Appendix D: Clearance Request Memo

Request for Clearance of Research Output

Center for Economic Studies and Research Data Centers

••

Project # :

Submitted by :

For CES reviewer to complete
Cleared for release :

Cleared by :

••

1. General information

a. Name of this request's subdirectory under the project's main clearance directory:

http://localhost:8080/ced2ar-web/codebooks/synlbdv2

b. Please provide a general description of the outputs you wish to clear:

In most countries, national statistical agencies do not release establishment-level business microdata, because doing so represents too large a risk to establishments' confidentiality. One approach with the potential for overcoming these risks is to release synthetic data; that is, the released establishment data are simulated from statistical models designed to mimic the distributions of the underlying real microdata. The Synthetic Longitudinal Business Database (SynLBD) is the synthetic data version of the Longitudinal Business Database (LBD), an annual economic census of establishments in the United States comprising more than 20 million records dating back to 1976. More information is available at https://www.census.gov/ces/dataproducts/synlbd/index.html. In this codebook, variables are noted as "blanked" if they are available on the confidential version but have been removed from the synthetic version; "synthetic" if the confidential values have been synthesized and released on the synthetic version.

c. Please state how the outputs are part of the research project as approved (You may summarize or copy descriptions from your proposal, with page references.)

Where should this come from?

2. Research

a. Descriptions of Research

Describe your Research sample(s) or "cuts" of data used in research output. For each sample, please describe your selection criteria and how the research sample differs from the samples underlying survey publications or other samples you have used. Take as much space as you need for each; add samples as needed.

Sampling from posterior predictive distribution

b. Relationship Between Samples

Describe how your samples relate to each other (e.g., if you have two samples, is one a subsample of another?) In the cases of samples and subsamples, there is an implicit third sample, the difference between the two. Please describe this sample above. We probably will need to examine any implicit samples as well.

c. Relationship to other Publications

Describe how your samples may relate to similar samples from other projects or from survey publications. (e.g., how your sample of an industry in the LRD differs from the Census of Manufactures or Annual Survey of Manufactures files in the LRD).

Related Materials
  I. https://www.census.gov/ces/pdf/SynLBD_Codebook.pdf

Related Studies
  I. Test study

Related Publications
  I. Kinney, Satkartar K., Jerome P. Reiter, Arnold P. Reznek, Javier Miranda, Ron S. Jarmin and John M. Abowd. 2011. CES WP-11-04 In most countries, national statistical agencies do not release establishment-level business microdata, because doing so represents too large a risk to establishments' confidentiality. One approach with the potential for overcoming these risks is to release synthetic data; that is, the released establishment data are simulated from statistical models designed to mimic the distributions of the underlying real microdata. In this article, we describe an application of this strategy to create a public use file for the Longitudinal Business Database, an annual economic census of establishments in the United States comprising more than 20 million records dating back to 1976. The U.S. Bureau of the Census and the Internal Revenue Service recently approved the release of these synthetic microdata for public use, making the synthetic Longitudinal Business Database the first-ever business microdata set publicly released in the United States. We describe how we created the synthetic data, evaluated analytical validity, and assessed disclosure risk.

3. Output Files

For each research output file to be removed, please enter the following information:

a. File name (e.g., output.rtf)

synlbd1997c.dta - Incomplete location provided: synlbd1997c.dta (Stata)

b. Description of file

    synlbd1997c.dta:
    synlbd1997c.dta all 28 Stata

c. Program that produced the file (e.g., output.sas or model.do)

    synlbdv2 was derived from lbdv2
    synlbdv2 was attributed to synlbdproc

4. Variable Definitions

Name : lbdnum

Label : (synthetic) LBD Number

Type : character

Description :

Longitudinal establishment identifier. Can be used to track establishment units over time.

Name : emp

Label : (synthetic) March 12 Employment

Type : numeric

Description :

Paid employment consists of full and part-time employees, including salaried officers and executives of corporations, who were on the payroll in the pay period including March 12. Included are employees on sick leave, holidays, and vacations; not included are proprietors and partners of unincorporated businesses. Employment refers to paid employment at the establishment where business is conducted. Note, this will correspond to firm employment only for single-unit establishment firms. Reported to the Internal Revenue Service (IRS) on Form 941. In some cases this value is imputed due to missing or invalid data.

Name : pay

Label : (synthetic) Reported Annual Payroll (in $1,000)

Type : numeric

Description :

Total annual payroll includes all forms of compensation, such as salaries, wages, commissions, bonuses, vacation allowances, sick-leave pay, and the value of payments in kind (e.g., free meals and lodgings) paid during the year to all employees. Sum of quarterly IRS Form 941 payroll for the year. Missing or invalid 941 data is replaced with imputed values.

Name :   sic3

Label :   (observed, computed) SIC3 code

Type :   character

Description :

   Three digit Standard Industrial Classification code.

Notes :

   #1: Test note

Name :   mu

Label :   (synthetic) Single-Multi Identifier

Type :   numeric

Description :

   Indicator for whether the establishment belongs to a firm composed of two or more establishments. A value of 1
   indicates the establishment is a member of a firm composed of two or more establishments. A value of 0
   indicates the establishment is the only member ofthe firm.

Name :   firstyear

Label :   (synthetic) First Year Establishment is Observed

Type :   numeric

Description :

   Indicator for the first year the establishment is observed in the data (birth year). This variable is left censored at
   1976. In conjunction with LASTYEAR, allows users to quickly determine the tenure of an establishment from any
   point in the data series.

Name :   lastyear

Label :   (synthetic) Last Year Establishment is Observed

Type :   numeric

Description :

   Indicator for the last year the establishment is observed (death year). This variable is right censored at the last
   year of the data. In conjunction with FIRSTYEAR, allows users to quickly determine the tenure of an
   establishment from any point in the data series.

Name :   act

Label :   (blanked) Activity Code

Type :   character

Description :

   Variable not present on Synthetic LBD, only available on confidential LBD

Name :   bestsic

Label :   (blanked) Best SIC code

Type :   character

Description :

   Variable not present on Synthetic LBD, only available on confidential LBD

Name :   bestnaics

Label :   (blanked) Best NAICS code

Type :   character

Description :

   Variable not present on Synthetic LBD, only available on confidential LBD

Name :   cbp

Label :   (blanked) Indicator: used in County Business Patterns

Type :   numeric

Description :

   Variable not present on Synthetic LBD, only available on confidential LBD. Indicates that an observation was used in the tabulation of the County Business Patterns

Name :   cfn

Label :   (blanked) Census File Number

Type :   character

Description :

   Variable not present on Synthetic LBD, only available on confidential LBD. Links to other Economic microdata

Name : county

Label : (blanked) County FIPS codes

Type : character

Description :

 Variable not present on Synthetic LBD, only available on confidential LBD. County FIPS code. It is not possible to compute statistics at the state or county level.

Name : firstflag

Label : (blanked) First Link Flag

Type : character

Description :

 Indicator for first link. Only available on confidential LBD.

Name : flaga

Label : (blanked) Type of Link Flag

Type : character

Description :

 Identifies the type of link. Only available on confidential LBD.

Name : flagb

Label : (blanked) Birth-Death-Continuer Link Flag

Type : character

Description :

 Identifies if the link is for birth/death/continuing establishment. Only available on confidential LBD.

Name : lastflag

Label : (blanked) Last Link Flag

Type : character

Description :

 Only available on confidential LBD.

Name :   lfo

Label :   (blanked) Legal Form of Organization

Type :   character

Description :

   Identifies the legal form of the organization. Only available on confidential LBD.

Name :   lfo1

Label :   (blanked) LFO1

Type :   character

Description :

   Processing variable. Only available on confidential LBD.

Name :   mfsic1

Label :   (blanked) Most Frequent SIC 1

Type :   character

Description :

   Variable not present on Synthetic LBD, only available on confidential LBD.

Name :   pdiv

Label :   (blanked) Processing (Economic) Division Code

Type :   character

Description :

   Variable not present on Synthetic LBD, only available on confidential LBD.

Name :   rapf

Label :   (blanked) Reported Annual Payroll Flag

Type :   character

Description :

   Only available on confidential LBD.

Name :   recnum

Label :   (blanked) SSEL Record Number

Type :   numeric

Description :

    Links to BR. Only available on confidential LBD, Variable not present on Synthetic LBD.

Name :   sic

Label :   (blanked) Standard Industrial Classification Code

Type :   character

Description :

    Variable not present on Synthetic LBD, only available on confidential LBD. Detailed SIC code

Name :   sone

Label :   processing variable (to be dropped in future version)

Type :   character

Name :   state

Label :   (blanked) State FIPS codes

Type :   character

Description :

    Variable not present on Synthetic LBD, only available on confidential LBD. It is not possible to compute statistics at the state or county level.

Name :   toc

Label :   (blanked) Type of Operation Code

Type :   character

Description :

    Variable not present on Synthetic LBD, only available on confidential LBD.

Name :   yr

Label :   (computed) Year

Type :   numeric

Description :

   Implicit in file name, was added on SynLBD 2.0.2