This document was generated: October 11, 2024 at 11:59 PM

NBER-CES Manufacturing Industry Database (2009) [SIC]

(24 variables)
Last update to metadata: 2015-11-26 10:30:29 (auto-generated)
Document Date: February 4, 2015

Codebook prepared by: Cornell NSF-Census Research Network

Data prepared by: Randy A. Becker , Wayne B. Gray , Jordan Marvakov , and Eric J. Bartelsman

Principal Investigator(s): Randy A. Becker , Wayne B. Gray , and Jordan Marvakov

Citation

Please cite this codebook as:

Cornell NSF-Census Research Network. Codebook for the NBER-CES Manufacturing Industry Database (2009) [Codebook file]. Cornell Institute for Social and Economic Research and Labor Dynamics Institute [distributor]. Cornell University, Ithaca, NY, 2013

Please cite this dataset as:

Becker, Randy A., Wayne B. Gray, Jordan Marvakov. NBER-CES Manufacturing Industry Database: June 2013 Revision [Computer file]. Boston, MA; National Bureau of Economic Research [distributor], Boston, MA, 2013

Abstract

This database is a joint effort between the National Bureau of Economic Research (NBER) and U.S. Census Bureau's Center for Economic Studies (CES), containing annual industry-level data from 1958-2009 on output, employment, payroll and other input costs, investment, capital stocks, TFP, and various industry-specific price indexes. Because of the change from SIC to NAICS industry definitions in 1997, the database is provided in two versions: one with 459 four-digit 1987 SIC industries and the other with 473 six-digit 1997 NAICS industries.

Datasets

sic5809.dta http://www.nber.org/data/nberces5809.html (Stata)

sic5809.sas7bdat http://www.nber.org/data/nberces5809.html (SAS)

sic5809.xls http://www.nber.org/data/nberces5809.html (Excel spreadsheet)

sic5809.csv http://www.nber.org/data/nberces5809.html (CSV)

Terms of Use

Access Levels

released

No description given

restricted

No description given

Access Conditions

The data providers ask that you please not forward this data on to others. Instead, please refer interested parties to http://www.nber.org/data/nberces5809.html . This ensures that everyone receives the latest version of the data that's available.

Citation Requirements

When using the data, please cite it according to bibliographic citation provided.

Contact

For questions regarding this data collection, please contact: Please forward all inquiries to: Wayne Gray, Clark University and NBER, wgray@clarku.edu or Randy Becker, Center for Economic Studies, U.S. Census Bureau, randy.a.becker@census.gov

Additional Information

Methodology

The NBER-CES Manufacturing Industry Database contains annual data from the United States manufacturing sector for the period from 1958 to 2009. The data used for the development of the database come from various sources, but chiefly from four government agencies: the Bureau of the Census, the Bureau of Economic Analysis, the Bureau of Labor Statistics, and the Federal Reserve Board. The goal is to provide a long time-series of data for a large number of industries, adjusting for changes in industry definitions, and creating price deflators, capital stocks, and productivity estimates. This paper describes issues related to the most recent update of the database (1997 to 2009). For a more detailed discussion of an older version of the data see Bartelsman and Gray (1996). The key feature (and complication) of recent updates is the adjustment to the NAICS industry definitions. It is published in two versions: one based on the 1987 Standard Industrial Classification (SIC), containing 459 industries, and the other based on the 1997 North American Industrial Classification System (NAICS), containing 473 industries. The boundaries of the manufacturing sector shifted between SIC and NAICS, with some SIC industries leaving manufacturing and other NAICS industries entering manufacturing. The exiting SIC industries don't have data after 1996, and the entering NAICS industries don't have data before 1997. To enable the calculation of price deflators in each version, the benchmark year for deflators in the NAICS version is 1997, while the SIC version continues to use 1987 as a benchmark. This version of the database recalculates all the data post-1996, and earlier updates of the database (1958-2002 and 1958-2005) are superseded and should not be used. For more information, see methodology in http://www.nber.org/nberces/t0205.pdf , technical

notes in http://www.nber.org/nberces/nberces5809/nberces_5809_technical_notes.pdf , information on Concordances (1987 SIC - 1997 NAICS in Stata , SAS , Excel , CSV , PDF , 1972 SIC - 1987 SIC in Stata , SAS , Excel , CSV ,).

Sources

I. Annual Survey of Manufactures (ASM)

II. Federal Reserve Board data (FRB)

III. Bureau of Economic Analysis (BEA) GDP-by-Industry data

IV. 2002 Manufacturing Energy Consumption Survey (MECS)

V. Bureau of Labor Statistics database

Related Publications

I. Eric J. Bartelsman and Wayne Gray, 1996. "The NBER Manufacturing Productivity Database," NBER Technical Working Papers 0205, National Bureau of Economic Research, Inc. http://www.nber.org/papers/t0205.pdf

Variable Groups - NBER-CES Manufacturing Industry Database (2009) [SIC]

Codebook does not contain variable groups.

Variable Name

sic

Label

1987 Stand. Ind. Classification

Concept

Type

numeric

Files

F1csv F1dta F1sas F1xls

Full Description

1987 SIC 4-digit Codes that identify observations in the dataset.

Summary Statistics

Valid values	23868
Invalid values	0
Minimum	2011
Maximum	3999
Mean	3039.194
Standard deviation	601.456

Value Ranges

Value Range

Range: [2011 , 3999]

Variable Name	year
Label	Year ranges from 1958 to 2009
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Summary Statistics

Valid values	23868
Invalid values	0
Minimum	1958
Maximum	2009
Mean	1983.5
Standard deviation	15.009

Value Ranges

Value Range

Range: [1958 , 2009]

Variable Name	emp
Label	Total Employment in 1000s
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

Number of employees (in 1000s). This includes both production workers (average of reported employment in March, May, August and November) and other workers (only March figure). It does not include employees in the auxiliary (administrative) units, which are reported separately.

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.2
Maximum	565.4
Mean	36.838
Standard deviation	51.612

Value Ranges

Value Range

Range: [0.2 , 565.4]

Variable Name	pay
Label	Total payroll in $1m
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

Total payroll in millions of dollars. This does not include social security or other legally mandated payments, or employer payments for some fringe benefits.

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	5
Maximum	22245.3
Mean	768.116
Standard deviation	1458.324

Value Ranges

Value Range

Range: [5 , 22245.3]

Variable Name	prode
Label	Production workers in 1000s
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

Number of production workers (in 1000s). This excludes supervisors above the line-supervisor level, clerical, sales, office, professional, and technical workers. Plants report the number of production workers for four separate pay periods (including the 12th day of March, May, August and November), and these are averaged together.

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.1
Maximum	464.5
Mean	26.609
Standard deviation	37.68

Value Ranges

Value Range

Range: [0.1 , 464.5]

Variable Name

prodh

Label

Production worker hours in 1m

Concept

Type

numeric

Files

F1csv F1dta F1sas F1xls

Full Description

Number of production worker hours (in millions of hours). This includes all hours worked or paid for, including actual overtime hours, but excluding vacation, holidays or sick leave.

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.2
Maximum	912.9
Mean	52.917
Standard deviation	75.255

Value Ranges

Value Range

Range: [0.2 , 912.9]

Variable Name

prodw

Label

Production worker wages in $1m

Concept

Type

numeric

Files

F1csv F1dta F1sas F1xls

Full Description

Production worker wages (millions of dollars).

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	3.5
Maximum	16566.6
Mean	459.051
Standard deviation	900.455

Value Ranges

Value Range

Range: [3.5 , 16566.6]

Variable Name	vship
Label	Total value of shipments in $1m
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

Value of industry shipments (millions of dollars). These are based on net selling values, f.o.b. plant, after discounts and allowances. This includes receipts for contract work and miscellaneous services provided by the plant to others. VSHIP is not adjusted for inventory changes, and can include large amounts of duplication across industries or even across plants in the same industry (especially machinery and transportation industries that often include production of parts as well as final products).

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	19.3
Maximum	732728.4
Mean	4980.543
Standard deviation	14831.598

Value Ranges

Value Range

Range: [19.3 , 732728.4]

Variable Name	matcost
Label	Total cost of materials in $1m
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

Cost of materials (millions of dollars). This includes the total delivered cost of raw materials, parts, and supplies put into production or used for repair and manteinance, along with purchased electric energy and fuels consumed for heat and power, and contract work done by others for the plant. This excludes the costs of services used, overhead costs, or expenditures related to plant expansion. Because MATCOST includes energy spending, to calculate spending on non-energy materials one must use (MATCOST-ENERGY).

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	7
Maximum	648048.4
Mean	2706.393
Standard deviation	10763.527

Value Ranges

Value Range

Range: [7 , 648048.4]

Variable Name

vadd

Label

Total value added in $1m

Concept

Type

numeric

Files

F1csv F1dta F1sas F1xls

Full Description

Value added by manifacture (millions of dollars). This equals VSHIP - MATCOST + change in finished goods and work-in-process inventories during the year.

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	10.2
Maximum	111187.9
Mean	2286.422
Standard deviation	5249.453

Value Ranges

Value Range

Range: [10.2 , 111187.9]

Variable Name	invest
Label	Total capital expendture in $1m
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

Total capital spending (millions of dollars). This variable underwent a substantial change in definition. The ASM reported new capital expenditures before 1997 (with used capital expenditures in an additional table), but switched to reporting the sum of new and used expenditures in 1997, with no separate breakout of new and used expenditures (even in Census years). In order to maintain consistency with the earlier data, we adjusted the post-1996 data to reflect new expenditures. The data on new and used expenditures from 1987, 1992, and 1996 were combined to derive a fraction new (new/(new+used)) for each industry, which was then used to adjust the post-1996 capital expenditure numbers to reflect only new expenditures.

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.1
Maximum	17601.6
Mean	162.085
Standard deviation	510

Value Ranges

Value Range

Range: [0.1 , 17601.6]

Variable Name

invent

Label

End-of-year inventories in $1m

Concept

Type

numeric

Files

F1csv F1dta F1sas F1xls

Full Description

End-of-year inventories (millions of dollars). Before 1982 this was based on any generally accepted accounting method. Beginning in 1982 this is based either at cost or at market, with LIFO users asked to report pre-adjustment values.

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	1.3
Maximum	40084.9
Mean	602.765
Standard deviation	1492.782

Value Ranges

Value Range

Range: [1.3 , 40084.9]

Variable Name

energy

Label

Cost of electricity and fuels in $1m

Concept

Type

numeric

Files

F1csv F1dta F1sas F1xls

Full Description

Expenditures on purchased fuels and electrical energy (millions of dollars).

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.1
Maximum	14201.5
Mean	100.26
Standard deviation	387.328

Value Ranges

Value Range

Range: [0.1 , 14201.5]

Variable Name

cap

Label

Total real capital stock in $1m

Concept

Type

numeric

Files

F1csv F1dta F1sas F1xls

Full Description

Real capital stock (millions of 1987 dollars). This equals (EQUIP + PLANT).

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	3.7
Maximum	95007.2
Mean	2399.676
Standard deviation	5750.494

Value Ranges

Value Range

Range: [3.7 , 95007.2]

Variable Name

equip

Label

Real capital: equipment in $1m

Concept

Type

numeric

Files

F1csv F1dta F1sas F1xls

Full Description

Real equipment capital stock (millions of 1987 dollars).

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.9
Maximum	69967.6
Mean	1443.159
Standard deviation	3740.439

Value Ranges

Value Range

Range: [0.9 , 69967.6]

Variable Name

plant

Label

Real capital: structures in $1m

Concept

Type

numeric

Files

F1csv F1dta F1sas F1xls

Full Description

Real structures capital stock (millions of 1987 dollars).

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	2
Maximum	30641.7
Mean	956.518
Standard deviation	2173.42

Value Ranges

Value Range

Range: [2 , 30641.7]

Variable Name

piship

Label

Deflator for VSHIP 1987=1.000

Concept

Type

numeric

Files

F1csv F1dta F1sas F1xls

Full Description

Shipment deflator (VSHIP). Bureau of Economic Analysis (BEA) GDP-by-Industry data for price indexes used in the calculation of the price deflator for value of shipments (equals 1 in 1987).

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.009
Maximum	9.473
Mean	0.903
Standard deviation	0.614

Value Ranges

Value Range

Range: [0.00899999961256981 , 9.47299957275391]

Variable Name	pimat
Label	Deflator for MATCOST 1987=1.000
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

The materials cost deflators are calculated using data from the use-make (input-output) tables and the GDP-by-Industry data of the Bureau of Economic Analysis (BEA). The raw data are available at the BEA website (http://www.bea.gov) under the Industry tab. The use-make tables disaggregate each industry's total materials cost into the amount spent on each specific material. The share of each material in the industry's total materials cost is then used to weight each material's price index, resulting in a weighted-average price index for the industry's total materials cost. The benchmark versions of the use-make tables, which are generated every 5 years, are used. A moving average of the weights from the two surrounding benchmark years is used to interpolate the weights for the use-make tables in the intervening years. Inputs coming from manufacturing industries use the price index in the database for that industry's shipments. Inputs coming from other (non-manufacturing) sectors use that sector's price index for gross output.

The main difficulty with implementing the above procedure is the lack of a clear one-to-one correspondence between the industry codes used in the use-make tables and those used in the ASM/CMF data. The use-make tables contain many fewer industries than the NAICS industries. BEA does provide concordance tables, which are supplemented with some manual matching based on industry definitions. We assume that all NAICS industries which map into the same use-make industry have the same structure of inputs.

To match MATCOST, this is a deflator for all materials, including energy. If you need a separate 'non-energy materials' deflator, you can construct an implicit one, based on (MATCOST/PIMAT) - (ENERGY/PIEN).

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.154
Maximum	3.549
Mean	0.845

Standard deviation 0.413

Value Ranges

Value Range

Range: [0.153999999165535 , 3.54900002479553]

Variable Name	piinv
Label	Deflator for INVEST 1987=1.000
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

Price deflator for new investment (equals 1 in 1987). This combines separate deflators for structures and equipment, based on the distribution of each type of asset in the industry. Note that this is a deflator for new investment flows, not for the existing capital stock. The ASM provides annual data for gross nominal capital expenditures on equipment and structures. However, it lacks information on depreciation rates and investment deflators, which are necessary in order to construct the real capital stocks in the database.

Depreciation rates are derived from Federal Reserve Board data (FRB). Depreciation rates are calculated for each industry, backing them out of a standard perpetual inventory equation: `K_t = (1-d_{t-1})K_{t-1} + I_t`, where `d` is the rate of depreciation, `K` is the real capital stock in years `t` and `t-1`, and `I` is the real capital investment in year `t`.

To calculate the investment deflator, price indexes for equipment and structures are computed from the ratios of nominal and real investment for the 4-digit NAICS industries in the FRB data. The final deflator values are created as the weighted mean of the two types of capital expenditures. The shares used for the weights are averaged over years `t` and `t-1`. For detailed information please refer to http://www.nber.org/nberces/nberces5809/nberces_5809_technical_notes.pdf

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.236
Maximum	1.605
Mean	0.803
Standard deviation	0.362

Value Ranges

Value Range

Range: [0.236000001430511 , 1.60500001907349]

Variable Name	pien
Label	Deflator for ENERGY 1987=1.000
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

Price deflator for energy (equals 1 in 1987). The energy deflator is based on each industry's expenditures on seven types of energy: electricity, residual fuel oil, light oils, liquefied petroleum, coal, coke, and natural gas. Previous updates used information for six types of energy (the distillates category was recently split into light oils and liquefied petroleum).

As with the materials cost deflator, we combine price indices for the different energy types with information on each energy type's share in the industry's total energy expenditures. Data on energy expenditures for six fuels (excluding electricity) come from 2002 Manufacturing Energy Consumption Survey (MECS). The data on energy expenditures for six of the fuels in the current update (excluding electricity) come from the 2002 Manufacturing Energy Consumption Survey (MECS), available at the Bureau of the Census website (http://www.census.gov). The ASM data provide information on electricity and non- electricity energy spending, and the MECS data are used to disaggregate each industry's non-electricity spending share into the share of each of the six non-electricity energy types. There are a number of missing values in the MECS data. Extremely small values (less than 0.5), as well as values with high relative standard error (more than 50%) are withheld from the official MECS release by Census. Additionally, some values are withheld to avoid disclosing information for individual establishments in cases where an industry relatively few plants. The extremely small values are replaced by 0.1. Price indexes for each non-electricity energy type are obtained from the Bureau of Labor Statistics database (http://www.bls.gov), using the industrial price indexes. The codes of the extracted BLS variables are:

- WPU051 Coal
- WPU052 Coke oven products
- WPU0532 Liquefied petroleum gas
- WPU054321 Industrial electric power
- WPU055321 Industrial natural gas
- WPU0573 Light fuel oils
- WPU0574 Residual fuels

Since the 1996 update, the WPU052 Coke oven products series has been discontinued. Thus, the coke price data used here are extrapolated from coal prices. Electricity prices are calculated from the industry- specific electricity consumption and expenditures data in the ASM tables, where available, and otherwise taken from the national-average values.

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.102
Maximum	2.565
Mean	0.835
Standard deviation	0.483

Value Ranges

Value Range

Range: [0.101999998092651 , 2.56500005722046]

Variable Name	dtfp5
Label	5-factor TFP annual growth rate
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

There are two versions of TFP in the NBER-CES Manufacturing Database: 4-factor and 5-factor. The 5- factor version separates out energy from non-energy materials; the 4-factor uses a single materials input (which includes energy). The TFP calculation requires definitions of the cost shares, the factors, the factor changes, and the output changes. The five cost shares (`alpha_i`) vary by industry by year, defined using the variable names in the dataset:

- (`alpha_1`) Non-production workers: (pay-prodw)/vship [i.e., their pay divided by shipments]
- (`alpha_2`) Production workers: prodw/vship [i.e., their pay divided by shipments]
- (`alpha_3`) Energy: energy/vship [i.e., energy expenditure divided by shipments]
- (`alpha_4`) Materials: (matcost-energy)/vship [i.e., non-energy materials divided by shipments]
- (`alpha_5`) Capital: 1 - (sum of the above shares) [i.e., the residual]

In calculating TFP growth from one year to the next, we use the average of the two years' cost shares: `bar alpha_{it} = 0.5alpha_{it} + 0.5alpha_{it-1}` The 5 factors (`X_i`) are defined as follows, using the variable names in the dataset:

- (`X_1`) Non-production workers: (emp-prode) [i.e., the number of non-production workers]
- (`X_2`) Production workers: prodh [i.e., production worker hours, not employees]
- (`X_3`) Energy: (energy/pien) [i.e., real energy expenditures]
- (`X_4`) Materials: ((matcost/pimat) - (energy/pien)) [i.e., real non-energy materials]
- (`X_5`) Capital: cap [i.e., total capital stock, already in real terms]

The change in factor usage between one year and the next is defined as the change in natural logs, (for example): `dX_{it} = ln(X_{it}) - ln(X_{it-1})`

We also need the change in real output: (Q) Real output: *vship/piship*

As with factor usage, we express output change in terms of natural logs, hence: `dQ_t = ln(Q_t) - ln(Q_{t-1})`

The change in 5-factor TFP (dTFP5) between this year and last is thus defined as: `dTFP5_t = dQ_t - sum_i(alpha'_{it}dX_{it})`, `i = 1,..., 5`

Given the series of dTFP5 values, one can then "roll up" these changes to form a TFP index (TFP5), by setting the index equal to 1.0 in some initial year `t` and then growing the index forward by the following equation: `TFP5_{t+1} = exp[ ln(TFP5_t) + (dTFP5_{t+1})]`

The values of 4-factor TFP growth (dTFP4) and the corresponding TFP index (TFP4) are calculated similarly, but using total materials cost spending rather than separating it into energy and non-energy materials.

Summary Statistics

Valid values	23300
Invalid values	568
Minimum	-0.641
Maximum	1.387
Mean	0.00301
Standard deviation	0.0684

Value Ranges

Value Range

Range: [-0.640999972820282 , 1.38699996471405]

Variable Name	tfp5
Label	5-factor TFP index 1987=1.000
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

There are two versions of TFP in the NBER-CES Manufacturing Database: 4-factor and 5-factor. The 5- factor version separates out energy from non-energy materials; the 4-factor uses a single materials input (which includes energy). The TFP calculation requires definitions of the cost shares, the factors, the factor changes, and the output changes. The five cost shares (`alpha_i`) vary by industry by year, defined using the variable names in the dataset:

- (`alpha_1`) Non-production workers: (pay-prodw)/vship [i.e., their pay divided by shipments]
- (`alpha_2`) Production workers: prodw/vship [i.e., their pay divided by shipments]
- (`alpha_3`) Energy: energy/vship [i.e., energy expenditure divided by shipments]
- (`alpha_4`) Materials: (matcost-energy)/vship [i.e., non-energy materials divided by shipments]
- (`alpha_5`) Capital: 1 - (sum of the above shares) [i.e., the residual]

In calculating TFP growth from one year to the next, we use the average of the two years' cost shares: `bar alpha_{it} = 0.5alpha_{it} + 0.5alpha_{it-1}` The 5 factors (`X_i`) are defined as follows, using the variable names in the dataset:

- (`X_1`) Non-production workers: (emp-prode) [i.e., the number of non-production workers]
- (`X_2`) Production workers: prodh [i.e., production worker hours, not employees]
- (`X_3`) Energy: (energy/pien) [i.e., real energy expenditures]
- (`X_4`) Materials: ((matcost/pimat) - (energy/pien)) [i.e., real non-energy materials]
- (`X_5`) Capital: cap [i.e., total capital stock, already in real terms]

The change in factor usage between one year and the next is defined as the change in natural logs, (for example): `dX_{it} = ln(X_{it}) - ln(X_{it-1})`

We also need the change in real output: (Q) Real output: *vship/piship*

As with factor usage, we express output change in terms of natural logs, hence: `dQ_t = ln(Q_t) - ln(Q_{t-1})`

The change in 5-factor TFP (dTFP5) between this year and last is thus defined as: `dTFP5_t = dQ_t - sum_i(alpha'_{it}dX_{it})`, `i = 1,..., 5`

Given the series of dTFP5 values, one can then "roll up" these changes to form a TFP index (TFP5), by setting the index equal to 1.0 in some initial year `t` and then growing the index forward by the following equation: `TFP5_{t+1} = exp[ ln(TFP5_t) + (dTFP5_{t+1})]`

The values of 4-factor TFP growth (dTFP4) and the corresponding TFP index (TFP4) are calculated similarly, but using total materials cost spending rather than separating it into energy and non-energy materials.

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.045
Maximum	52.76
Mean	1.006
Standard deviation	1.035

Value Ranges

Value Range

Range: [0.0450000017881393 , 52.7599983215332]

Variable Name	dtfp4
Label	4-factor TFP annual growth rate
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

There are two versions of TFP in the NBER-CES Manufacturing Database: 4-factor and 5-factor. The 5- factor version separates out energy from non-energy materials; the 4-factor uses a single materials input (which includes energy). The TFP calculation requires definitions of the cost shares, the factors, the factor changes, and the output changes. The five cost shares (`alpha_i`) vary by industry by year, defined using the variable names in the dataset:

- (`alpha_1`) Non-production workers: (pay-prodw)/vship [i.e., their pay divided by shipments]
- (`alpha_2`) Production workers: prodw/vship [i.e., their pay divided by shipments]
- (`alpha_3`) Energy: energy/vship [i.e., energy expenditure divided by shipments]
- (`alpha_4`) Materials: (matcost-energy)/vship [i.e., non-energy materials divided by shipments]
- (`alpha_5`) Capital: 1 - (sum of the above shares) [i.e., the residual]

In calculating TFP growth from one year to the next, we use the average of the two years' cost shares: `bar alpha_{it} = 0.5alpha_{it} + 0.5alpha_{it-1}` The 5 factors (`X_i`) are defined as follows, using the variable names in the dataset:

- (`X_1`) Non-production workers: (emp-prode) [i.e., the number of non-production workers]
- (`X_2`) Production workers: prodh [i.e., production worker hours, not employees]
- (`X_3`) Energy: (energy/pien) [i.e., real energy expenditures]
- (`X_4`) Materials: ((matcost/pimat) - (energy/pien)) [i.e., real non-energy materials]
- (`X_5`) Capital: cap [i.e., total capital stock, already in real terms]

The change in factor usage between one year and the next is defined as the change in natural logs, (for example): `dX_{it} = ln(X_{it}) - ln(X_{it-1})`

We also need the change in real output: (Q) Real output: *vship/piship*

As with factor usage, we express output change in terms of natural logs, hence: `dQ_t = ln(Q_t) - ln(Q_{t-1})`

The change in 5-factor TFP (dTFP5) between this year and last is thus defined as: `dTFP5_t = dQ_t - sum_i(alpha'_{it}dX_{it})`, `i = 1,..., 5`

Given the series of dTFP5 values, one can then "roll up" these changes to form a TFP index (TFP5), by setting the index equal to 1.0 in some initial year `t` and then growing the index forward by the following equation: `TFP5_{t+1} = exp[ ln(TFP5_t) + (dTFP5_{t+1})]`

The values of 4-factor TFP growth (dTFP4) and the corresponding TFP index (TFP4) are calculated similarly, but using total materials cost spending rather than separating it into energy and non-energy materials.

Summary Statistics

Valid values	23300
Invalid values	568
Minimum	-0.64
Maximum	1.387
Mean	0.003
Standard deviation	0.0684

Value Ranges

Value Range

Range: [-0.639999985694885 , 1.38699996471405]

Variable Name	tfp4
Label	4-factor TFP index 1987=1.000
Concept	
Type	numeric
Files	F1csv F1dta F1sas F1xls

Full Description

There are two versions of TFP in the NBER-CES Manufacturing Database: 4-factor and 5-factor. The 5- factor version separates out energy from non-energy materials; the 4-factor uses a single materials input (which includes energy). The TFP calculation requires definitions of the cost shares, the factors, the factor changes, and the output changes. The five cost shares (`alpha_i`) vary by industry by year, defined using the variable names in the dataset:

- (`alpha_1`) Non-production workers: (pay-prodw)/vship [i.e., their pay divided by shipments]
- (`alpha_2`) Production workers: prodw/vship [i.e., their pay divided by shipments]
- (`alpha_3`) Energy: energy/vship [i.e., energy expenditure divided by shipments]
- (`alpha_4`) Materials: (matcost-energy)/vship [i.e., non-energy materials divided by shipments]
- (`alpha_5`) Capital: 1 - (sum of the above shares) [i.e., the residual]

In calculating TFP growth from one year to the next, we use the average of the two years' cost shares: `bar alpha_{it} = 0.5alpha_{it} + 0.5alpha_{it-1}` The 5 factors (`X_i`) are defined as follows, using the variable names in the dataset:

- (`X_1`) Non-production workers: (emp-prode) [i.e., the number of non-production workers]
- (`X_2`) Production workers: prodh [i.e., production worker hours, not employees]
- (`X_3`) Energy: (energy/pien) [i.e., real energy expenditures]
- (`X_4`) Materials: ((matcost/pimat) - (energy/pien)) [i.e., real non-energy materials]
- (`X_5`) Capital: cap [i.e., total capital stock, already in real terms]

The change in factor usage between one year and the next is defined as the change in natural logs, (for example): `dX_{it} = ln(X_{it}) - ln(X_{it-1})`

We also need the change in real output: (Q) Real output: *vship/piship*

As with factor usage, we express output change in terms of natural logs, hence: `dQ_t = ln(Q_t) - ln(Q_{t-1})`

The change in 5-factor TFP (dTFP5) between this year and last is thus defined as: `dTFP5_t = dQ_t - sum_i(alpha'_{it}dX_{it})`, `i = 1,..., 5`

Given the series of dTFP5 values, one can then "roll up" these changes to form a TFP index (TFP5), by setting the index equal to 1.0 in some initial year `t` and then growing the index forward by the following equation: `TFP5_{t+1} = exp[ ln(TFP5_t) + (dTFP5_{t+1})]`

The values of 4-factor TFP growth (dTFP4) and the corresponding TFP index (TFP4) are calculated similarly, but using total materials cost spending rather than separating it into energy and non-energy materials.

Summary Statistics

Valid values	23759
Invalid values	109
Minimum	0.044
Maximum	52.736
Mean	1.006
Standard deviation	1.034

Value Ranges

Value Range

Range: [0.0439999997615814 , 52.7360000610352]