Skewed Business Cycles*

Replication of Empirical Results

Sergio Salgado† Fatih Guvenen‡ Nicholas Bloom§

April, 2020

1 Introduction

This document provides additional details for the replication of the empirical results of the paper. The replication packet of two main sets of files. The first uses the set of moments already calculated from the raw data to create the main figures of the paper. The second takes raw data and creates cross-sectional moments.

Replication of Main Empirical Results

To replicate the main figures and tables of the paper, use the following files:

- *SBC_AGGREGATE.xls* which contains aggregate time-series of annual GDP growth from the Federal Reserve Economic Data (FRED) for the United States, quarterly GDP growth from FRED for the United States, cross-country annual GDP growth from the World Development Indicators from the World Bank, and cross-country quarterly GDP growth from OECD Stats.
- *SBC_CENSUS_LBD.xls* which contains the cross-sectional moments of the distribution of employment growth for establishments and firms from the Census Bureau's Longitudinal Business Database (LBD). Section 2.2 of this document describes the available material in additional detail.
- *SBC_CENSUS_ASM.xls* which contains the data used to generate the scatter plot of firm-level productivity shocks constructed using data from the US Census of Manufacturing (CMF) and the Annual Survey of Manufacturing (ASM).

*Part of this research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1694. All results based on US Census data have been reviewed to ensure that no confidential information is disclosed. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the US Census. The replication packet for the empirical results of the paper is available on the authors' websites, and in GitHub at the following link: https://github.com/salga010/SBC-Replication.

†The Wharton School, University of Pennsylvania; ssalgado@wharton.upenn.edu

‡University of Minnesota, FRB of Minneapolis, and NBER; guvenen@umn.edu

§Stanford University, SIEPR, and NBER; nbloom@stanford.edu

- *SBC_ USA_ AND_ CROSSCOUNTRYDEC2019.xls* which contains the cross-sectional moments of the distribution of annual growth of quarterly sales, annual sales, annual employment, and stock returns for publicly traded firms from Compustat; Annual sales growth and employment growth for publicly traded firms from the Bureau van Dijk's (BvD) Osiris database; Stock returns from Global Compustat; And annual employment growth, sales growth, and firm-level productivity measures from the BvD's Amadeus database. Section 2.3 describes the available material in more detail.

- *SBC_ VAR.xls* which contains the monthly macro-time series used to estimate the VAR results presented in the paper.

- *SBC_ ReplicaEmpirics.do* takes the .xls files and replicates the figures and tables in the paper.

- *SBC_ ReplicaVAR.do* takes *SBC_ ReplicaVAR.xls,* estimates the VAR and plots the VARs of the paper.

- *SBC_ SPCOVID.do* takes stock price data from WRDS and replicates the stock returns analysis for Figure 2.

Compustat, CRSP, Global Compustat, BvD Osiris, and BvD Amadeus were accessed via the Wharton Research Data Services (WRDS). LBD, CMF, and ASM were accessed at a Federal Statistical Research Data Center under FSRDC Project Number 1694; Underlying data used in our calculations using Census information is not disclosed.

Replication from Raw Data

To replicate the cross-sectional moments from the raw data and do your own calculations, use the following files:

- SBC_ Clean_ CSTAT.do takes the raw .dta at the annual and quarterly frequency for the universe of firms in Compustat and applies the selection criteria; the code generates a clean dataset at the annual and quarterly frequencies of the time-series of cross-sectional moments used in the paper for sales, employment, stock returns, and others.

- SBC_ Clean_ BvDOsiris.do takes the raw .csv at the annual frequency for the universe of firms available in the BvD Osiris dataset and applies the selection criteria; the code generates a clean dataset at the country/year frequency containing the cross-sectional moments used in the paper for sales and employment.

- SBC_ Clean_ GCSTAT.do takes the raw .dta at daily stock returns for the universe of firms in the Global Compustat database and applies the selection criteria; the code generates a clean dataset at the country/quarterly level and the time series of the cross-sectional moments of stock returns used in the paper.

- SBC_ Clean_ BvDAmadeus.do takes raw .dta at the annual frequency for the universe of firms available in the BvD Amadeus dataset for the following countries: AUT, BEL, BLR, CHE, DEU,

DNK, ESP, FIN, FRA, GBR, GRC, HUN, IRL, ISL, ITA, NLD, NOR, POL, PRT, SWE, and UKR. The code applies the selection criteria and generates a clean dataset at the country/year level containing the cross-sectional moments for sales and employment used in the appendix of the paper.

- SBC_Clean_BvDAmadeusTFP.do takes raw .dta at the annual frequency for the universe of firms available in the BvD Amadeus dataset for the following countries: DEU, DNK, ESP, FIN, FRA, GBR, GRC, HUN, IRL, ISL, ITA, NLD, NOR, POL, PRT, SWE, and UKR. The dataset contains data on value-added, capital expenditures, and labor input used to calculate firm-level TFP for selected countries. The code applies the selection criteria and generates a clean dataset at the country/year- and country/industry-level containing the cross-sectional of the TFP shocks distribution using four different estimation methods.

The do-files contain additional comments and we made the best of our efforts to make them bug-free. In case you have any questions or suggestions to improve the code, please contact us. Be aware that the clean datasets created by these do-files might be larger than the raw dataset because of the large number of variables generated. The Excel file SBC_DataVariableList.xls contains the list of variables in the raw data used in the paper following the names descriptions from WRDS for each of the datasets. Finally, the raw data can be accessed through WRDS or, for replication purposes, in the following links:

- Annual Compustat replication data here, Quarterly Compustat replication data here, and Daily Stock Price replication data here.
- BvD Osiris replication data, here.

The Global Compustat and the BvD Amadeus datasets are substantially larger than the rest of the datasets. In case you need the raw data to replicate our results, please contact Sergio Salgado.

2 Replication of Main Empirical Results

2.1 SBC Replica Empirics

The do-file *SBC_ReplicaEmpirics.do* takes the data from the Excel files and replicates the figures and tables of the paper.

2.2 Census LBD

The file *SBC_Census_LBD.xls* contains different moments of the employment growth distribution across firms and establishments. To avoid the disclosure of sensitive information, all the reported percentiles and percentiles-based measures (e.g. Kelley skewness) are calculated as the employment-weighted average within a 1% centered band around the corresponding percentile (for instance, the 90th percentile is the average across all observations between the 89th and 91st percentiles of the employment growth distribution). See Appendix A for additional details on sample selection. The excel file *Census_LBD.xls* contains the following sheets and variables,

- Firm Employment Growth Moments: Moments of the employment growth distribution at the firm level across the entire non-farm private sector. The variables in the sheet are the following,
  - Year
  - p90, p50, p10: 90th, 50th, and 10th percentiles of the distribution
  - ksk: Kelley skewness calculated as $KSK_t = \frac{P90-P50}{P90-P10} - \frac{P50-P10}{P90-P10}$
  - ksk2: alternative measure of Kelley skewness calculated as $KSK2_t = \frac{P95-P50}{P95-P5} - \frac{P50-P5}{P95-P5}$.
  - ksk3: alternative measure of Kelley skewness calculated as $KSK3_t = \frac{P97.5-P50}{P97.5-P2.5} - \frac{P50-P2.5}{P97.5-P2.5}$.
  - p9050: 90th-to-50th percentiles spread
  - p5010: 50th-to-10th percentiles spread
  - ksk AC: Kelley skewness of the employment growth distribution considering entry and exit
  - ksk 3 years: Kelley skewness of employment growth distribution of three year log change
- Firm Employment Growth Kelley: Kelley skewness calculated as $KSK_t = \frac{P90-P50}{P90-P10} - \frac{P50-P10}{P90-P10}$ at the firm level within firm size and age groups. The variables in the sheet are the following,
  - Year
  - group: population group for which the moments are calculated. The categories are
    - fize1: firms with average employment between 1 and 19 where average employment is calculated as $\overline{E}^f_{j,t} = 0.5 \times \left(E^f_{j,t} + E^f_{j,t+1}\right)$.
    - fsize2: firms with average employment between 20 and 49
    - fsize3: firms with average employment between 50 and 99
    - fsize4: firms with average employment between 100 and 499
    - fsize6: firms with average employment between 500 and 999
    - fsize7: firms with average employment of 1000 or more
    - fage1: firms of less than 5 years old
    - fage2: firms between 5 and 10 years old
    - fage3: firms more than 10 years old.[1]
  - ksk: Kelley skewness of the employment growth distribution within the relevant group.
- Estab. Employment Growth Kelley: Kelley skewness calculated as $KSK_t = \frac{P90-P50}{P90-P10} - \frac{P50-P10}{P90-P10}$ at the establishment level for entire US nonfarm private sector and within establishment size and age groups. The variables in the sheet are the following,
  - Year
  - group: population group for which the moments are calculated. The categories are
    - All: Entire US nonfarm private sector
    - eize1: establishment with average employment between 1 and 19 where average employment is calculated as $\overline{E}^e_{j,t} = 0.5 \times \left(E^e_{j,t} + E^e_{j,t+1}\right)$.

[1] Firms already present in the LBD in 1976 were not considered in any of the age groups.

- esize2: establishments with average employment between 20 and 49
- esize3: establishments with average employment between 50 and 99
- esize4: establishments with average employment between 100 and 499
- esize6: establishments with average employment between 500 and 999
- esize7: establishments with average employment of 1000 or more
- eage1: establishments of less than 5 years old
- eage2: establishments between 5 and 10 years old
- eage3: establishments of more than 10 years old.[2]

– ksk: Kelley skewness of the employment growth distribution within the relevant group.

- USA Employment Scatter: mean and Kelley skewness within different quantiles of the industry-year distribution of average employment growth in the LBD. These moments were generated using Census data and no additional replication material was disclosed.

2.3 United States and Cross Country

The *SBC_ USA_ AND_ CROSSCOUNTRYDEC2019.xls* file contains cross-sectional moments of the distribution of sales growth, employment growth, stock returns, and productivity shocks for the United States and for a cross section of countries. The excel file *SBC_ USA_ AND_ CROSSCOUNTRYDEC2019.xls* contains the following sheets:

- USA Kernel Density: estimated empirical density for recession and expansion years of the annual sales growth distribution for a sample of publicly traded firms in the United States
- USA Quarter Sales Growth: moments of the four quarters growth rate of quarterly sales for a sample of publicly traded firms in the United States
- USA Quarter Stock Returns: moments of the four quarters growth rate of end-of-quarter stock price for a sample of publicly traded firms in the United States
- USA Annual Sales Growth: moments the one-year growth rate of annual sales in the United States
- USA Annual Employment Growth: moments the one-year growth rate of annual employment in the United States
- Country Kernel Density: estimated empirical density for recession and expansion years of the annual sales growth distribution for a sample of publicly traded firms from BvD's Osiris across ~40 countries
- Country Annual Sales Growth: moments the one-year growth rate of annual sales for a sample of publicly traded firms from BvD's Osiris across ~40 countries
- Country Annual Employment Growth: moments the one-year growth rate of annual sales for a sample of publicly traded firms from BvD's Osiris across ~40 countries

[2] Establishment already present in the LBD in 1976 were not considered in any of the age groups.

- Country Stock Returns: moments the four quarters growth rate of the end-of-quarter stock price for a sample of publicly traded cross ~40 countries from Global Compustat
- Amadeus Sales Growth: moments of one-year growth rate of annual sales for a sample of firms from the BvD Amadeus dataset.
- Amadeus Employment Growth: moments of the one-year growth rate of annual employment for a sample of firms from the BvD Amadeus dataset.
- Amadeus TFP Shocks: moments of the distribution of firms' productivity shocks using different estimation methods.

The sheets contain the following variables

- group: defines the sample over which the moments are calculated. This takes the categories all (all sample) or NAICqq where qq is a two-digit industry group. This is only available for USA data
- ISO3: country ISO code
- Year
- Quarter: Only available for Sales and Stock returns for firms in the United States
- num: number of observations used to calculate the moments
- me: cross-sectional average
- sd: cross-sectional standard deviation
- sk: coefficient of skewness (third standardized central moment)
- ku: coefficient of kurtosis (fourth standardized central moment)
- p025: 2.5th percentile of the distribution
- p10: 10th percentile of the distribution
- p25: 25th percentile of the distribution
- p50: 50th percentile of the distribution
- p75: 75th percentile of the distribution
- p90: 90th percentile of the distribution
- p975: 97.5th percentile of the distribution
- p9010: 90th-to-10th percentiles spread
- p7525: 75th-to-25th percentiles spread

- p9050: 90th-to-50th percentiles spread
- p5010: 50th-to-10th percentiles spread
- ksk: Kelley skewness calculated as $KSK_t = \frac{P90-P50}{P90-P10} - \frac{P50-P10}{P90-P10}$
- cku: Crow-Sidiqqui kurtosis calculated as $CKU_t = \frac{P97.5-P2.5}{P75-P25}$

The sheet Amadeus TFP Shocks in *SBC_ USA_ AND_ CROSSCOUNTRY.xls* contains the moments of the distribution of TFP shocks at the firm-level measured using four different methods (see Appendix in the paper and the do-file SBC_Clean_BvDAmadeusTFP.do for additional details). The data is available for bins defined at the country, industry, and year, with more than 100 firm level observations (nobs1 $>$ 100). The sheet contains the following variables, all measured within a country/industry/year.

- iso3: Country iso code
- naics2: Two-digit NAIC code
- year
- nobsx: number of firm level observations used to calculate firm-level productivity; The value of x can take four different values depending of the method used to estimate productivity as described in the Appendix C of the paper; nobs1 corresponds to productivity shocks estimated using Factor Shares; nobs2 corresponds to OLS regression; nobs3 is Olley and Pakes method; nobs4 is labor productivity; nobsg corresponds the distribution of sales growth calculated as the log-change of real sales between years t and t-1.
- p10x: is the 10th percentile of the distribution of TFP shocks within a country/industry/year bin.
- p50x, p90x, mex, kskx, p9010x, p5010x and p9050x: are 50th, 90th, average, Kelley skewness, 90th to 10th percentiles differential, 50th to 10th percentiles differential, and the 90th to 50th percentile differential respectively where x follows the same classifications as in p10.

2.4 Census of Manufacturing

The file *SBC_ CENSUS_ ASM.xls* contains the quantiles of the industry-year distribution of average TFP innovations and the corresponding Kelley skewness used in Figure 6 of the paper. The underlying data used to generate this plot is not reported.

2.5 VAR

To replicate the VAR results use *SBC_ ReplicaVAR.do* that reads the excel file *SBC_ VAR.xls*. The variables present in the excel file are the following

- year
- month

- nstock: S&P500 stock market index (closing value of last trading date of the month)
- fedfunds: FRED variable FEDFUNDS
- wage: FRED variable AHETPI
- cpi: FRED variable CPIAUCSL
- hours: FRED variable AWHMAN
- empm: FRED variable PAYEMS
- ipm: FRED variable INDPRO
- p9010m: 90th-to-10th percentiles spread of monthly stock returns constructed from CRSP
- kskm: Kelley Skewness of monthly stock returns constructed from CRSP
- p9010d: 90th-to-10th percentiles spread of daily stock returns constructed from CRSP
- kskd: Kelley Skewness of monthly stock returns constructed from CRSP