

Supplemental data for “The Nature of Firm Growth”: Autocovariance of Firm- and Establishment-level Employment

Benjamin W. Pugsley
University of Notre Dame

Petr Sedlacek
University of Oxford & CEPR

Vincent Sterk
University College London & CEPR

July 2020

Summary

These supplemental data provide the full autocovariance matrices of firm- and establishment-level residual log employment for firm and establishment ages 0 to 19, estimated from the U.S. Census Bureau Longitudinal Business Database (LBD). We provide these data for interested researchers to calibrate and estimate models of firm dynamics. If you use the moments, we would ask you to please cite Pugsley, Sedlacek and Sterk (2020). We also provide a small set of example programs that estimate the specification of the statistical model from the paper along with some alternatives.

Autocovariance matrices (./data/ folder)

This folder includes the autocovariance matrices for firms and establishments. Each is provided as a separate excel spreadsheet, which contains the autocovariance formatted as a matrix and stacked (column-wise from lower triangle) as a vector. Each also provides the unbalanced and balanced estimates used in the paper. We include a brief summary of the autocovariance measures below. For additional details on measurement, please refer to Pugsley, Sedlacek and Sterk (2020) and the online appendix.

The matrices report the cross sectional autocovariance of log employment residuals, $\log \widetilde{n_{iat}}$, obtained by regressing firm (establishment) log employment, $\log n_{iat}$ on a birth cohort fixed effect and an industry fixed effect (NAICS6) and computing residuals. Firm (establishment) employment is measured for each firm (establishment) i at each age a .

To compute the autocovariance matrix, we pool all observations from the entering cohorts from 1979 to 1993 so that we can observe ages 0 to 19 in the 2012 vintage of the LBD. After pooling the cohorts, we drop the t notation. Each cell of the covariance matrix will correspond to an age pair a and $h = a - j$ for $0 \leq a \leq 19$ and $0 \leq j \leq a$:

$$\text{Cov} [\log \widetilde{n_{ia}}, \log \widetilde{n_{ia-j}}] = E [\log \widetilde{n_{ia}} \log \widetilde{n_{ia-j}}] - E [\log \widetilde{n_{ia}}] E [\log \widetilde{n_{ia-j}}].$$

We report the covariance formula only to emphasize that the autocovariance between a particular age pair is calculated by averaging *across* all firms (establishments) observed at both ages, rather than by averaging over all ages that are j periods apart. In other words, it is a cross sectional measure of autocovariance, rather than a time-series measure that assumes the data are covariance stationary. As we discuss in the paper, the estimated (cross sectional) autocovariance reveals the data are definitively not covariance stationary.

For firms and establishments, we provide two estimates of the autocovariance matrix. The first is an “unbalanced” measure that computes each covariance cell using whatever firms are operating at those ages

in the sample. In the unbalanced autocovariance the set of firms used between a and h may be different than the set of firms used between a' and h' because of exit and spells of inactivity. The second measure is a “balanced” one. This is based on restricting the calculation to firms (establishments) that survive at least 19 years. It is not based on a strictly balanced panel. Here, the set of firms between a and h and a' and h' will be mostly the same, but there may still be some difference because of temporary spells of inactivity.

Example (./example/ folder)

This folder provides a simple Matlab example that estimates the baseline statistical model for log employment from the paper as well as a statistical model restricted to only a heterogeneous initial condition and persistent ex-post shocks, which is commonly used in the literature, c.f., Hopenhayn and Rogerson (1993). The files are set to estimate these models using the balanced sample for firms.¹ You can adjust this with the settings at the top of the script. The contents are listed below:

Main file

<code>run_example.m</code>	script that estimates the baseline and an example of a restricted model and plots their autocovariance functions against the data
----------------------------	---

Supporting files

<code>autocov_ar1_general.m</code>	return model autocovariance function as a matrix for some age range
<code>gmm_ar1_general.m</code>	return weighted squared distance between statistical model and data
<code>autocovgmm_light.m</code>	estimates the model, allows for multiple initial guesses, different solvers, etc...
<code>getmodelfit.m</code>	computes several alternative measures of model fit for the estimated model

¹The example only computes point estimates since it is not possible to release the 210 x 210 covariance matrix of the moment conditions necessary to compute standard errors. The standard errors in the paper were estimated within the Census research environment.