README file for data and programs for Manuscript No. AEJMacro-2009-0101

"Business Volatility, Job Destruction and Unemployment" by Steven J. Davis, R. Jason Faberman, John Haltiwanger, Ron Jarmin, and Javier Miranda

Overview:

This document describes the files in the zip file AEJ_Macro_2009_0101_data.zip. The description below is organized as follows. Section I describes the final datasets and programs used to produce the figures and tables in the main text as well as the final datasets and programs used to produce the web appendix for the paper. Section II describes the intermediate files in the zip file, a description of the data sources and the steps required to generate the intermediate and final files.

Note that *.dta are STATA datasets, *.do are STATA programs, *.sas7bdat are SAS datasets, *.sas are SAS programs, and *.xls are EXCEL spreadsheets.

I. **Final Datasets and Programs**

Tables

For the results in each table, the list below provides the final dataset used to produce the table as well as the code used to generate the results presented in the table.

Table 1:
1. Data: bed_cps.sas7bdat (SAS dataset)
2. Program: bed_cps_stats.sas (SAS code)
Table 2:
1. Data: lbd_cps.sas7bdat (SAS dataset)
2. Program: lbd_cps_stats.sas (SAS code)
Tables 3 and 4:
1. BED_CPS results:
    a. Data: bed_cps_3year.dta, bed_cps_3year.sas7bdat
    b. Programs: tables_3_4_bed_cps.do, bed_cps_stats.sas Note: STATA code is primary code, SAS code generates within r-squared.
2. LBD_CPS results:
    a. Data: lbd_cps_3year.dta, lbd_cps_3year.sas7bdat
    b. Programs: tables_3_4_lbd_cps.do, lbd_cps_stats.sas
    c. Note: STATA code is primary code, SAS code generates within r-squared.
Table W1
1. BED_CPS results:
    a. Data: bed_cps_outlier.dta, bed_cps_outlier.sas7bdat
    b. Programs: table_w1a_bed_cps.do, bed_cps_without_cons_outlier.sas
2. LBD_CPS results:
    a. Data: lbd_cps_3year_nofire.dta, lbd_cps_3year_nofire.sas7bdat
    b. Programs: table_w1b_lbd_cps, lbd_cps_nofire.sas

Five-year horizon robustness check (results discussed in text)
1. BED_CPS results:
   a. Data: bed_cps_5year.dta
   b. Program: regress_bed_5year.do
2. LBD_CPS results:
   a. Data: lbd_cps_5year.dta
   b. Program: regress_lbd_5year.do

Figures

Figures are excel spreadsheets and are mostly self-explanatory (e.g., Figure1_2.xls includes the figure and underlying data points for Figures 1 and 2). Details about the source data used are in section II below (and also see text for additional description). A few notes on specific figures are as follows:
1. The data for the scatterplots in Figures 9 and 10 and W1 and W2 is generated from the sas programs bed_cps_stats.sas and lbd_cps_stats.sas.
2. The counterfactuals for Figures 11 and 12 are programmed into the excel spreadsheets in Figures11_12.xls.

II. Source Data and Intermediate Steps

1. Source Data:
   a. CPS data:
      i. The CPS data are from LABSTAT and can be downloaded directly from the BLS web site at http://data.bls.gov/cgi-bin/srgate. The files "unemployment codes.xls" and "duration codes by industry and occupation.xls" in the zip file describe the series codes for the analysis in this paper. For the analysis in this paper, only the duration codes by industry are used. Several series are downloaded included SIC based series from 1976-2002 and NAICS based series from 2000-2005. A description of how these series are integrated is described below in the intermediate steps description. As described in the text, the numbers of unemployed by industry and numbers of unemployed less than five weeks by industry are based on the experienced unemployed (i.e., new entrants are not included). The industry is for the prior job.
   b. BED data:
      i. The BED job flow series are available from http://www.bls.gov/bdm/. We use and extended version of the BED job flow series created by Faberman (2008). This series is similar to the publicly available statistics, but also includes estimates for the 1990-92 period produced using establishment microdata housed at the BLS. See the text for more description and this reference.
   c. LBD/BDS data:

i. The job flow series created from the LBD are available from the Business Dynamic Statistics (BDS) at http://www.ces.census.gov/index.php/bds/bds_home. We have used a slightly modified version of the BDS for the job flows in decomposing manufacturing into durables and nondurables manufacturing.

   ii. In addition, we have used the underlying LBD establishment level data to generate the establishment-volatility and establishment dispersion measures. The methodology for generating these measures is described in the text as well as in greater detail in Davis, Haltiwanger, Jarmin and Miranda (2007) (the latter describes the methodology for generating firm volatility and firm dispersion measures).

   iii. The LBD is available for access at Census/NSF/RDC centers for approved projects. Go to http://www.ces.census.gov/ for information on how to apply for access.

   d. Spliced MTD/LRD/BED Manufacturing series:

   i. This series is taken from Davis, Faberman and Haltiwanger (2006) and Faberman (2008). The series represents an econometric splicing of BED data from Faberman (2008) with a merged MTD/LRD series of estimates created by Davis and Haltiwanger (1999). The methodology and sources used are described in detail in Faberman (2008).

   e. JOLTS data for Figures 1 and 2:

   i. Figures 1 and 2 are based on tabulations of worker flow rates from JOLTS micro data housed at the BLS and are replications of the analyses conducted in Davis, Faberman and Haltiwanger (2006) and Davis, Faberman and Haltiwanger (2009). See the latter for details about the JOLTS data and the methodology used to create Figures 1 and 2.

2. Intermediate steps:

   a. The annual LBD_CPS integrated data are in lbd_cps.sas7bdat. The 3 year version of the dataset is created in lbd_cps_stats.sas. Five year versions or other robustness checks are easily created by modifying lbd_cps_stats.sas.

   b. The quarterly BED_CPS integrated data are in bed_cps.sas7bdat. The 3 year version of the dataset is create in bed_cpst_stats.sas

   c. Details on intermediate steps in creating CPS tabulations for above datasets:

   i. Starting with the monthly raw data we create an SIC 1976-2005 dataset for integrating with the LBD and a NAICS 1990-2005 dataset for integrating with the BED. The details of this splicing are below but before splicing we make the Shimer (2005) adjustment for the CPS 1994 redesign by multiplying the duration less than 5 weeks by 1.1 for series from 1994 forward. The splicing is then completed as follows:

1. There is overlap on a monthly basis for both SIC and NAICS based CPS data for 2000-2002.
2. To create splice factors for converting SIC to NAICS, we run regressions of the NAICS series on the SIC series. We use those factors to convert the SIC series to NAICS series going backwards.
3. For most broad sectors, there is essentially a one-to-one mapping between SIC and NAICS so this is a straightforward matter. The only complex broad sector is services where for NAICS we combine sectors Information, Professional and Business Services, Education and Health, Leisure and Hospitality and Services for the overall service sector that approximately matches the broad sector Services in the SIC sector. Since some of these NAICS sectors came from nondurables manufacturing and from transportation, communications and utilities we include those SIC sectors as additional RHS variables for the NAICS services regression. We found the results were not sensitive to adding these variables.
4. We did the reverse for creating SIC based 2002-2005 versions of the CPS series.

ii. After splicing, we ran x-11 on each industry, month series to create seasonally adjusted monthly series.
iii. Using the monthly series, we create the inflow and escape rates using the number employed, number unemployed and number unemployed less than five weeks for each industry, month using the formulas in the text.
iv. We then took quarterly and annual averages using the appropriate time intervals for the BED and LBD respectively (e.g., for LBD we use the April through February averages since LBD is March to March changes).
v. With the quarterly average and annual average CPS series, we merge to the respective BED and LBD/BDS based series by quarter/industry or year/industry as appropriate. The resulting datasets are the bed_cps.sas7bdat and lbd_cps.sas7bdat datasets described above.