Productivity Estimation Methods in Stata

November 30, 2024

1 Introduction

This document provides an overview of the Stata code used to estimate firm-level productivity using different econometric methods, specifically focusing on the Olley-Pakes (OP) and Levinsohn-Petrin (LP) methods, including their implementations with the Ackerberg-Caves-Frazer (ACF) correction. The code utilizes the `prodest` package in Stata to perform these estimations. We also compare the results obtained from these methods.

2 Data Preparation: Columbia Annual Manufacturing Survey

2.1 Variables Used

The analysis uses firm-level panel data with the following variables (all monetary variables are defalted to thousands of 1982 pesos):

- `pro`: Output.
- `k`: Capital stock.
- `labor`: Number of employees.
- `mat`: Material inputs.

2.2 Generating Investment

For the OP method, investment data is required. Investment is calculated as the difference between the current capital stock and the depreciated capital stock from the previous period:

$$\texttt{inv} = \texttt{k} - (1 - \delta) \times L.\texttt{k} \tag{1}$$

where $\delta = 0.074$ is the depreciation rate (a weighted average of BEA depreciation rates for equipment and structures), and $L.\texttt{k}$ denotes the lagged capital stock.

Note: More than 25% of the calculated investment values are non-positive, which is a potential drawback for the OP method as it requires positive investment values for identification.

2.3 Creating Log Variables

We take the natural logarithm of the relevant variables to linearize the production function:

- `log_y` $= \ln($`pro`$)$
- `log_k` $= \ln($`k`$)$
- `log_l` $= \ln($`labor`$)$
- `log_m` $= \ln($`mat`$)$
- `log_inv` $= \ln($`inv`$)$

Naturally, the logarithm is undefined for non-positive values, so observations with non-positive values in any of these variables will be excluded from the analysis.

3 Productivity Estimation Methods

3.1 Olley-Pakes (OP) Estimation

3.1.1 Sample Selection and Trimming

To prepare the sample for OP estimation:

1. Create an indicator variable `op` to identify observations with non-missing log variables, including investment:

   ```
   gen op = 0
   replace op = 1 if log_y != . & log_l != . & log_k != .
   & log_m != . & log_inv != .
   ```

2. Calculate input-output ratios and the investment-capital ratio in logs:

$$\begin{aligned}\texttt{log\_k\_y} &= \texttt{log\_k} - \texttt{log\_y}\\ \texttt{log\_l\_y} &= \texttt{log\_l} - \texttt{log\_y}\\ \texttt{log\_m\_y} &= \texttt{log\_m} - \texttt{log\_y}\\ \texttt{log\_inv\_k} &= \texttt{log\_inv} - \texttt{log\_k}\end{aligned}$$

3. Winsorize these ratios at the 1st and 99th percentiles to reduce the influence of outliers:

```
gstats winsor log_k_y log_l_y log_m_y log_inv_k if op == 1,
suffix(_op_tr) trim cuts(1 99)
```

4. Update the `op` indicator to exclude observations that became missing after winsorization:

```
replace op = 0 if (log_k_y_op_tr == . | log_l_y_op_tr == . |
log_m_y_op_tr == . | log_inv_k_op_tr == .) & op == 1
```

3.1.2 Estimation

Using the `prodest` command:

```
prodest log_y if op == 1, free(log_l log_m) state(log_k) proxy(log_inv) met(op)
opt(dfp) reps(50) id(id) t(year) fsresiduals(op_fsres)
```

- `free(log_l log_m)`: Labor and materials are treated as freely variable inputs.
- `state(log_k)`: Capital is the state variable (slowly adjusting input).
- `proxy(log_inv)`: Investment is used as the proxy variable.
- `met(op)`: Specifies the Olley-Pakes estimation method.
- `opt(dfp)`: Optimization method used is the Davidon-Fletcher-Powell algorithm.
- `reps(50)`: Number of replications for bootstrapping standard errors.
- `id(id)` and `t(year)`: Panel identifiers.
- `fsresiduals(op_fsres)`: Saves first-stage residuals.

3.2 Levinsohn-Petrin (LP) Estimation

3.2.1 Sample Selection and Trimming

The process is similar to OP estimation but excludes investment:

1. Create an indicator variable `lp` for observations with non-missing log variables:

```
gen lp = 0
replace lp = 1 if log_y != . & log_l != . & log_k != .
& log_m != .
```

2. Winsorize input-output ratios:

```
gstats winsor log_k_y log_l_y log_m_y if lp == 1, suffix(_lp_tr)
trim cuts(1 99)
```

3. Update lp indicator:

```
replace lp = 0 if (log_k_y_lp_tr == . | log_l_y_lp_tr == . |
log_m_y_lp_tr == .) & lp == 1
```

3.2.2 Estimation

Using `prodest`:

```
prodest log_y if lp == 1, free(log_l) state(log_k) proxy(log_m) met(lp)
opt(dfp) reps(50) id(id) t(year) fsresiduals(lp_fsres)
```

- `free(log_l)`: Labor is treated as a freely variable input.
- `proxy(log_m)`: Materials are used as the proxy variable.
- Other options are similar to the OP estimation.

3.3 OP Estimation with ACF Correction

The Ackerberg-Caves-Frazer (ACF) correction addresses potential collinearity issues in the first stage of OP and LP methods.

```
prodest log_y if op == 1, free(log_l log_m) state(log_k) proxy(log_inv) met(op)
acf opt(dfp) reps(50) id(id) t(year) fsresiduals(op_acf_fsres)
```

- `acf`: Implements the ACF correction.

3.4 LP Estimation with ACF Correction

Due to optimization challenges with the ACF correction in LP estimation, it may be necessary to specify starting values and adjust optimization options:

```
prodest log_y if lp == 1, free(log_l) state(log_k) proxy(log_m) met(lp)
acf reps(50) id(id) t(year) init(".4,.5,.3") fsresiduals(lp_acf_fsres)
```

- `init(".4,.5,.3")`: Provides starting values for the optimization.

4 Comparison of Different Methods

To assess the consistency and differences between the TFP estimates obtained from various methods, we create binscatter plots comparing the TFP estimates pairwise:

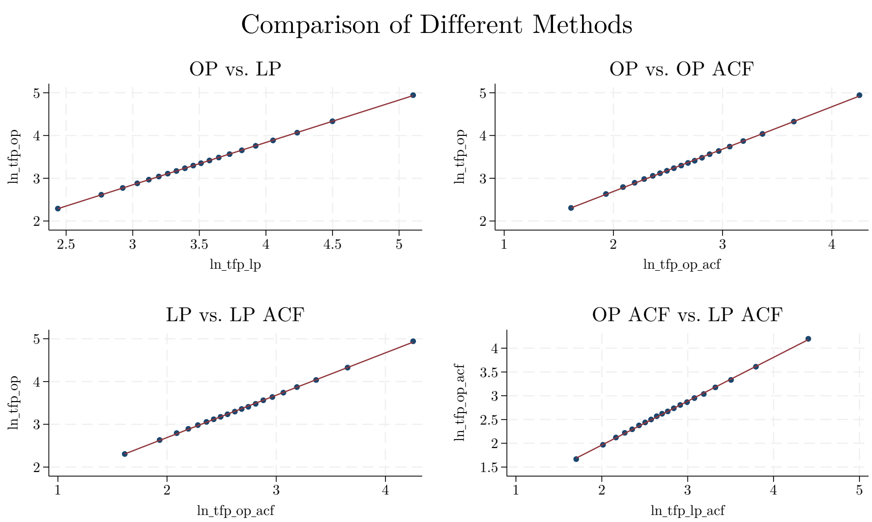

Figure 1: Comparison of TFP Estimates from Different Methods