A problem set on production function estimation

Helsinki GSE, Empirical Industrial Organization II: Topics (ECON-L1350)

In this problem set you will estimate a value added production function using the estimation procedure suggested by Levinsohn and Petrin (2003, ReStud). You are provided annual data on plants in the textiles industry in Chile. A similar dataset has been used in several productivity papers (e.g., Levinsohn and Petrin (2003, ReStud), Ackerberg, Caves and Frazer (2015, Econometrica), and Raval (2023, ReStud)). The observables are: log of value added (output), log of capital input, log of labour input, log of materials input, and lags of all of these observables. The data set is in .dta format.

The production function is a value added production function of Cobb-Douglas form:

$$y_{it} = \beta_0 + \beta_K k_{it} + \beta_L l_{it} + \omega_{it} + \eta_{it}$$

where y_{it} is the log of value added, k_{it} is the log of capital input, l_{it} is the log of labour input, ω_{it} is productivity that is observable to the firm's decision-maker at time t, and η_{it} is a shock to productivity that is unobservable to the firm's decision-maker.

1. Estimate the production function using OLS. Report the estimates and discuss how the estimates are likely to be biased, and why.

2. Use materials as the proxy variable to control for the unobservable ω_{it}. Make the same assumptions on the timing of inputs, and their dynamic nature, as LP do. Also assume that the materials input demand is strictly monotonic in ω_{it}, and defined as follows:

   $$m_{it} = f_M(k_{it}, \omega_{it})$$

   When the unobservable ω_{it} is controlled for by using the inversion of the materials demand function, the production function may be rewritten as:

   $$y_{it} = \beta_0 + \beta_K k_{it} + \beta_L l_{it} + h(k_{it}, m_{it}) + \eta_{it}$$

   Consider approximating $h(k_{it}, m_{it})$ by a polynomial and estimating the above equation by OLS. Explain why only β_L can be identified.

3. Do the first stage of LP. By approximating $h(k_{it}, m_{it})$ with a second order polynomial and combining terms, the production function can be rewritten as:

$$y_{it} = \beta_L l_{it} + \gamma_0 + \gamma_1 k_{it} + \gamma_2 m_{it} + \gamma_3 k_{it}^2 + \gamma_4 k_{it} m_{it} + \gamma_5 m_{it}^2 + \eta_{it}$$

   Estimate this equation by OLS. Report the estimate of the identified parameter. To prepare for the second stage, use your estimates to compute $\phi_{it} = \gamma_0 + \gamma_1 k_{it} + \gamma_2 m_{it} + \gamma_3 k_{it}^2 + \gamma_4 k_{it} m_{it} + \gamma_5 m_{it}^2$, which comprises all the terms with non-identified parameters. Also compute ϕ_{it-1}.

4. Parameters β_0 and β_K are identified in the second stage. This requires making the assumption that ω_{it} evolves as a first-order Markov process, i.e., ω_{it} can be decomposed as

$$\omega_{it} = E\left[\omega_{it} | \omega_{it-1}\right] + \xi_{it}$$

   where ξ_{it} is a shock to productivity. By construction, $E\left[\xi_{it} | I_{it-1}\right] = 0$. Write down two moments conditions that identify β_0 and β_K. Note: regarding the residuals, there are (at least) two ways to write the moments: using ξ_{it}, or $\xi_{it} + \eta_{it}$. You can choose either way.

5. Do the second stage of LP, which is non-linear GMM estimation, to estimate β_0 and β_K. When writing for the residual for the moment conditions:

- remeber that you've already obtained estimates of ϕ_{it}, ϕ_{it-1} (and β_L) in the first stage; use these to write for the "implied" $\widehat{\omega}_{it}\left(\beta_0, \beta_K\right)$ and $\widehat{\omega}_{it-1}\left(\beta_0, \beta_K\right)$
- you may assume that $g\left(\omega_{it-1}\right)$ takes the form of a first-order polynomial, i.e., $g\left(\omega_{it-1}\right) = \rho_0 + \rho_1 \omega_{it-1}$; this can be estimated by OLS
- also remember that for any "guess" on β_0, β_K, the implied $\widehat{\omega}_{it}$ and $\widehat{\omega}_{it-1}$ change, and you have to estimate $g\left(\widehat{\omega}_{it}\right)$

  Estimate β_0 and β_K by minimising

$$\min_{\beta_0, \beta_1} G_N\left(\beta_0, \beta_K\right)' G_N\left(\beta_0, \beta_K\right)$$

  where G_N is the sample analogue of the moment conditions you defined above:

$$G_N\left(\beta_0, \beta_K\right) = \frac{1}{N}\frac{1}{T}\sum_{it}\left[\begin{array}{c} ? \\ ? \end{array}\right]$$

Please submit your work, including the estimation code, to nelli.valmari@etla.fi. The deadline for this problem set is February 2nd, 2023.