Skewed Business Cycles*

Replication of Empirical Results

Sergio Salgado† Fatih Guvenen‡ Nicholas Bloom§

July 28, 2019

1 Introduction

This document describes the replication packet for the empirical results of the paper. The replication material consists of two main sets of files, the first used a set of moments already calculated from the raw data to creates the main figures of the paper. The second takes raw data and creates the cross sectional moments.

Replication of Main Empirical Results

First, to replicate the main empirical results and replicate the figures and tables of the paper, use the following files:

- *SBC_AGGREGATE.xls* which contains aggregate time-series of annual GDP growth from FRED for the United States, quarterly GDP growth from FRED for the United States, cross-country annual GDP growth from the World Development Indicators collected by the World Bank, and cross-country quarterly GDP growth from the 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_USA_AND_CROSSCOUNTRY.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

*All results based on Census data have been reviewed to ensure that no confidential information is disclosed. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the US Census Bureau. The replication packet of 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 and NBER; nbloom@stanford.edu

Compustat; Annual employment growth, sales growth, and firm-level productivty measures from the BvD's Amadeus database. Section 2.3 describes the available material in detail.

- *SBC_ReplicaEmpirics.do* takes the .xls files and replicates the main figures and empirical results of the paper.

Compustat, Global Compustat, BvD Osiris, and BvD Amadeus were accessed via the Wharton Research Data Services (WRDS).

Replication from Raw Data

Second, 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, and stock returns.
- SBC_Clean_BvDOsiris.do takes the raw .csv at the annual frequency for the universe of firms available in Bureau van Dijk's (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 used in the paper for stock returns.
- 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. The code generates a clean dataset at the country-annual frequency 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, ESP, FIN, FRA, ITA, PRT, and SWE. The dataset contains data on value added, capital expenditures, and labor input to use to calculate firm-level TFP.

The do-files are heavily commented and we made the best of our efforts to make them bug-free. In case you of any question or find an error in the code, please contact us. Be aware that the clean dataset 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 fromWRDS for each of the datasets. Finally, the underlying data is not in this packet but can be accessed through WRDS or, for replication purposes, in the following links:

- Annual Compustat replication data here and Quarterly Compustat replication data here
- BvD Osiris replication data, here

Global Compustat and 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 Do Files for Replication of 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. The do file

2.2 Census LBD

The *SBC_ Census_ LBD.xls* file 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 $\overline{E}^e_{j,t} = 0.5 \times \left(E^e_{j,t} + E^e_{j,t-1}\right)$.
    - 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.

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

2.3 United States and Cross Country

The *SBC_ USA_ AND_ CROSSCOUNTRY.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_ CROSSCOUNTRY.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
- 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 BvD's Amadeus dataset.
- Amadeus Employment Growth: moments of the one-year growth rate of annual employment for a sample of firms from BvD's Amadeus dataset.

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 C 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, year, with more than 100 firm level observations. The sheet contains the following variables, all measured within a country/industry/year.

- iso3: Country iso code
- naics2: Two-digit NAIC code
- year

- nobs: number of firm level observations used to calculate firm-level productivity
- p10x: is the 10th percentile of the distribution of TFP shocks within a country/industry/year bin. 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; p101 corresponds to productivity shocks estimated using Factor Shares; p102 corresponds to OLS regression; p103 is Olley and Pakes method; p104 is labor productivity; p10g corresponds the distribution of sales growth calculated as the log-change of real sales between years t and t-1.
- 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.