Redesigning the Longitudinal Business Database: Online Appendices*

Melissa Chow† Teresa C. Fort‡ Christopher Goetz†   Nathan Goldschlag†   James Lawrence†   Elisabeth Ruth Perlman†   Martha Stinson†   T. Kirk White†

May 2021

Appendix A   Data Dictionaries

Table 1: lbd{year} Data Dictionary

Variable Name	Type	Description
act	2	Active status flag
act_src	2	Source of act
alpha	2	Multi-unit identifier
alpha_src	2	Source of alpha
bds_cty	2	Time invariant county
bds_denom	1	Average employment t and $t-1$ with BDS smoothing
bds_denom_noise	1	Average employment t and $t-1$ with noise
bds_emp	1	Employment with BDS smoothing
bds_emp_noise	1	Employment with BDS smoothing, with noise
bds_emp_tm1	1	Employment in $t-1$ with BDS smoothing
bds_emp_tm1_noise	1	Employment in $t-1$ with BDS smoothing, with noise
bds_entry	2	Entry flag for BDS tabulations, 1 if employment in t-1 is 0 and employment in t is >0, 0 otherwise
bds_exit	2	Exit flag for BDS tabulations, 1 if employment in t-1 is >0 and employment in t is 0, 0 otherwise
bds_firmsize_emp	1	Average firm employment in t and $t-1$ with BDS smoothing
bds_ifirmsize_emp	1	Firm employment in $t-1$ with BDS smoothing
bds_msa	2	MSA geographic identifier, Feb2013 vintage, matched using the time invariant state and county codes.
bds_neg	1	Negative employment changes, job destruction, with BDS smoothing
bds_pos	1	Positive employment changes, job creation, with BDS smoothing
bds_st	2	Time invariant state code
bds_tab	2	BDS scope control flag
bds_vcnaics	2	Vintage consistent industry code with inter-censal adjustments
cbp_tab	2	CBP scope flag
cty	2	County
cty_src	2	Source of county
ein	2	Employer Identification Number

ein_rorg	2	Employer Identification Number for reorg cases.
ein_src	2	Source of ein
emp	1	Employment
emp_src	2	Source of employment
estabid	2	Establishment identifier (cfn or empunit_id_char)
estabid_flg	2	Flag that captures how estabid was selected from links file. Digits of the flag are as follows 1. Linkfile ID selected (1/2) 2. ID2 non-missing (0/1) 3. Emp ID1 non-missing (0/1) 4. Emp ID2 non-missing (0/1). Example: 2101 ESTABID is ID2, ID2 was present, ID1 employment was missing, and ID2 employment was non-missing.
estabid_rorg	2	Establishment identifier (cfn or empunit_id_char) associated with reorg ID not used for LBD attributes (id2 if id1 selected or id1 if id2 selected).
estabid_tm1	2	Establishment identifier for $t-1$
firmid	2	Firm identifier, "0"+ein for single units, alpha+"0000" for multi units
firmid_rorg	2	Firm identifier, "0"+ein_rorg for reorg cases.
firmid_tm1	2	Firm identifier for establishment in $t-1$
firstyear_emp	1	First year lbdnum is observed with positive employment
firstyear_pay	1	First year lbdnum is observed with positive payroll
flag_retime	1	Retiming flag
lastyear_emp	1	Last year lbdnum is observed with positive employment
lastyear_pay	1	Last year lbdnum is observed with positive payroll
lbdfid	2	Longitudinal firm identifier.
lbdfid_tm1	2	Longitudinal firm identifier for t-1.
lbdnum	2	Longitudinal establishment identifier
lfo	2	Legal form of organization
lfo_src	2	Source of lfo, BR or CBP
link_flg	2	Flag that captures what type of link was used to create lbdnum
merge_tm1	2	Flag to capture whether lbdnum exists in t-1, "tm1" is only in t-1, "t" is only in "t", "both" are both in t-1 and t
mu	2	Multi unit indicator
naics	2	NAICS code
naics_src	2	Source of naics

nfactor	1	Noise factor
nfactor_tm1	1	Noise factor $t-1$
pay	1	Final annual payroll, using both CBP, BR, and t+1 information
pay_flg	2	Flag for source of final annual payroll
pdiv	2	Processing division code
pdiv_src	2	Source of pdiv, BR or CBP
qp1	1	Quarter 1 payroll
qp2	1	Quarter 2 payroll
qp3	1	Quarter 3 payroll
qp4	1	Quarter 4 payroll
rec_src_flg	2	Record source flag to capture whether an estab was found in the BR and/or the CBP
sic	2	Standard Industrial Classification
sic_src	2	Source of sic
st	2	State
st_src	2	Source of state
toc	2	Type of operation code.
toc_src	2	Type of operation code source.
year	1	Year of observation.
zip	2	Zip code
zip_src	2	Source of zip code

Notes: Variable type 1 is numeric and variable type 2 is character.

Table 2: lbdfirm{year} Data Dictionary

Variable Name	Type	Description
firmage_emp	1	Firm employment-based age.
firmage_pay	1	Firm payroll-based age.
firmdeath_emp	1	Firm employment-based death flag.
firmdeath_pay	1	Firm payroll-based death flag.
firmid	2	Firmid, "0"+ein or alpha+"0000".
firm_firstyear_emp	1	First year firm appears with positive employment.
firm_firstyear_pay	1	First year firm appears with positive payroll.

		lbdfirm{year} Data Dictionary (cont.)
firm_initialyear_emp	1	Implied employment-based first year for firm based upon oldest establishment. Equal to min(firstyear_emp) of estabs linked to the firm in the firm's firm_firstyear_emp.
firm_initialyear_pay	1	Implied payroll-based first year for firm based upon oldest establishment. Equal to min(firstyear_pay) of estabs linked to the firm in the firm's firm_firstyear_pay.
firm_lastyear_emp	1	Last year firm appears with positive employment.
firm_lastyear_pay	1	Last year firm appears with positive payroll.
firmsize_bds_emp	1	Firm total BDS employment (bds_emp) in current year for in scope estabs (bds_tab=1).
firmsize_bds_emp_tm1	1	Firm total BDS employment (bds_emp) in (t-1) for in scope estabs (bds_tab=1).
firmsize_bds_pay	1	Firm total payroll in current year for in scope estabs (bds_tab=1).
firmsize_bds_pay_tm1	1	Firm total payroll (pay) in (t-1) for in scope estabs (bds_tab=1).
firmsize_emp	1	Firm total employment (emp) in current year for all estabs.
firmsize_emp_tm1	1	Firm total employment (emp) in (t-1) for all estabs.
firmsize_pay	1	Firm total payroll (pay) in current year for all estabs.
firmsize_pay_tm1	1	Firm total payroll (pay) in (t-1) for all estabs.
lbdfid	2	Longitudinal firm identifier.
year	1	Year of observation.

Notes: Variable type 1 is numeric and variable type 2 is character.

Appendix B   Additional Tables

Table 3: Edited Flag Codes on Linkwide File

Code	Description	Example	Fix
I01	Continuer: Erroneous Year1 Reorg	Estab A is a continuer in t-2/t-1 link file, but a t-1 re-org from B to A in the t-1/t file.	The progression from A to B to A is deemed erroneous. A is a normal continuer and is decoupled from B in t-1.
I02	Continuer: erroneous Year2 Reorg	Estab A re-orgs to B in t-2/t-1 link file, but is either a normal continuer or re-orgs to C in t-1/t.	New t-1/t link file is deemed correct, and re-org from t-2/t-1 link file erroneous. A is decoupled from B in t-1.
I03	Death: Decoupled Reorg	In the above case (I01), the re-org is erroneous and estab B does not continue into year t.	B is decoupled from A, and represents an exit in t.
I04	Continuer: Decoupled from erroneous Reorg	In the above case (I02), the re-org is erroneous and decoupled estab B is present in t-2 and t-1.	B is decoupled from A in t-1, and is a normal t-2/t-1 continuer.
I05	Continuer: Re-linked to a decoupled t-1 re-org	In the above case (I02), the re-org is erroneous and decoupled estab B is present in year t-1/t	B is decoupled from A in t-1, and is a normal t-1/t continuer.
I06	Birth: Decoupled from erroneous Reorg	In any of the above cases, B is only present in year t-1, and not in t-2.	B is decoupled from A, and represents a birth in t-1
L01	Reactivation from t-2: erroneous Reorg	Estab re-orgs from A to B in year t-2, but then becomes inactive. Returns in year t as A.	Link on A, forming a 2-yr reactivation. Decouple A from B in t-2.
L02	Reactivation from t-3: erroneous Reorg	Estab re-orgs from A to B in year t-3, but then becomes inactive. Returns in year t as A.	Link on A, forming a 3-yr reactivation. Decouple A from B in t-3.
L03	Reactivation from t-4: erroneous Reorg	Estab re-orgs from A to B in year t-4, but then becomes inactive. Returns in year t as A.	Link on A, forming a 4-yr reactivation. Decouple A from B in t-4.
L04	Reactivation from t-5: erroneous Reorg	Estab re-orgs from A to B in year t-5, but then becomes inactive. Returns in year t as A.	Link on A, forming a 5-yr reactivation. Decouple A from B in t-5.
L05	Reactivation from t-6: erroneous Reorg	Estab re-orgs from A to B in year t-6, but then becomes inactive. Returns in year t as A.	Link on A, forming a 6-yr reactivation. Decouple A from B in t-6.
L06	Reactivation from t-7: erroneous Reorg	Estab re-orgs from A to B in year t-7, but then becomes inactive. Returns in year t as A.	Link on A, forming a 7-yr reactivation. Decouple A from B in t-7.

Table 4: flag_match_pass_{year} Values

A: Continuers with ID matches	
A01: Year 1-Year 2 ID = Continuer, ID match	
A02: Year 1-Year 2 PPN = Continuer, PPN match	
A03: Year 1-Year 2 OLDID = Continuer, OLDID match	
A04: Year 1-Year 2 PS_ID = Continuer, PS_ID match	
B: Single-unit Deaths	
B01: Single-units No Year 1 REORG match = Death, no reorg matches	
B02: Single-units No Year 1 REORG, no address – non-match = Death, no reorg matches	

flag_match_pass_{year} Values
C: Single-unit Births
C01: Single-units No Year 2 REORG match = Birth, no reorg matches
C02: Single-units No Year 2 REORG, no address – non-match = Birth, no reorg matches
D: Transition to MU establishment
D01: Year 1-Year 2 SU-MU EIN+estab match = Continuer, SU-MU match
D02: Year 1-Year 2 SU-MU OLDEI+estab match = Continuer, SU-MU match
D03: Year 2 MU estabs match at EIN level but not estab level = MU estab birth
D04: Year 2 MU estabs match at OLDEI level but not estab level = MU estab birth
D05: Year 2 MU estabs that don't match by EIN or OLDEI to year 1 SU estab = MU estab birth
E: Transition from MU establishment
E01: Year 1-Year 2 MU-SU EIN+estab match = Continuer, MU-SU match
E02: Year 1 MU estabs match at EIN level but not estab level= MU estab death
E03: Year 1 MU estabs that dont match by EIN or OLDEI to year 2 SU estab = MU estab death
F: Year-to-year Name and Address Match (BITS and LBD)
F01: Year1-Year2 Exact NAME SU-SU = Continuer, BITS match
F02: Year1-Year2 Pseudo NAME SU-SU = Continuer, BITS match
F03: Year1-Year2 Standardized NAME SU-SU = Continuer, BITS match
F04: Year 1-Year 2 ADDRESS SU-SU = Continuer, BITS match
F05: Year1-Year2 LBD SU-SU = Continuer, LBD match
G: Year 2 Reorg match (BITS and LBD)
G01: Year 2 REORG Exact name match = Mid-year2 Reorg, BITS match
G02: Year 2 REORG Pseudo name match = Mid-year2 Reorg, BITS match
G03: Year 2 REORG Standardized name match = Mid-year2 Reorg, BITS match
G04: Year 2 REORG Address match = Mid-year2 Reorg, BITS match
G05: Year2 REORG LBD = Mid-year2 Reorg, LBD match
H: Year 1 Reorg match (BITS and LBD)
H01: Year 1 REORG Exact name match = Mid-year1 Reorg, BITS match
H02: Year 1 REORG Pseudo name match = Mid-year1 Reorg, BITS match
H03: Year 1 REORG Standardized name match = Mid-year1 Reorg, BITS match
H04: Year 1 REORG Address match = Mid-year1 Reorg, BITS match
H05: Year1 REORG LBD = Mid-year1 Reorg, LBD match
I: Records edited due to Erroneous reorg
I01: Continuer, erroneous Year1 Reorg (year t value, when t-1 reorg deemed erroneous)
I02: Continuer, erroneous Year2 Reorg (year t-1 value, edited during year t processing)
I03: Death: Decoupled Reorg (year t value created during year t processing)
I04: Continuer: Decoupled from erroneous Reorg (year t-1 value edited during year t processing)
I05: Continuer, re-linked to a decoupled t-1 re-org (year t value, when t-1 reorg deemed erroneous)
I06: Birth: Decoupled from erroneous Reorg (year t-1 value, when t-1 reorg deemed erroneous)
J: Reactivations with no reorg
J01: Reactivation from t-2, no Reorg
J02: Reactivation from t-3, no Reorg
J03: Reactivation from t-4, no Reorg
J04: Reactivation from t-5, no Reorg
J05: Reactivation from t-6, no Reorg
J06: Reactivation from t-7, no Reorg

flag_match_pass_{year} Values
K: Reactivations with Reorg
K01: Reactivation from t-2, w/ Reorg
K02: Reactivation from t-3, w/ Reorg
K03: Reactivation from t-4, w/ Reorg
K04: Reactivation from t-5, w/ Reorg
K05: Reactivation from t-6, w/ Reorg
K06: Reactivation from t-7, w/ Reorg
L: Reactivations with Reorgs later deemed to be erroneous
L01: Reactivation from t-2, erroneous Reorg (t-1 value edited during year t processing)
L02: Reactivation from t-3, erroneous Reorg (t-1 value edited during year t processing)
L03: Reactivation from t-4, erroneous Reorg (t-1 value edited during year t processing)
L04: Reactivation from t-5, erroneous Reorg (t-1 value edited during year t processing)
L05: Reactivation from t-6, erroneous Reorg (t-1 value edited during year t processing)
L06: Reactivation from t-7, erroneous Reorg (t-1 value edited during year t processing)

Table 5: pdiv Values

Processing Division Code	1987-1988	1988-1996	1997-1999	2000-2001	2002	2003-2014
American Samoa (inscope)				0	0	0
Mining	1	1	1	1	1	1
Construction	2	2	2	2	2	2
Manufacturing (Non-ASM)	3	3	3	3	3	3
Transportation (inscope)	4	4	4	4	4	4
Wholesale	5	5	5	5	5	5
Retail	6	6	6	6	6	6
Service (inscope)	7	7				
Professional, scientific, and technical services			7	7	7	7
Auxiliaries (inscope)					a	8
Central administrative offices and auxiliaries	8	8				
Puerto Rico (inscope)	9	9	9	9	9	9
Foreign	A	A	A	A	A	A
Utilities				B	B	B
Communications						
Information				C	C	C
Health care and social assistance				D	D	D
Arts, entertainment, and recreation				E	E	E
Agriculture	F	F	F	F	F	F
Government (out of scope)	G	G	G	G	G	G
Government (inscope)	H	H	H	H	H	H
Manufacturing (ASM)	I	I	I	I	I	I
Finance						
Finance and insurance				J	J	J
Insurance						
Accommodation and food services				K	K	K
Real estate						
Real estate and rental and leasing				L	L	L
Submaster	M	M	M	M	M	M
Northern Mariana Islands (inscope)				N	N	N
Northern Mariana Islands (out of scope)				O	O	O

	pdiv Values					
Puerto Rico (out of scope)	P	P	P	P	P	P
Management of companies and enterprises			Q	Q	Q	Q
Other services (except public administration)			R	R	R	R
Out of scope	S	S				
Out of scope (includes agriculture services, forestry, fishing, and hunting)			S	S	S	S
Educational services			T	T	T	T
Guam (inscope)	U	U	U	U	U	U
Virgin Islands (inscope)	V	V	V	V	V	V
Guam (out of scope)	W	W	W	W	W	W
Unclassified	X	X	X	X	X	X
Administrative and support, waste management and remediation services			Y	Y	Y	Y
Virgin Islands (out of scope)	Z	Z	Z	Z	Z	Z

Table 6: 2018 BDS Tabulations

By-Variable Combinations
year
year, bds_fage
year, bds_eage
year, bds_fsize
year, bds_esize
year, bds_ifsize
year, bds_iesize
year, bds_metro
year, bds_st
year, bds_msa
year, bds_cty
year, bds_sector
year, bds_vcnaics3
year, bds_vcnaics4
year, bds_fage, bds_fsize
year, bds_fage, bds_ifsize
year, bds_eage, bds_esize
year, bds_eage, bds_iesize
year, bds_st, bds_sector
year, bds_st, bds_metro
year, bds_st, bds_fage
year, bds_st, bds_fsize
year, bds_st, bds_ifsize
year, bds_st, bds_eage
year, bds_st, bds_esize
year, bds_st, bds_iesize
year, bds_sector, bds_fage
year, bds_sector, bds_fsize
year, bds_sector, bds_ifsize
year, bds_sector, bds_eage
year, bds_sector, bds_esize
year, bds_sector, bds_iesize
year, bds_vcnaics3, bds_fage
year, bds_vcnaics3, bds_fsize
year, bds_vcnaics3, bds_ifsize
year, bds_vcnaics3, bds_eage
year, bds_vcnaics4, bds_fage
year, bds_vcnaics4, bds_fsize
year, bds_vcnaics4, bds_ifsize
year, bds_vcnaics4, bds_eage
year, bds_metro, bds_fage
year, bds_metro, bds_fsize
year, bds_metro, bds_ifsize

BDS Tables
year, bds_metro, bds_eage
year, bds_metro, bds_esize
year, bds_metro, bds_iesize
year, bds_metro, bds_sector
year, bds_msa, bds_fagecoarse
year, bds_msa, bds_fsizecoarse
year, bds_msa, bds_ifsizecoarse
year, bds_msa, bds_eagecoarse
year, bds_msa, bds_sector
year, bds_cty, bds_fagecoarse
year, bds_cty, bds_fsizecoarse
year, bds_cty, bds_ifsizecoarse
year, bds_cty, bds_eagecoarse
year, bds_cty, bds_sector
year, bds_msa, bds_sector, bds_fagecoarse
year, bds_msa, bds_sector, bds_fsizecoarse
year, bds_msa, bds_sector, bds_ifsizecoarse
year, bds_msa, bds_sector, bds_eagecoarse
year, bds_st, bds_sector, bds_fage
year, bds_st, bds_sector, bds_fsize
year, bds_st, bds_sector, bds_ifsize
year, bds_st, bds_sector, bds_eage
year, bds_metro, bds_sector, bds_fage
year, bds_metro, bds_sector, bds_fsize
year, bds_metro, bds_sector, bds_ifsize
year, bds_metro, bds_sector, bds_eage
year, bds_metro, bds_st, bds_fage
year, bds_metro, bds_st, bds_fsize
year, bds_metro, bds_st, bds_ifsize
year, bds_metro, bds_st, bds_eage
year, bds_metro, bds_st, bds_sector
year, bds_st, bds_metro, bds_sector, bds_fage
year, bds_st, bds_metro, bds_sector, bds_fsize
year, bds_st, bds_metro, bds_sector, bds_ifsize
year, bds_st, bds_metro, bds_sector, bds_eage

Appendix C   Vintage-consistent NAICS

In this appendix section, we provide additional details on the methods used to assign the codes in the `bds_vcnaics` variable.

C.1   Vintage-consistent NAICS assignment flags

The *naics{year}_flag* files contain flags with information on how the codes in the `bds_vcnaics` variable were assigned. The source flag describes the data source for the code, which are presented in Table 7. The method flag indicates how the code was assigned, and the year flag indicates the

year from which the assigned code originates. Table 8 displays the potential assignment methods for `bds_vcnaics`. The NASS memo in the LBD documentation directory contains more information on each of these methods.

Assignment of the code to `bds_vcnaics` often involves several steps. The current source and method flags only document the last source or method in the process. We aim to track all sources and methods in these flags in future versions of the codes.

Table 7: Data sources for vc_naics_bds

Source	Description
LBD	Longitudinal Business Database
Census	Economic Census
FK	Fort-Klimek assignment

Notes: Table describes the values of the variable `fk_naics{year}_source` in the *naicsyear_flags* files. `fk_naics{year}_source` provides information on the final source of the `bds_vcnaics` variable.

Table 8: VC NAICS Assignment Methodologies

Method	Description
A	Original code
B	Unique concordance
C	Census timing fix
D	Replace consistent partials
E	Consistent old codes, roll new code
F	Roll unique, consistent
G	End, roll closest code
H	Gap, consistent code ends
I	Multiple mapping, roll unique consistent
K	Out of scope
L	Incorrect vintage rolled back
M	CFN roll
N	Proximity rolls
R	Random assignment
T	Potential auxiliary, roll 2002 back
U	Census of Auxiliaries
V	SIC auxiliary bridge code
W	TOC=81 in majority of years

Notes: Table describes the values of the variable `fk_naics{year}_method` in the *naics{year}_flags* files. `fk_naics{year}_method` provides information on the final assignment method of the `bds_vcnaics` variable.

As described in Section **??**, the variable `fk_{indcode}_splits` denotes the potential number of new vintage codes to which the old vintage code ("indcode") could have been assigned. Splits variables are available for each vintage transition, and are populated for observations that were randomly assigned an industry code. Researchers can obtain the total number of splits across all vintage

transitions for an establishment in a given year by summing across all the `fk_{indcode}_splits` variables.

C.2 Vintage-consistent NAICS auxiliary assignments

In this section, we provide details on the methods and data sources we use to identify auxiliary establishments in the SIC era.

Table 9: TOC value mappings

TOC value	NAICS
80	551110
81	551114
82	541700
83	493100
87	551110
88	551114
89	551114

Notes: Table describes the TOC variable values and corresponding NAICS codes.

Table 10: 1997 SIC codes and mappings to auxiliary information

code	NAICS
05	518210
15	541700
35	811000
40	493100
45	484000
55	541200
65	541800
70	541114
89	541114
90	541114

Notes: Table describes the last two digits of the internal SIC97 codes and their mappings to NAICS codes. We thank John Murphy for this information.