BDS CODEBOOK AND GLOSSARY (2023 Release)

BDS VARIABLE DEFINITIONS AND CODES:

year

Values: 1978-2023

Description: Record year

sector

Values: 11, 21, 22, 23, 31-33, 42, 44-45, 48-49, 51, 52, 53, 54, 55, 56, 61, 62, 71, 72, 81

Description: 2-digit NAICS 2017 Codes:

11 - Agriculture, Forestry, Fishing and Hunting
21 - Mining, Quarrying, and Oil and Gas Extraction
22 - Utilities
23 - Construction
31-33 - Manufacturing
42 - Wholesale Trade
44-45 - Retail Trade
48-49 - Transportation and Warehousing
51 - Information
52 - Finance and Insurance
53 - Real Estate and Rental and Leasing
54 - Professional, Scientific, and Technical Services
55 - Management of Companies and Enterprises
56 - Administrative and Support and Waste Management and Remediation Services
61 - Educational Services
62 - Health Care and Social Assistance
71 - Arts, Entertainment, and Recreation
72 - Accommodation and Food Services
81 - Other Services (except Public Administration)

vcnaics3

Values:
Description: Vintage consistent 3-digit NAICS 2017 codes

vcnaics4

Values:
Description: Vintage consistent 4-digit NAICS 2017 codes

esize

Values: a) 1 to 4, b) 5 to 9, c) 10 to 19, d) 20 to 99, e) 100 to 499, f) 500 to 999, g) 1000+

Description: Establishment size. A categorical variable defined as the average of the establishment's year t-1 and year t employment.

esizecoarse

Values: a) 1 to 19, b) 20 to 499, c) 500+

Description: Establishment size coarse. Coarse version of establishment size variable defined above. Includes 3 broader size classes.

iesize

Values: a) 1 to 4, b) 5 to 9, c) 10 to 19, d) 20 to 99, e) 100 to 499, f) 500 to 999, g) 1000+

Description: Initial establishment size. A categorical variable defined for any given consecutive two-year period as the size at year t-1 except in cases when year t-1 employment is equal to zero in which case initial size is year t employment.

iesizecoarse

Values: a) 1 to 19, b) 20 to 499, c) 500+

Description: Initial establishment size coarse. Coarse version of initial establishment size variable defined above. Includes 3 broader initial establishment size classes.

fsize

Values: a) 1 to 4, b) 5 to 9, c) 10 to 19, d) 20 to 99, e) 100 to 499, f) 500 to 999, g) 1000 to 2499, h) 2500 to 4999, i) 5000 to 9999, j) 10000+

Description: Firm size. A categorical variable defined as the average of the firm's year t-1 and year t employment.

fsizecoarse

Values: a) 1 to 19, b) 20 to 499, c) 500+

Description: Firm size coarse. Coarse version of firm size variable defined above. Includes 3 broader firm size classes.

ifsize

Values: a) 1 to 4, b) 5 to 9, c) 10 to 19, d) 20 to 99, e) 100 to 499, f) 500 to 999, g) 1000 to 2499, h) 2500 to 4999, i) 5000 to 9999, j) 10000+

Description: Initial firm size. A categorical variable defined for any given consecutive two-year period as the size at year t-1 except in cases when year t-1 employment is equal to zero in which case initial size is year t employment.

ifsizecoarse

Values: a) 1 to 19, b) 20 to 499, c) 500+

Description: Initial firm size coarse. Coarse version of initial firm size variable defined above. Includes 3 broader initial firm size classes.

eage

Values: a) 0, b) 1, c) 2, d) 3, e) 4, f) 5, g) 6 to 10, h) 11 to 15, i) 16 to 20, j) 21 to 25, k) 26+, l) Left Censored

Description: Establishment age. This is an employment-based measure of establishment age. An estab born before 1977 is of unknown age and is assigned an age of "Left Censored". One born during or after 1977 is assigned an age of zero in the first year it reports positive employment, and it ages each year thereafter regardless of employment.

eagecoarse

Values: a) 0, b) 1 to 5, c) 6 to 10, d) 11+, e) Left Censored

Description: Establishment age coarse. This is the coarse version of the establishment age variable defined above. Includes 5 broader establishment age classes.

fage

Values: a) 0, b) 1, c) 2, d) 3, e) 4, f) 5, g) 6 to 10, h) 11 to 15, i) 16 to 20, j) 21 to 25, k) 26+, l) Left Censored

Description: This is an employment-based measure of firm age. A firm born before 1977 is of unknown age and is assigned a firm age of "Left Censored". A firm born during or after 1977 is assigned a firm age as follows:
1. One or more pre-existing establishments with prior employment, the firm will be born with a non-zero age based on the earliest year any establishment reported employment.
2. One or more pre-existing establishments with no prior employment in any establishment, the firm is born with no firm age. The firm will be age zero in the first year any establishment reports employment.
3. One or more newly founded establishments with employment, the firm will be born with age of zero.
4. One or more newly founded establishments with no employment in any establishment, the firm will be born with no firm age. The firm will be age zero in the first year any establishment reports employment.

fagecoarse

Values: a) 0, b) 1 to 5, c) 6 to 10, d) 11+, e) Left Censored

Description: Firm age coarse. This is a coarse version of the firm age variable described above. Includes 5 broader firm age classes.

firms

Values:

Description: A simple count of the number of firms in the cell. For state level tables, a firm with establishments in multiple states be counted multiple times, once in each state, irrespective of the portion of the firm residing in that state.

estabs

Values:

Description: A simple count of the number of establishments in the cell.

emp

Values:

Description: Paid employment consists of full and part-time employees, including salaried officers and executives of corporations, who are on the payroll in the pay period including March 12. Included are employees on paid sick leave, holidays, and vacations.

denom

Values:

Description: Davis-Haltiwanger-Schuh (DHS) denominator. For time t, denom is the average of employment for times t and t-1. This variable attempts to prevent transitory shocks from creating a bias to the relationship between net growth from t-1 to t and size.

estabs_entry

Values:

Description: A count of establishments born within the cell during the last 12 months.

estabs_entry_rate

Values:

Description: 100 * (estabs_entry at time t divided by the average of estabs at t and t-1).

estabs_exit
Values:
Description: A count of establishments exiting from within the cell during the last 12 months.

estabs_exit_rate
Values:
Description: 100 * (estabs_exit at time t divided by the average of estabs at t and t-1).

job_creation
Values:
Description: Count of all employment gains within the cell from expanding and opening establishments between the week of March 12 of the prior year to the current year.

job_creation_births
Values:
Description: Count of employment gains within the cell from establishment openings (births) between the week of March 12 of the prior year to the current year.

job_creation_continuers
Values:
Description: Count of employment gains within the cell from continuing establishments that expanded between the week of March 12 of the prior year to the current year.

job_creation_rate_births
Values:
Description: 100 * (job_creation_births / denom)

job_creation_rate
Values:
Description: 100 * (job_creation / denom)

job_destruction
Values:
Description: Count of all employment losses within the cell from contracting and closing establishments between the week of March 12 of the prior year to the current year.

job_destruction_deaths
Values:
Description: Count of all employment losses within the cell from establishment closings (deaths) between the week of March 12 of the prior year to the current year.

job_destruction_continuers
Values:
Description: Count of employment losses within the cell from continuing establishments that contracted between the week of March 12 of the prior year to the current year.

job_destruction_rate_deaths
Values:
Description: 100 * (job_destruction_deaths / denom)

job_destruction_rate
Values:
Description: 100 * (job_destruction / denom)

net_job_creation
Values:
Description: job_creation - job_destruction

net_job_creation_rate
Values:
Description: job_creation_rate - job_destruction_rate

reallocation_rate
Values:
Description: job_creation_rate + job_destruction_rate - abs(net_job_creation_rate). This is often referred to as an "excess" reallocation rate since it measures the rate of job reallocation over and above that needed to accommodate the net job creation in the cell. A gross reallocation measure is simply the sum of the job_creation_rate and the job_destruction_rate.

firmdeath_firms
Values:
Description: Count of firms that have exited in their entirety during the period. All establishments owned by the firm must exit to be considered a firm death. This definition of firm death is narrow and strictly applied, so that a firm with 100 establishments would not qualify as a firm death if 99 exited while 1 continued under different ownership. Note firm legal entities that cease to exist because of merger and acquisition activity are not classified as firm deaths in the BDS data.

firmdeath_estabs
Values:
Description: Count of establishments associated with firm deaths.

firmdeath_emp
Values:
Description: Count of employment associated with firm deaths.

metro
Values: M, N, SW

Description: Establishments located in Metropolitan or Micropolitan Statistical Areas are in cells designated with "M", while those in non-metropolitan areas are aggregated into cells designated with "N". Cells designated as "SW" (signifying "statewide") are assigned to establishments without a fixed location such as employee leasing establishments.

msa
Values:
Description: Numeric code for metropolitan and micropolitan statistical areas, as established by Office of Management and Budget (OMB). The definitions from July 2023 are used: https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/delineation-files.html

cty
Values:
Description: Three-character numeric code for county within a given state. Users can refer to the CBP county reference files for listings of valid county codes (https://www.census.gov/programs-surveys/cbp/technical-documentation/reference/state-county-geography-reference.html). The BDS also includes statewide (999) county codes. Statewide county codes are assigned to establishments without a fixed location such as employee leasing establishments. County codes are vintage consistent and currently reflect 2023 geography.

st

Values: 01, 02, 04, 05, 06, 08, 09, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 44, 45, 46, 47, 48, 49, 50, 51, 53, 54, 55, 56

Description: Two-character numeric code corresponding to the Federal Information Processing Series (FIPS) code for each state. These codes are vintage consistent and currently reflect 2023 geography. The state codes take on the values listed in the table below:

01-AL	15-HI	26-MI	37-NC	49-UT
02-AK	16-ID	27-MN	38-ND	50-VT
04-AZ	17-IL	28-MS	39-OH	51-VA
05-AR	18-IN	29-MO	40-OK	53-WA
06-CA	19-IA	30-MT	41-OR	54-WV
08-CO	20-KS	31-NE	42-PA	55-WI
09-CT	21-KY	32-NV	44-RI	56-WY
10-DE	22-LA	33-NH	45-SC	72-PR
11-DC	23-ME	34-NJ	46-SD	
12-FL	24-MD	35-NM	47-TN	
13-GA	25-MA	36-NY	48-TX	

BDS GLOSSARY:

Business Dynamics Statistics (BDS) – A public dataset based on the LBD that describes United States business dynamics across a wide range of measures. Disclosure analysis is performed prior to release to the public to protect the confidentiality of the underlying LBD data.

Business Register (BR) – A comprehensive database of all U.S. business establishments developed and maintained by the U.S. Census Bureau, with data beginning in 1975 and continuing to the present. This is restricted data and is the source data for the Longitudinal Business Database (LBD).

Censoring – A statistical term indicating that a value cannot be known with certainty. Within the BDS, all firms/establishments born prior to 1976 have an unknown birth year and are therefore of an unknown age and are grouped into the age category "Left Censored".

Data Quality Suppression – Data quality suppressions are made when a cell is determined to be unreliable due to its time series characteristics. Cells suppressed due to data quality concerns will appear as "S".

Disclosure Suppression – Disclosure suppressions are made when a cell has too few firms. Cells suppressed due to containing too few firms will appear as "D".

Structurally Missing Flag – Structurally missing flags are applied to cells that are "structurally zero" or "structurally missing." These are cells in the firm and establishment age datasets where activity is not possible given the nature or structure of the BDS data. These cells will appear as "X". An example of a "structurally zero" cell is the variable 'estab_exit' for firms age '0'. These cells will always be '0', given the nature of data for firms age '0'. An example of a "structurally missing" cell is any variable for firms age '5' in the years 1978 to 1981. These cells will always be missing for the years 1978 to 1981 because the source data for the BDS, the LBD, begins in 1976.

Rate Not Available – Rates that cannot be calculated due to a denominator of '0' will appear as "N".

North American Industry Classification System (NAICS) – The North American Industry Classification System (NAICS) is the standard used by Federal statistical agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy. This system is used by the United States, Canada, and Mexico. Note that the 2023 BDS release is based on the 2017 Vintage NAICS.