Using partially synthetic microdata to protect sensitive cells in business statistics

Javier Miranda[2] Lars Vilhuber[1]

[1]Labor Dynamics Institute, ILR, Cornell University, United States

[2]Center for Economic Studies, U.S. Census Bureau, United States

January 2016
NCRN Virtual Seminar

Funding

- Vilhuber's work is partially funded by NSF Grant #1042181 and #0941226.
- This work is part of the Census Bureau's LBD Initiative.

Disclaimer

"

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review by the Census Bureau than its official publications. This report is released to inform interested parties and to encourage discussion. Any findings, conclusions or opinions are those of the authors. They do not necessarily reflect those of the Center for Economic Studies, the U.S. Census Bureau, or the National Science Foundation.

"

Business Dynamics

”The U.S. economy is comprised of over 6 million establishments with paid employees. The population of these businesses is constantly churning – some businesses grow, others decline and yet others close. New businesses are constantly replenishing this pool.”[*]

Statistics at great detail on

- job creation and destruction
- establishment births and deaths
- firm startups and shutdowns

by establishment and firm characteristics (age, size, location)

Business Dynamic Statistics (BDS)

www.census.gov/ces/dataproducts/bds/

Firm and Establishment Characteristics

- Sector
- Firm Size
- Firm Age
- Initial Firm Size
- Geography (State, Metro/Non-metro, MSA)
- Cross-tabulations by up to three of these characteristics

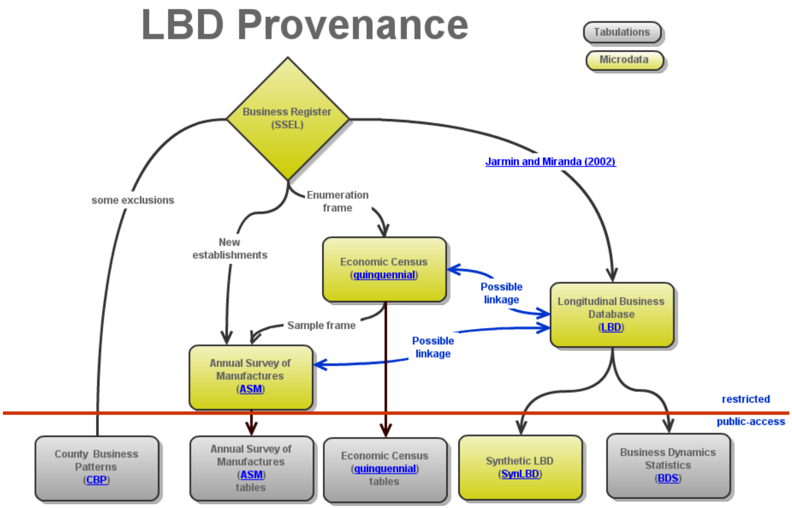
Lots of detail

Currently 62 very detailed tables, latest release September 2015

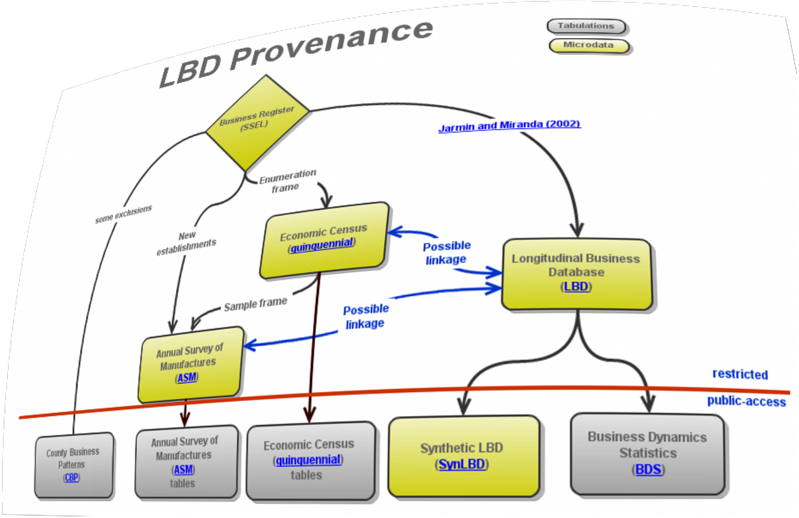
Business Microdata at the Census Bureau

LBD-BDS complex

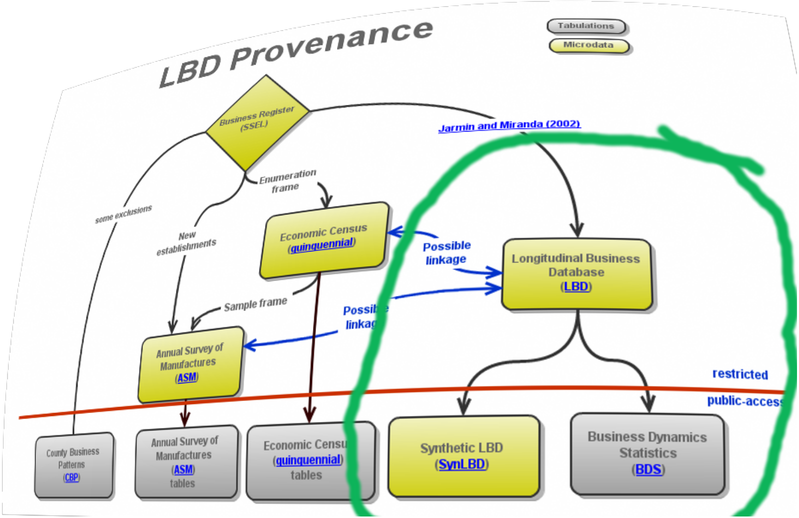
Business Microdata at the Census Bureau



Business Microdata at the Census Bureau



Business Microdata at the Census Bureau



Disclosure avoidance in the BDS

P-percent rule with secondary suppressions

- Cells where the top 2 firms account for more than P percent of the total value of the cell are flagged for suppression

Disclosure avoidance in the BDS

P-percent rule with secondary suppressions

- Cells where the top 2 firms account for more than P percent of the total value of the cell are flagged for suppression
- P value is not disclosed

Disclosure avoidance in the BDS

P-percent rule with secondary suppressions

- Cells where the top 2 firms account for more than *P* percent of the total value of the cell are flagged for suppression
- *P* value is not disclosed
- Trivially: cells with fewer than 3 firms represented are always suppressed

Disclosure avoidance in the BDS

P-percent rule with secondary suppressions

- Cells where the top 2 firms account for more than P percent of the total value of the cell are flagged for suppression
- P value is not disclosed
- Trivially: cells with fewer than 3 firms represented are always suppressed
- Secondary suppressions: “minimize the amount of information loss in a given table row or column”.

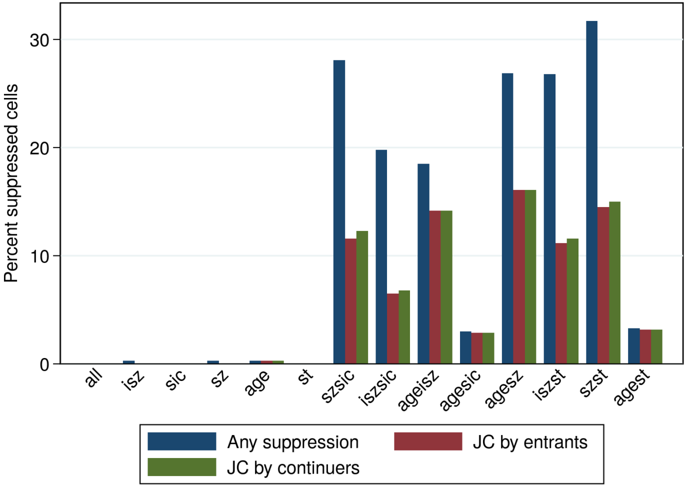
Extent of suppression

Table: Suppressions in establishment-level BDS

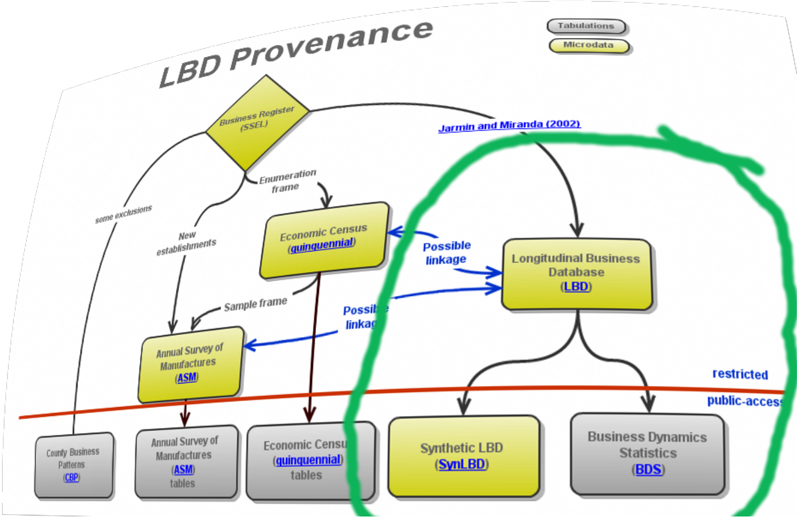
Type	Level	Number of cells	Suppressions (%) Any	Job creation by entrants	Job creation by continuers
Age	e	337	0.3	0.3	0.3
Age-Initial Size	e	3033	18.5	14.2	14.2
Age-SIC	e	3033	3	2.9	2.9
Age-State	e	19023	3.3	3.2	3.2
Age-Size	e	3033	26.9	16.1	16.1
All	e	36	0	0	0
Initial Size	e	324	0.3	0	0
Initial Size-SIC	e	2916	19.8	6.5	6.8
Initial Size-State	e	18357	26.8	11.2	11.6
SIC	e	324	0	0	0
State	e	1836	0	0	0
Size	e	324	0.3	0	0
Size-SIC	e	2915	28.1	11.6	12.3
Size-State	e	18358	31.7	14.5	15

Note: Cells are year *x* categories, where the number of categories varies by published table.

Extent of suppression



Business Microdata at the Census Bureau



Purpose of SynLBD

The SynLBD is

- synthetic establishment (and soon firm) microdata

Purpose of SynLBD

The SynLBD is

- synthetic establishment (and soon firm) microdata
- derived from confidential Longitudinal Business Database (LBD, [5])

Purpose of SynLBD

The SynLBD is

- synthetic establishment (and soon firm) microdata
- derived from confidential Longitudinal Business Database (LBD, [5])
- designed to facilitate researcher access to establishment microdata (LBD) (see http://vrdc.cornell.edu/sds)

Purpose of SynLBD

The SynLBD is

- synthetic establishment (and soon firm) microdata
- derived from confidential Longitudinal Business Database (LBD, [5])
- designed to facilitate researcher access to establishment microdata (LBD) (see http://vrdc.cornell.edu/sds)
- while preserving the confidentiality of establishment/business data.

Purpose of SynLBD

The SynLBD is

- synthetic establishment (and soon firm) microdata
- derived from confidential Longitudinal Business Database (LBD, [5])
- designed to facilitate researcher access to establishment microdata (LBD) (see http://vrdc.cornell.edu/sds)
- while preserving the confidentiality of establishment/business data.
- part of a larger strategy by the Census Bureau to provide *better statistics on business dynamics* CNSTAT [9]

Contents of (Syn)LBD

Data elements

- longitudinal establishment identifiers (created using probabilistic matching [5])
- information on birth, death
- employment and payroll over time
- location
- industry
- firm affiliation of employer establishments

Contents of (Syn)LBD

Data elements

- longitudinal establishment identifiers (created using probabilistic matching [5]) Masked
- information on birth, death Synthesized
- employment and payroll over time Synthesized
- location Suppressed
- industry Released
- firm affiliation of employer establishments $\rightarrow$ next version

Contents of (Syn)LBD

Data elements

- longitudinal establishment identifiers (created using probabilistic matching [5]) Masked
- information on birth, death Synthesized
- employment and payroll over time Synthesized
- location Suppressed
- industry Released
- firm affiliation of employer establishments → next version

Complete description

Kinney et al [7]

[more]

Putting two and two together...

V2.0 of SynLBD released by Census Bureau's Disclosure Review Board in 2011

Putting two and two together...

- V2.0 of SynLBD released by Census Bureau's Disclosure Review Board in 2011
- Let's combine public-use data to fill in suppressions

Goal is two-fold

Retro-active utility

A mechanism that can fill in existing suppressions.

Improving disclosure avoidance going forward

Evaluate future disclosure avoidance mechanisms:

- Suppression
- This proposition
- Noise infusion (not here)

Analytic validity

Figures

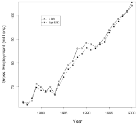


Figure 1: Gross Employment Level by Year, LBD vs Synthetic

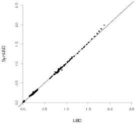


Figure 3: Share of Employment by Industry Sector and Year, 1976-2000

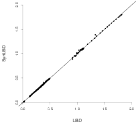


Figure 2: Share of Establishments by Industry Sector and Year, 1976-2000.

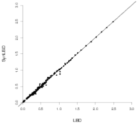


Figure 4: Share of Payroll by Industry Sector and Year, 1976-2000

37

Analytic validity

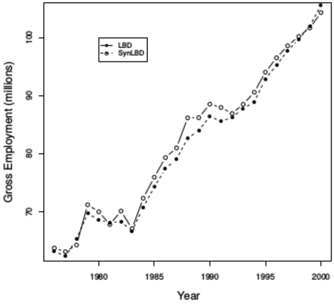


Figure 1: Gross Employment Level by Year, LBD vs Synthetic

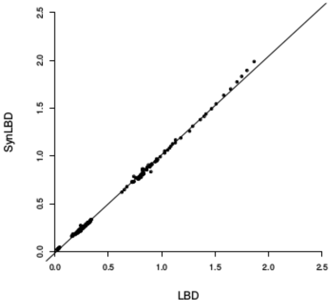


Figure 3: Share of Employment by Industry Sector and Year, 1976-2000

Analytic validity

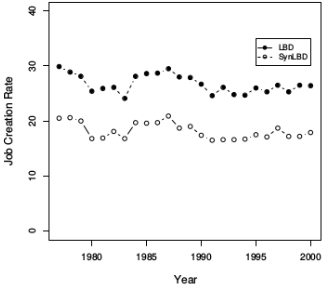


Figure 8: Job Creation Rate by Year, LBD vs Synthetic

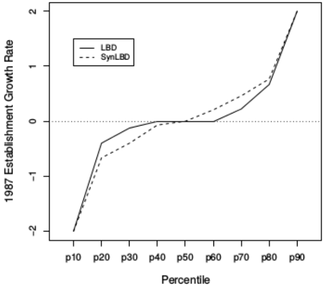


Figure 9: Distribution of Job Creation Rates, LBD vs Synthetic

Notation

Base variable

Establishment employment e_{jt}.

Example

$$birth_{jt} = \begin{cases} 1 & \text{if } e_{jt} > 0 \text{ and } e_{jt-s} = 0 \ \forall s \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$jcbirth_{jt} = \begin{cases} e_{jt} - e_{jt-1} & \text{if } e_{jt} > 0 \text{ and } e_{jt-s} = 0 \ \forall s \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Notation

Synthetic values

Synthesized version of variable x_{jt} is denoted $\tilde{x}_j t$.

Cells

- Collections of characteristics $k_t(j)$ (industry, geography, establishment or firm age and size)
- $j \in K_t'$ describes the set of firms at time t such that $k_t(j) = k'$.

Notation

Aggregations

Generically in capital letters:

$$E_{\cdot t} = \sum_{j=1}^{J} e_{jt}, \tag{3}$$

Aggregations across establishments having characteristics k' at time t

$$X_{k't} = \sum_{j \in K_t'} x_{jt} \tag{4}$$

Suppression rules

Suppression rules

for (aggregate) variable X are captured by I_t^X, such that the releasable variable $X^{(0)}$ under the current regime can be described by

$$X_{k't}^{(0)} = \begin{cases} X_{k't} & \text{if } I_{kt}^X = 1 \\ \text{missing} & \text{otherwise} \end{cases} \quad (5)$$

Algorithm 1

We can now express the simple “drop-in” algorithm, leading to the released variable $X^{(i)}$, as:

BDS$^{(in)}$

---

if $I_t^X = 0$ **then**

$X_{k't}^{(i)} = \tilde{X}_{k't}$

else

$X_{k't}^{(i)} = X_{k't}$

end if

Weighted Algorithm 1

Time-consistency

Because no time-consistency is imposed, this method can lead to seam biases or higher intertemporal variance

Weighted Algorithm 1

Time-consistency

Because no time-consistency is imposed, this method can lead to seam biases or higher intertemporal variance

Smoothing the time series

In periods that follow a period with suppressions ($I_t^X = 1$), we average synthetic tabulations with non-suppressed tabulations, for up to n periods.

Weighted Algorithm 1

BDS$^{(i)}$

Algorithm 1: Weighted Drop-in

$s^* = min_{s\in[0,n]}$ s.t. $I^X_{t-s} = 0$
if $n > 0$ and $\exists s^*$ **then**
$$X^{(i)}_{k't} = \frac{s^*}{n} X_{k't} + \left(1 - \frac{s^*}{n}\right) \tilde{X}_{k't}$$
else if $n = 0$ and $I^X_t = 0$ **then**
$$X^{(i)}_{k't} = \tilde{X}_{k't}$$
else
$$X^{(i)}_{k't} = X_{k't}$$
end if

Algorithm 2

Similar idea, at microdata level

Replace sensitive establishments with synthetic establishments.

Smooth the replacement

- per-establishment weight $w_{js} \in [0, 1]$, applied to the observed data, that increases from 0 in t to 1 in $t + n$,
- a per-establishment weight $\tilde{w}_{js}$, applied to the synthetic data, that decreases from 1 in t to 0 in $t + n$,
- thus “blending in” the real establishments, and “blending out” the synthetic establishments.

Algorithm 2: notation

$J^{-}_{k't}$ establishments excluded from tabulations at time t

- We construct $J^{-}_{k't}$ by first adding establishment identifiers that meet the suppression conditions I^{X}_{kt} at time t.
- Then add those same establishments to "future" I^{X}_{ks}, for $s \in [t+1, t+n]$ if $n > 0$.
- At any point in time t, the set $J^{-}_{k't}$ contains establishments that met suppression conditions now and in the *past*, i.e., in $[t-n, t]$.

Algorithm 2: notation

$J^-_{k't}$ establishments excluded from tabulations at time t

- We construct $J^-_{k't}$ by first adding establishment identifiers that meet the suppression conditions I^X_{kt} at time t.
- Then add those same establishments to "future" I^X_{ks}, for $s \in [t+1, t+n]$ if $n > 0$.
- At any point in time t, the set $J^-_{k't}$ contains establishments that met suppression conditions now and in the *past*, i.e., in $[t-n, t]$.

$J^+_{k't}$ synthetic establishments

added to tabulations as replacements

Algorithm 2

BDS$^{(ii)}$

Algorithm 2: Forward-longitudinal

Compute: $X_{k't} = \sum_{j \in K'_t} x_{jt}$

Compute: I_t^X

if $I_t^X = 0$ **then**

// Suppression condition met for cell k'

Assign all $j \in K'_t$ to $J^-_{k's}$ for $t \leq s \leq t+n$

Assign all $j \in \tilde{K}'_t$ to $J^+_{k't}$ for $t \leq s \leq t+n$

end if

Compute:

$$X^{(iiw)}_{k't} = \sum_{j \in \{K'_t \cap J^+_{k't}\}} \tilde{w}_{jt}\tilde{x}_{jt} + \sum_{j \in K'_t \wedge j \in J^-_{k't}} w_{jt}x_{jt} + \sum_{j \in K'_t \wedge j \notin J^-_{k't}} x_{jt}$$

Algorithm 2

BDS$^{(ii)}$

Algorithm 2: Forward-longitudinal

Compute: $X_{k't} = \sum_{j \in K'_t} x_{jt}$
Compute: I^X_t
if $I^X_t = 0$ **then**
  // Suppression condition met for cell k'
  Assign all $j \in K'_t$ to $J^-_{k's}$ for $t \leq s \leq t+n$
  Assign all $j \in \tilde{K}'_t$ to $J^+_{k't}$ for $t \leq s \leq t+n$
end if
Compute:

$$X^{(iiw)}_{k't} = \sum_{j \in \{K'_t \cap J^+_{k't}\}} \tilde{w}_{jt}\tilde{x}_{jt} + \sum_{j \in K'_t \wedge j \in J^-_{k't}} w_{jt}x_{jt} + \sum_{j \in K'_t \wedge j \notin J^-_{k't}} x_{jt}$$

Algorithm 2

BDS$^{(ii)}$

Algorithm 2: Forward-longitudinal

Compute: $X_{k't} = \sum_{j \in K'_t} x_{jt}$

Compute: I_t^X

if $I_t^X = 0$ **then**

// Suppression condition met for cell k'

Assign all $j \in K'_t$ to $J^-_{k's}$ for $t \leq s \leq t+n$

Assign all $j \in \tilde{K}'_t$ to $J^+_{k't}$ for $t \leq s \leq t+n$

end if

Compute:

$$X^{(iiw)}_{k't} = \sum_{j \in \{K'_t \cap J^+_{k't}\}} \tilde{w}_{jt}\tilde{x}_{jt} + \sum_{j \in K'_t \wedge j \in J^-_{k't}} w_{jt}x_{jt} + \sum_{j \in K'_t \wedge j \notin J^-_{k't}} x_{jt}$$

Algorithm 2

BDS$^{(ii)}$

Algorithm 2: Forward-longitudinal

Compute: $X_{k't} = \sum_{j \in K'_t} x_{jt}$
Compute: I_t^X
if $I_t^X = 0$ **then**
// Suppression condition met for cell k'
Assign all $j \in K'_t$ to $J^-_{k's}$ for $t \leq s \leq t + n$
Assign all $j \in \tilde{K}'_t$ to $J^+_{k't}$ for $t \leq s \leq t + n$
end if
Compute:

$$X^{(iiw)}_{k't} = \sum_{j \in \left\{K'_t \cap J^+_{k't}\right\}} \tilde{w}_{jt}\tilde{x}_{jt} + \sum_{j \in K'_t \wedge j \in J^-_{k't}} w_{jt}x_{jt} + \sum_{j \in K'_t \wedge j \notin J^-_{k't}} x_{jt}$$

Subtleties

Careful treatment of border cases

- Setting $n = 0$ is similar to the "Drop-in" case, but margins add up
- Setting $w_{js} = 0$ for $s \in (t, t + n]$ simply replaces real establishments with synthetic establishments, no phase-in
- Synthetic establishments that are in cell k' in t but are in cell k'' in $t + 1$: should they receive $\tilde{w}_{jt+1} > 0$?

Analysis

Analysis

- We implemented Algorithm 1 and 2 for Business Dynamics Statistics (BDS) tabulations by establishment age and size (`bds_e_agesz`).

Analysis

Analysis

- We implemented Algorithm 1 and 2 for BDS tabulations by establishment age and size (`bds_e_agesz`).
- Variations of w and n

Analysis

Analysis

- We implemented Algorithm 1 and 2 for BDS tabulations by establishment age and size (`bds_e_agesz`).
- Variations of w and n
- For good measure, also added a simple multiplicative noise-infused BDS$^{(n)}$ tabulation (no suppressions)

Analysis

Analysis

- We implemented Algorithm 1 and 2 for BDS tabulations by establishment age and size (`bds_e_agesz`).
- Variations of w and n
- For good measure, also added a simple multiplicative noise-infused BDS$^{(n)}$ tabulation (no suppressions)
- About 26% of all cells have some suppression

Analysis

Analysis

- We implemented Algorithm 1 and 2 for BDS tabulations by establishment age and size (`bds_e_agesz`).
- Variations of w and n
- For good measure, also added a simple multiplicative noise-infused $\text{BDS}^{(n)}$ tabulation (no suppressions)
- About 26% of all cells have some suppression
- Here: variable, “Job Creation by establishment births” (`job_creation_births`) and “Job Creation by establishment continuers” (`job_creation_continuers`)

Protection: From Kinney et al

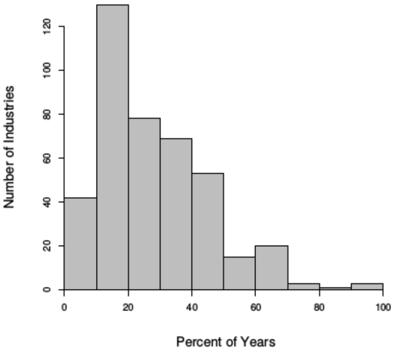


Figure 13: Histogram: Percent Distance Between Actual and Synthetic Employment

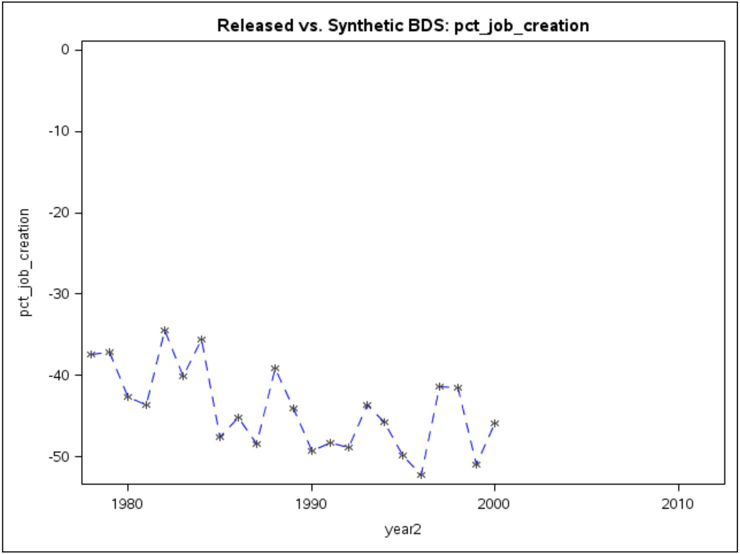
The comparison is for individual establishments, not within cells

Cell-wise comparison

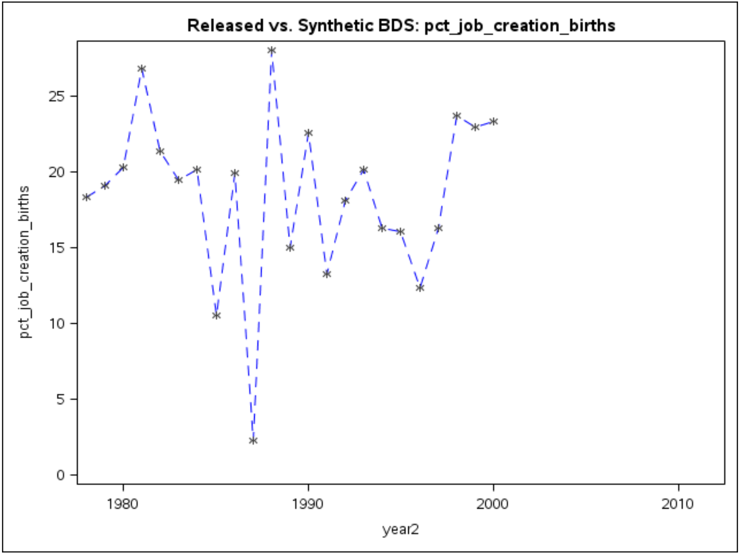
Criteria for cell-wise comparison

- Differences in count of establishment in a cell
- Differences in values of cells

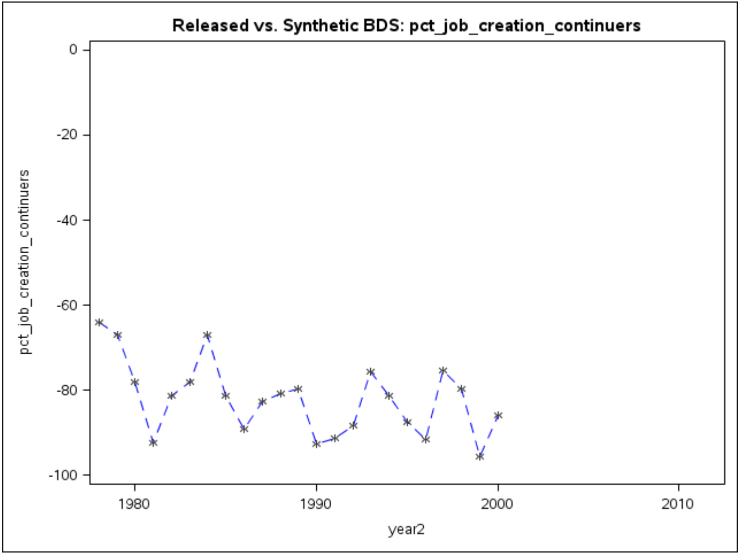
Cell-wise comparison



Cell-wise comparison



Cell-wise comparison



Analytic validity: time-series

Setup

Estimate an AR(2) process for each of (confidential) $X_{k't}$, (synthetic) $X_{k't}^{(s)}$, $X_{k't}^{(i)}$, and $X_{k't}^{(ii)}$ (and their variants)

Metrics

- number of missing time-series estimates/feasible regressions
- the number of significant coefficients for the first lag ρ_1 of the AR(2)
- *coverage*, the percentage of regressions where the true ρ_1 lies within the confidence band around the coefficient estimated from the comparison ρ_1^s and $\rho_1^{(i)}$,
- interval overlap measure J_k [6]

J_k

Consider the overlap of confidence intervals (L, U) for ρ_1 (estimated from the confidential data) and (L^*, U^*) for ρ_1^*. Let $L^{over} = \max(L, L^*)$ and $U^{over} = \min(U, U^*)$. Then the average overlap in confidence intervals is

$$J_k^* = \frac{1}{2}\left[\frac{U^{over} - L^{over}}{U - L} + \frac{U^{over} - L^{over}}{U^* - L^*}\right]$$

We then average J_k^* over all estimated AR(2) regressions.

Analytic validity: Percent missing

Table: Analytic validity: Feasibility of AR(2) regressions

Variable	Number feasible	Percent Infeasible							
	$X_{k't}$	$X^{(s)}_{k't}$	$X^{(0)}_{k't}$	$X^{(i)}_{k't}$	$X^{(in)}_{k't}$	$X^{(ii)}_{k't}$	$X^{(iiw)}_{k't}$	$X^{(iin)}_{k't}$	$X^{(n)}_{k't}$
emp	90	0	0	0	0	0	0	0	0
estabs	90	0	0	0	0	0	0	0	0
estabsentry	64	59.4	0	0	0	0	0	0	0
jobcreation	90	0	0	0	0	0	0	0	0
jobcreationbirths	90	25.6	18.9	13.3	13.3	1.1	2.2	1.1	0
jobcreationcontinuers	81	0	6.2	0	0	0	0	0	0

Analytic validity: Percent missing

Improvement in feasible regressions

- ... but not completely.
- Algorithm 2 performs better (noise-infused performs best)
- Possibly due to poor analytic validity of the underlying synthetic data for these variables (Column 2)

Analytic validity: Coverage

Table: Analytic validity: AR(2) regressions: Coverage

Variable	Coverage							
	$\rho_1^{(s)}$	$\rho_1^{(0)}$	$\rho_1^{(i)}$	$\rho_1^{(in)}$	$\rho_1^{(ii)}$	$\rho_1^{(iiw)}$	$\rho_1^{(iin)}$	$\rho_1^{(n)}$
emp	88.9	100	100	100	100	100	100	100
estabs	88.9	100	100	100	100	100	100	100
estabsentry	92.3	90.6	90.6	90.6	100	100	100	100
jobcreation	82.2	100	100	100	100	100	100	100
jobcreationbirths	89.6	91.8	91	89.7	97.8	97.7	98.9	100
jobcreationcontinuers	76.5	100	81.5	87.7	87.7	88.9	86.4	100

Analytic validity: Coverage

Improvement in coverage under Algorithm 2

- no improvement when using Algorithm 1 (but coverage of underlying synthetic data is poor)
- Only small difference between Algorithm 2 and noise-infused tabulations

Analytic validity: Overlap

Table: Analytic validity: AR(2) regressions: Interval overlap

Variable	Interval overlap							
	$J_k^{(s)}$	$J_k^{(0)}$	$J_k^{(i)}$	$J_k^{(in)}$	$J_k^{(ii)}$	$J_k^{(iiw)}$	$J_k^{(iin)}$	$J_k^{(n)}$
emp	83.4	99.4	100	100	100	100	100	97.7
estabs	80.4	97.6	100	100	100	100	100	97.8
estabsentry	78.7	82.6	82.6	82.6	100	100	100	95.8
jobcreation	73.3	94.4	100	100	100	100	100	96
jobcreationbirths	72.9	80.9	81.5	79.9	91.9	91.9	91.8	94.5
jobcreationcontinuers	70.7	92.6	77.5	81.6	85.1	85.3	85	95.9

Analytic validity: Overlap

Similar picture to the Coverage statistics

- no improvement when using Algorithm 1 (but coverage of underlying synthetic data is poor)
- bigger difference between Algorithm 2 and noise-infused tabulations (but notice deterioration in non-sensitive cells)

Open issues

Unexplored issues

- SynLBD is synthesized independently within industry

Open issues

Unexplored issues

- SynLBD is synthesized independently within industry
- Geography is not synthesized, not considered within synthesis process (and not released) - unclear how geography subtabulations will fare, what the disclosure avoidance implications are

Open issues

Unexplored issues

- SynLBD is synthesized independently within industry
- Geography is not synthesized, not considered within synthesis process (and not released) - unclear how geography subtabulations will fare, what the disclosure avoidance implications are
- Firm-level characteristics go into a bit more detail, and require availability of SynLBD v3

Open issues

Unexplored issues

- SynLBD is synthesized independently within industry
- Geography is not synthesized, not considered within synthesis process (and not released) - unclear how geography subtabulations will fare, what the disclosure avoidance implications are
- Firm-level characteristics go into a bit more detail, and require availability of SynLBD v3
- Time consistency of the series

Open issues

Unexplored issues

- SynLBD is synthesized independently within industry
- Geography is not synthesized, not considered within synthesis process (and not released) - unclear how geography subtabulations will fare, what the disclosure avoidance implications are
- Firm-level characteristics go into a bit more detail, and require availability of SynLBD v3
- Time consistency of the series
- Comparison to alternative “outside-the-firewall” imputation mechanisms ([4, 2])

Conclusion

Early in the process

- Desirable a-priori properties (use of public-use data to fill in blanks)
- May not work for other variables
- Assumes suppression as primary disclosure avoidance mechanism...

Thank you

More info:

- For information on the SynLBD, see goo.gl/eyrv7w
- Access through the Synthetic Data Server, www.vrdc.cornell.edu/sds/

Extra slides

Bibliography

J. M. Abowd, K. Gittings, K. L. McKinney, B. E. Stephens, L. Vilhuber, and S. Woodcock, “Dynamically consistent noise infusion and partially synthetic data as confidentiality protection measures for related time-series,” Federal Committee on Statistical Methodology, Tech. Rep., January 2012. [Online]. Available: http://www.fcsm.gov/events/papers2012.html

J. R. Bradley, S. H. Holan, and C. K. Wikle, “Mixed Effects Modeling for Areal Data that Exhibit Multivariate-Spatio-Temporal Dependencies,” *ArXiv e-prints*, Jul. 2014.

R. K. Gittings, “Essays in labor economics and synthetic data methods,” Ph.D., Cornell University, 2009.

S. H. Holan, D. Toth, M. A. R. Ferreira, and A. F. Karr, “Bayesian multiscale multiple imputation with implications for data confidentiality,” *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 564–577, 2010. [Online]. Available: http://dx.doi.org/10.1198/jasa.2009.ap08629

R. Jarmin and J. Miranda, “The Longitudinal Business Database,” U.S. Census Bureau, Center for Economic Studies, Discussion Paper CES-WP-02-17, 2002.

A. F. KARR, C. N. KOHNEN, A. OGANIAN, J. P. REITER, and A. P. SANIL, “A framework for evaluating the utility of data altered to protect confidentiality,” *The American Statistician*, vol. 60, no. 3, pp. 1–9, 2006.

S. K. Kinney, J. P. Reiter, A. P. Reznek, J. Miranda, R. S. Jarmin, and J. M. Abowd, “Towards unrestricted public use business microdata: The Synthetic Longitudinal Business Database,” *International Statistical Review*, vol. 79, no. 3, pp. 362–384, December 2011. [Online]. Available: http://ideas.repec.org/a/bla/istatr/v79y2011i3p362-384.html

J. Miranda and L. Vilhuber, “Using partially synthetic data to replace suppression in the business dynamics statistics: Early results,” in *Privacy in Statistical Databases*, ser. Lecture Notes in Computer Science, J. Domingo-Ferrer, Ed. Springer International Publishing, 2014, vol. 8744, pp. 232–242. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-11257-2_18

Panel on Measuring Business Formation, Dynamic

Acronyms

BDS Business Dynamics Statistics