particularly useful in our analyses.

The main issue is that the overlap between SPC and ASM has been declining in recent years. Specifically, SPC no longer samples from ASM. Instead SPC sample firms from CM and additionally firms from some industries (e.g., printing) that are in scope for the SPC but not the CM. Since for many firms capital stock is a slow moving variable we focus on the capital stock (historical cost) reported in the Census years. In other words, for a firm in SPC 1994 we use the capital stock this firm reported in 1992 Economic Census.

We created measures of real capital stock based on perpetual inventory. When we construct real capital stock we use perpetual inventory separately for equipment and structures:

$$K_{t+1} = (1-\delta)K_t + I_t/PI_t, \qquad (1)$$

where K is capital stock, I is nominal investment, and PI is the price index for investment goods.[1] The price deflators for investment in new and used capital are from NBER productivity

[1] In earlier years ASM/CM collected information on retirements and sales of capital (structures and machines). We do not use this information to adjust our measures of capital because we do not have this information after early 1990s. So we do not take into account disinvestment in our perpetual inventory calculations.

dataset. This dataset in turn is based on the data from BEA. The NBER productivity data set is at 4-digit SIC level. 2-digit SIC industry depreciation rates are from BEA.

To start the perpetual inventory, we set the initial capital stock equal to the historical cost of capital in this firm times the ratio of replacement value of capital to historical value of capital in the industry to which the given firm belongs. If the firm does not report the initial historical stock of capital (it happens mainly in later years when BAE and MAE were collected only in Census year), we impute the real capital stock based on a regression of historical cost of capital on polynomials of investment, sales and employment. These regressions have all variables in logs, have year fixed effects and are industry and type-of-capital specific. In other words, for each industry and separately for structures and machines, we regress real capital stock on polynomials of sales and employment (incl. year fixed effects) and then take predicted values from this regression as initial values for perpetual inventory. When firms report only BAE+MAE, we use historical average in the industry for the share of structures in total capital stock to split BAE+MAE into BAE and MAE.[2]

Because ASM and SPC change the sample of firms every five years, we have firms with 5-year gaps in reporting of investment. ASM and SPC also add firms to the sample to compensate for attrition of the sample. For these types of firms, we have gaps in reported investment. We impute investment using several methods: i) industry median (conditional on positive investment), ii) regression based, iii) zero investment. In the first approach, we use the historical probability of making a positive investment to determine whether a firm makes an investment.[3] In the second approach, we use Tobit-type regression where explanatory variables are polynomials in sales and employment. In the third approach we set investment equal to zero if it is missing.

We run the perpetual inventory equation (1) forward and backward. The resulting real capital stock is in fixed 1987 dollars. See Table 2 for details on data availability in ASM/CM.

[2] Information on balance sheet value of equipment (structures) was collected annually before 1992 and only in census years afterwards. After 1997, only information on total assets is collected (no split between equipment and structures). Even for census years, only firms with "long forms" report assets.

[3] In other words, we draw a random value from U[0,1] and if the drawn value is greater than some threshold we assign a given firm with a positive investment. If the draw value is below the threshold, we assign zero investment to the firm. Conditional on positive investment, the firm receives the median investment in its 4-digit industry. This approach is aimed at capturing the fact that investment is bunched in spikes.

IV. NBER Productivity dataset

NBER/BLS/BEA data is coded in SIC-1987. This data set has level of real capital stock, labor, materials, output and price indices. The base year in price indices is 1987. The level of disaggregation is 4-digit SIC. The data runs from 1958 to 2002. Data for 2003-2004 are extrapolated using industry-specific AR(1) regression estimated on previous years.

We also use BEA data on historical and current value of capital stock (in current and fixed 1996 dollars) and depreciation rates by industries to initiate perpetual inventory when we compute capital stock. The data runs from 1972 to 2002. Data for 2003-2004 is extrapolated using projection on the constant and time trend. This regression is estimated industry-by-industry on last 10 years of the data.

V. Crosswalks between ASM/CM and SPC

We merge ASM/CM and SPC using permanent plant numbers (PPN). In recent years (after 2002 Census), US Census switched to a new system of plan identifiers. We use 2002 Census when both new and old plant identifiers are available to assign old PPNs to establishments after 2002. Note that SPC switched to new plant identifiers in 2004. In rare cases when PPN was missing or incorrectly coded, we use other firm identifiers (employer identification number, alpha-code, regional information, etc.) to find the correct PPN.

Pulling all years together, we construct a bridge between this plant identifier to the permanent plant number. This dictionary has about 300,000 firms that ever appeared in SPC.

In non-census years, about 6,000 to 10,000 plants in SPC can be matched to plants in ASM. In census years, the match is almost complete (in the worst 1997 year, about 1,000 firms in SPC cannot be matched to CM). Because surveys used different plant identifiers in 2002-2004, the quality of the match is somewhat lower than for other years. The worst match is for 2004, when SPC plant identifiers allow to match only about 6,000 firms. The cross-walk is available in the project folder.

VI. Industry classification

We use 1987 Standard Industrial Classification to associate plants with industries. We had to recode industries from NAICS and other vintages of SIC. Our recoding procedure has several steps. First, for industries with one-to-one bridges between 1987 SIC and other industry classification, we assigned the industry based on the industry classification (either NAICS or other vintages of SIC) reported in the corresponding year. Second, for industries with no one-to-one bridges, we use the 1987 SIC code available in the closest adjacent year. Third, for firms which had no 1987 SIC code and had one-to-many mapping from the reported industry code to 1987 SIC, we use probabilities assignment into 1987 SIC industries.

For example, a plant in industry X in NAICS could be in industries Y1 and Y2 in 1987 SIC. From 1997 Census, we know the shares of Y1 and Y2 plants. Suppose that industry Y1 is twice as large as industry Y2. Then the probability of any given plant with NAICS classification to be assigned into industry Y1 is twice as large as the probability of being assigned into industry Y2. In short, with probabilities equal to shares of Y1 and Y2 we assign this plant into Y1 or Y2 1987 SIC industry. This probabilistic assigned was done for relatively few firms before 2004 and for about 50% of firms in 2004. We ensure that plants once assigned a 1987 SIC code keep the code as long as they continue to have the same NAICS or other-vintage SIC code. This prevents plants from jumping across industries over time when their original industry codes do not change.

The level of disaggregation for industry is 4-digit SIC. However, we are more confident in the correctness of the classification at the 3-digit level SIC. There is little (if any) space for industry misclassification at the 2-digit level.

VII. Cleaning capital utilization measures

Table 1 summarizes data availability.
Four variables are constructed from the raw data.

- plant hours per week (phw)
- plant hours per day (phd)
- days per week in operation (pod)
- weeks per Q4 in operation (pow)

Output in the 4th quarter is available for all years. For all surveys, we use the data edited by the US Census staff. The edited data removes inconsistencies and errors in the raw data.

1974-1988

Variables:

- pod is taken directly from the data.
- phd is taken directly from the data.
- pow is taken directly from the data (for 1979-1988)
- phw is computed as pod $\times$ phd.

Cleaning/Recoding:

- Recode pod to 7 if pod is greater than 7.

- Recode hours per day to 24 if hours per day is greater than 24 and reported hours per week is equal to 24 times the number of days in operation.
- Recode hours per day to 16 if hours per day is greater than 24 and reported hours per week is equal to 16 times the number of days in operation
- Recode hours per day to 8 if hours per day is greater than 24 and reported hours per week is equal to 8 times the number of days in operation
- Recode hours per day to the rounded value of reported hours divided by days in operation if the number of hours is greater than 24.
- Recode the number of shifts to three if the number of shifts is greater than 3.
- Recode the number of weeks to 13 if the number of weeks is greater than 13.

1989-1994

Special features:

- The survey was collected biannually. The data for odd years was collected retrospectively.

Variables:

- pod is taken directly from the data.
- phd is taken directly from the data.
- pow is not available.
- phw is computed as pod × phd.

1995-1996

Variables:

- pod is taken directly from the data.
- phd is computed as (phw/pod).
- pow is not available.
- phw is taken directly from the data.

Cleaning/Recoding:

- Recode phd to 24 if phd is greater than 24.
- Recode phw to 40 if phw equals 200 and pod equals 5.
- Recode phw to missing value if phw is greater than 168.

1997

The data file in the RDC was corrupted. We requested re-creation of the file and obtained the new file in Nov 2008.

Special features:

- From this year onward firms are classified according to NAICS. In 1997 survey, there is a bridge between SIC(1987) and NAICS. For some of the subsequent years, the survey keeps SIC(1987) classification.

Variables:

- pod is taken directly from the data.
- phd is computed as (phw/pod).
- pow is taken directly from the data.
- phw is taken directly from the data.

Cleaning/Recoding:
- Recode phd to 24 if phd is greater than 24.
- Recode phw to 40 if phw equals 200 and pod equals 5.
- Recode phw to missing value if phw is greater than 168.

1998-2004

Special features:
- The data file for SPC 1998 in the RDC was corrupted. We requested re-creation of the file and obtained the new file in Nov 2008.
- 1998 was the last year based on 1992 CM sample.
- 1998 was the first year when firms were requested to report hours, employment, etc. by ***shift***.
- In 2004, the SPC survey changes the plant identifier. For about 3,000 plants the link to past years or ASM/CM is hard to establish.

Variables:
- pod is taken directly from the data. Because information is reported by shift, total pod is computed as the *maximum* number of days in operation across shifts.
- phd is computed as (phw/pod).
- pow is taken directly from the data. Because information is reported by shift, total pow is computed as the *maximum* number of weeks in operation across shifts.
- phw is taken directly from the data. Because information is reported by shift, total phw is computed as the *sum* of plant hours for all shifts.

Cleaning/Recoding:
- Starting with 1998, we do ***massive*** cleaning of the data. The main problem is that firms often report hours that are not consistent with 168 maximum workweek. Common instances are when for each shift the respondents put 168 hours/week. Other common problems are when each shift is below 168 but the total over all shifts is above 168: 100 hours/week 1st shift, 70 hours/week 2nd shift, 40 hours/week 3rd shift. Also there are massive inconsistencies between reported days in operation and reported hours per shift. For example, a firm may report 100 hours/week in the first shift, but only 5 days in operation. Sometimes, hours in the 1st shift is equal to the sum of hours in the 2nd and 3rd shifts, as if respondents put the total in the 1st column of the survey form.[4]

  Below we describe the sequence of step we take to clean/recode the data
  1) If 168 plant hours per week are reported in any shift, recode each shift to 56 plants hours per week.

[4] We strongly recommend to change the form and introduce the column total (over all shifts) and clearly indicate natural limits on reported statistics.

2) If 144 plant hours per week are reported in any shift, recode each shift to 48 plants hours per week.
3) If 120 plant hours per week are reported in any shift, recode each shift to 40 plants hours per week.
4) If 84 plant hours per week are reported in any shift and only two shifts are indicated as active, recode hours in the inactive shift as zero.
5) If 72 plant hours per week are reported in any shift and only three workdays per week, recode hours in each shift to 24 hours/week.
6) If a firm reported zero days in operation for a given shift and total hours across shifts exceed 168 per week, set shift plant hours per week in inactive shifts to zero.
7) If a firm reports hours in the 1st shift ***equal*** to the sum of hours in the 2nd and 3rd shifts and total hours across shifts exceed 168 hours/week, we set hours reported in the 1st shift to hours in the 2nd shift, 2nd shift equal to 3rd shift, and 3rd shift to missing. [It looks like firms put total in the first column]
8) If a firm reports hours in the 1st shift ***greater*** than the sum of hours in the 2nd and 3rd shifts, total hours across shifts exceed 168 hours/week and hours in the 1st sift exceed 100, we set hours reported in the 2nd and 3rd shifts to zero.
9) If a firm reports hours in the 1st shift ***greater*** than the sum of hours in the 2nd and 3rd shifts and hours in the first shift is greater than (days in operation for the 1st shift)x12, we set hours reported in the 2nd and 3rd shifts to zero.
10) If a firm reports hours in the 1st shift ***greater*** than (days in operation for the 1st shift)x12, we set hours reported in the 2nd and 3rd shifts to zero.
11) If a firm reports hours in the 1st shift ***greater*** than 56 and equal to hours in the 2nd and 3rd shifts, we set hours reported in the 2nd and 3rd shifts to zero.
12) If a firm reports hours equal to 168, the same number of days in operation for each shift and the number of days in operation is less than 7, set hours in each shift equal to 8x(days in operation per week).
13) If a firm reports hours equal to 168 and one of the shifts has less than 7 days in operation, set hours in each shift equal to 8x(days in operation per week for a given shift).
14) If plant hours per week are greater than 168, set hours for each shift equal to 8x(days in operation).
15) If plant hours per week are greater than 168, set plant hours per week to 168.

- Recode phd to 24 if phd is greater than 24.
- Recode pod to 7 if pod>7.
- Recode pow to 13 if pow is greater than 13.

Cleaning has large effects on the moments of plant hours per week. In the raw data, the mean plant hours per week exceed 200. In the cleaned data, the mean plant hours is about 90-100 which is comparable to plant hours per week reported in previous years.

VIII. Construction of capital utilization series

We construct three series of capital utilization:

- Average plant hours per week (PHW)

- Average plant hours per day (PHD)
- Average number of plant days in operation (POD)

As we discuss in the proposal, there are several ways to construct the series at different levels of aggregation. Specifically, we consider several weights to construct the series. Let *i* and *t* index establishments and time (year). Denote a reported measure of capital utilization with KU_{it} (i.e., KU can be PHW, PHD or POD); sample weights with w_{it}; a measure of output with Y_{it}; a measure of capital stock with K_{it}; a measure of employment with L_{it}. We construct the following series:

- unweighted: $\overline{KU}_t^{(nowgt)} = (1/N_t)\sum_i KU_{it}$;
- weighted by sample weight: $\overline{KU}_t^{(wgt)} = \sum_i KU_{it}w_{it} \Big/ \sum_i w_{it}$;
- weighted by sample weights and output (here, total value of shipments from the most recent Census of Manufacturers and from the current Survey of Plant Capacity): $\overline{KU}_t^{(Y)} = \sum_i KU_{it}w_{it}Y_{it} \Big/ \sum_i w_{it}Y_{it}$;
- weighted by sample weights and capital stock (here, balance sheet value of fixed assets from the most recent Census of Manufacturers; replacement value of capital with and without imputations for missing values): $\overline{KU}_t^{(K)} = \sum_i KU_{it}w_{it}K_{it} \Big/ \sum_i w_{it}K_{it}$;
- weighted by sample weights and employment (here, total number of employees and total hours of production workers): $\overline{KU}_t^{(L)} = \sum_i KU_{it}w_{it}L_{it} \Big/ \sum_i w_{it}L_{it}$.

Our series $\overline{KU}_t^{(nowgt)}, \overline{KU}_t^{(wgt)}, \overline{KU}_t^{(Y)}, \overline{KU}_t^{(K)}, \overline{KU}_t^{(L)}$ (plus associated variance, standard errors, and the number of plants) are for 1974-2004 at three levels of disaggregation: macroeconomic, 2-digit SIC industry level, selected 3-digit SIC industry level.

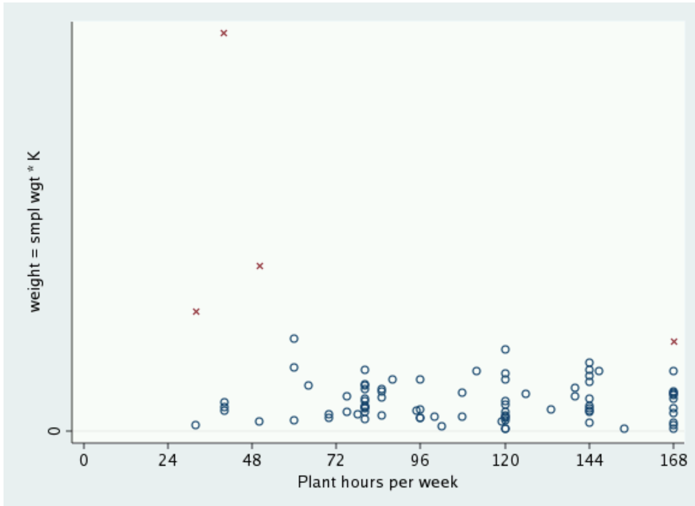
Since the distributions of sales, capital and employment are strongly skewed, we also construct series adjusted for influential observations. To understand the reason for this request, consider $\overline{KU}_t^{(Y)}$. The effective weight for plant *j* in this measure is given by $w_{jt}Y_{jt}/\sum_i w_{it}Y_{it}$. Note that sample weight w_{jt} and measure of output Y_{jt} can lead to a very large weight $w_{jt}Y_{jt}$. This is not desirable because time series can be dominated by reporting errors, unusual events and other irregularities so that the resulting aggregate time series can be choppy.

We apply the following procedure to limit the effect of extreme observations. In the first screening step, we jackknife (i.e., drop one plant at a time) capital utilization for a given industry/year and compute the statistics of capital utilization. Suppose that this industry and year has N plant observations. After applying the jackknife, we have N values of the capital utilization statistic. The distribution of this statistic informs us about the effect of any given observation on the statistic we want to report (e.g., mean value of capital utilization rate). We say that an observation is potentially influential if the statistic computed without this observation is outside the range of median plus/minus four interquantile ranges where median and interquantile range are computed on the basis of N statistics from the jackknife. We say that an observation is influential if it is potentially influential and its weight $w_{jt}Y_{jt}$ is above 90th percentile of the weights $w_{jt}Y_{jt}$ for a given industry and year. In short, influential observations have very large weights $w_{jt}Y_{jt}$ and move the industry level measure of capital utilization by large amounts. For influential observations, we censor the weight $w_{jt}Y_{jt}$ to be equal to the 90th percentile of the weights $w_{jt}Y_{jt}$ for a given industry and year.

Note that we do not drop influential observations and we do not recode firm-level measures of capital utilization for influential observations.

Figure 1 illustrates the importance of controlling for observations with unusually large effective weights for some industry in 1991: censoring the weights at the 90^{th} percentile increases average plant hours by almost 10 hours.

Figure 1. Effect of influential observations.



Notes: Crosses denote influential observations with large weights. Circles are regular observations. Mean plant hours per week in this industry is 106 without influential observations and 97 with influential observations.

Table 1. Data availability in SPC.

Survey year	Temp idle flag	Hours per day	Days per week	Plant hours per week	Weeks in Q4	Number of shifts per day	Total hours	Overtime hours	Temporary workers	Production workers
1974-1978	X	X	X				X	X		X
1979-1988	X	X	X		X		X	X		X
1990	X		X	X						
1992	X		X	X						
1994	X		X	X						
1996	X		X	X						
1997	X		X	X	X	X	X	X		X
1998-2004	X		X	X	X	X	X	X	X	X

Table 2. DATA AVAILABILITY IN THE ANNUAL SURVEY/CENSUS OF MANUFACTURES.

Capital		Employment		Materials		Output	
Variable	Years available	Variable	Years available	Variable	Years available	Variable	Years available
Building assets b.o.p.	74-85,87,92	Total employment	all	Cost of materials	all	Total value of shipments	all
Machine assets b.o.p.	74-85,87,92	Non-production employees	all	Cost of materials: parts	all	Receipts from contract work	all
Total assets b.o.p.	74-85,87,92	Production workers (average)	72-95, CM	Cost of resales	all	Misc. receipts	all
Building assets e.o.p.	72-85,87,92	Production workers Mar	72-95, CM	Cost of contract work	all	Value added	all
Machine assets e.o.p.	72-85,87,92	Production workers May	72-95, CM	Material delivered costs	all	Export	all
Total assets e.o.p.	72-85,87,92	Production workers Aug	72-95, CM	Cost of fuels	all	Product value of shipments	CM years
Buildings retirement	77-85,87, 92	Production workers Nov	72-95, CM	Cost of purchased electricity	all	Product quantity shipments	CM years
Machine retirement	77-85,87, 92	Hours of production workers	72-95, CM	Purchased Electricity	all		
Buildings depreciation	77-85,87, 92	Hours of prod. workers Mar	72-80, 87, CM	Generated electricity	all		
Machine depreciation	77-85,87,92	Hours of prod. workers May	72-80, 87, CM	Fuels: similar to electricity	all		
Expenditures on new machines	all years	Hours of prod. workers Aug	72-80, 87, CM				
Expenditures on new buildings	all years	Hours of prod. workers Nov	72-80, 87, CM				
Expenditures on used buildings	77-96	Total salaries and wages	all				
Expenditures on used machines	77-96	Production worker wages	all				
Expenditures on used capital	all years	Non production worker wages	all				
Total capital expenditures	all years	Total supplemental labor costs	ASM years				
Rental payments on buildings	72-85,87, 92						
Rental payments on machines	72-85,87, 92						
Total rental payments	72-85,87, 92						