ECON8862: Monetary Theory II, Fabio Schiantarelli

Problem Set

Learning how to Use Compustat,
Estimate Investment Models by GMM,
and Solve and Estimate Models with Fixed Adjustment Costs

Fall 2021, Boston College

Part 1 is an adapted version from S. Terry Problem set 1 for the Ec741 course at BU. Part 2 requires students to estimate investment models using dynamic panel GMM. Part 3 (written by Luigi Pollio) teaches students how to solve and estimate an investment model with fixed adjustment costs, using value function iteration and the Simulated Methods of Moments (SMM).

1 Firm-Level Data from Compustat

Email me the response to Part 1 by November 1, as a written up set of solutions and answers. Use Tex to write up your answers, and use any statistical language or package you'd like for the analysis, such as STATA. With your formatted answers, also include your code.

1.1 WRDS + NBER Preliminaries

(a) Your first task is to get access to Compustat's North America Fundamentals Annual database. Compustat annual includes standardized information from US-listed public firms' financial statements at fiscal year end, and it's an important source of US firm- level data. You will access Compustat through the Wharton Research Data Services (WRDS) site. WRDS is a business data warehouse to which BC has a subscription. Go to the website here and register for an account using your BC email. Approval can take a day or two, so sign up immediately. Do not wait until some future date before the problem set is due. Once you're set up, Compustat documentation is readily accessible via Google or the WRDS website.

(b) Your next task is to download the NBER-CES manufacturing database here. This database, compiled in a consistently measured fashion, contains data on cost shares, prices deflators, output, etc. for individual manufacturing industries in the US. Down- load the 1987 SIC version. You can use whatever file format you'd like, e.g., STATA or CSV. For each SIC-4 digit industry and year, compute and/or save the following information:

- *labshare*: the labor share in value added, using a 10-year moving average of payroll costs relative to value added
- *capshare*: the capital share, as the residual of the implied labor share
- *vaddfrac*: a 10-year moving average of the ratio of value added to gross output (shipments)
- *piship*: the shipments price deflator
- *piinv*: the investment price deflator

The result should be a panel dataset by sic X year with each of the variables above

1.2 Download, Clean and Merge Compustat

In this step, you'll download the Compustat dataset itself. To select the right dataset, first login to WRDS. Then navigate to "Compustat-Capital IQ," then "Compustat Monthly Updates - North America," then "Fundamentals Annual." At that point, make the following selections:

(a) Select 1980-01 to 2018-12 for "Step 1. Choose your date range."

(b) Select "Search the entire database" to download the full dataset under "Step 2. Apply your company codes."

(c) Under "Screening Variables," uncheck "FS" from "Industry Format," which will avoid duplication of certain observations

(d) Under "Screening Variables," uncheck "CAD" from "Currency," which will avoid duplication of certain observations

(e) Select "All" under "Step 3. Choose variable types" in order to download all available data. This will result in a large file, but you'll have more information to play with later in Part 2.

(f) Choose your output format under "Step 4. Select query output." I prefer to download files in STATA format with .zip compression, although you can use whatever approach you like.

(g) WRDS will construct your dataset and notify you when it's available for download.

Now that you have the Compustat database in hand, read the data in and perform the following basic cleaning exercises:

(h) Keep only US-based firms, i.e., keep if *fic*="USA".

(i) Keep only final versions of statements, i.e., keep if *final*="Y".

(j) To avoid firms with strange production functions, drop regulated utilities and financial companies, i.e., drop if the 4-digit sic code is in the range [4900,5000) or [6000,7000).

(k) Get rid of years with extremely large values for acquisitions to avoid the influence of large mergers, i.e., drop if acquisitions *aqc* is greater than 5% of assets *at*.

(l) Since Compustat records end-of-year capital values, shift the reported book value *ppent* forwards one year, i.e. set $ppent_{it} = ppent_{i,t-1}$.

(m) Keep only if the book value of assets at, book value of capital *ppent*, number of em- ployees *emp*, capital investment *capxv*, and revenues *sale* are nonmissing and positive.

(n) Keep only if the firm exists in the data for greater than or equal to two years.

(o) Merge in the NBER-CES manufacturing data by *sic x year*, where you can take the Compustat fiscal year *fyear* = NBER-CES year. Keep only the matched observations.

The result of this exercise should be an unbalanced panel of manufacturing firm-years, with tens of thousands of observations. The exact number of observations may differ based on when you download the Compustat data, which is continually updated.

1.3 Compute Some Simple Series of Interest

Now,you will compute some crude versions of economically interesting series.

(a) *investment*: Set the investment series equal to capital expenditures *capxv* deflated by the investment price *piinv*.

(b) *capital*: Set the capital series equal to the book value of capital *ppent* deflated by the investment price *piinv*. How does this approach for computing capital differ from, and when is it likely to be incorrect relative to, the perpetual inventory method?

(c) *irate*: Set the investment rate equal to the ratio of investment and capital.

(d) *outputva*: Set the value added output series equal to sale times *vaddfrac*, deflated by the shipments deflator *piship*. Under what assumptions on the production function does this approach make sense?

(e) *outputgrowth*, *empgrowth*: Compute output and employment growth using the Davis-Haltiwanger formulas.

(f) *tfpr*: At this point you have value added, capital, and labor. Use the industry-year specific labor and capital cost shares from the NBER-CES database to compute the log of revenue TFP based on these two inputs. You should not need to estimate any production functions here. Under what assumptions does such a cost share-based TFP measure make sense?

(g) *tfprgrowth*: Compute the growth rate, i.e., the log difference, of TFP.

(h) *operating income rate* : operating income before depreciation/*ppent*

(i) *cash flow rate*: [operating income before depreciation - net interest payments - taxes + non operating income +special items (or net income before extraordinary items + depreciation)]/*at*.

(j) *output*, (another measure of output): Set the output series equal to sales deflated by the shipments deflator *piship*.

(k) *output_cap_r*: (output capital ratio) *output/capital*.

At this point, if you compute a histogram or moments of any of these variables, you'll immediately see that plenty of outliers dominate mean calculations. You'll also see that there are big, persistent, differences in scaling across firms of different sizes. Your next step will deal with both of these challenges to inference:

(l) Drop firms which are in the panel for less than 5 years.

(m) Winsorize the top and the bottom 5% of the variables *empgrowth, outputgrowth, tfprgrowth, operating income rate, cash flow rate* while for the variable *irate* winsorize only the top 5%. Hint: in Stata you can do it by using the command winsor2 (downloaded the module *winsor2*): winsor2 *var1 var2*, cuts(*5 95*). It creates two new variables with suffix *_w* for which values below the 5st or above the 95th percentile are replaced by such percentiles.

(n) Plot histograms of each of the cleaned series from subpart (m), pooled over all years. Label your figures and all axes.

(o) Report the mean and standard deviation of each of the cleaned series from subpart (l), pooled over all years. Interpret units in all cases.

2 Estimate Q Models and Euler Equations for the Capital Stock by panel GMM

Part 2 requires you to estimate a few investment models. I will announce later when this is due.

(a) First estimate a Q-model of investment using variants of equation (16) in lecture 11. Define Q as the sum of the market value of shares plus debt divided by the value of capital stock (If you want to be fancy subtract from the numerator the value of inventories and current assets. Be careful that the timing is right) and drop the outliers for Q following the procedure outlined above. Use the beginning of period Q as the regressor (not Q at time t+1, as we have in the lecture handout). Then, augment the Q-model with a cash flow variable. We suggest you use the lagged cash flow rate.

(b) Estimate the augmented Q-model using the within estimator (Least Square Dummy Variables estimator). Always include common year effect as well. Then estimate the model by GMM using xtabond2 in Stata (allowing for common year effects).

(c) Try both the GMM difference estimator and the system estimator and discuss the relative merits. Try both the one step and two step estimator and discuss the relative merits. When using two step estimators, compare the Windmeijer corrected with the uncorrected standard errors. I suggest you use (in the difference model) lags 2-4 or 3-4 as instruments. In all cases calculate the test of over-identifying restrictions (Hansen-Sargan test)

(d) Comment on sign, magnitude and significance of the various coefficients, as well as on the test for serial correlation of the residuals in difference (Arellano-Bond Test), as well as on the test of over-identifying restrictions (Hansen-Sargan Test).

(e) Divide the sample into two types of firms based on whether employment *emp* is lower or greater than 250. Re-estimate the Q-model allowing the coefficients to be different for small and large firms. We suggest you to create a dummy that takes value 1 if a firm in period t has lower that 250 employees and 0 otherwise and estimate the heterogeneous effect by interacting the lagged dummy for size with both lagged cash rate and lagged Q. Interpret the results. What can you learn from these regressions (if anything) about the role of financing constraints and under which assumptions?

(f) Estimate the Euler equations for capital in the parametrization of the lecture notes. See equation (18) in lecture (11). First assume perfect competition, then allow for imperfect competition. Use lecture (11) to figure how you can allow for imperfect competition). Assume a common mark-up across firms for simplicity. Estimate by GMM only. Repeat what you have done for the Q-model in points (b) to (c). If you want to try to do (e) as well first talk to us on how to define the instruments.

3 Model solution

This part is optional and it is not due in class. Consider a standard investment model with adjustment cost that contain a quadratic and fix component, stochastic productivity and linear equity cost issuance. Each firm solve the following dynamic profit maximization problem subject to a set of constraints

$$V(z,k) = \max_{k'} \left[\Pi(k,z) - AC\left(z,k,k'\right) - \eta\left(d\left(k,k',z\right)\right) + \frac{1}{1+r}\mathbb{E}\left(V\left(z',k'\right) \mid z\right) \right] \tag{1}$$

$$\text{s.t. } \Pi(k,z) = zk^{\alpha} - pk' + p(1-\delta)k \tag{2}$$

$$AC\left(z,k,k'\right) = \frac{\gamma_c}{2}\left(\frac{k'-(1-\delta)k}{k}\right)^2 k + \gamma_f y \mathbb{I}\left(k' \neq (1-\delta)k\right) \tag{3}$$

$$\eta\left(d\left(k,k',z\right)\right) = \left(\eta_0 + \eta_1 |d\left(k,k',z\right)|\right)\mathbb{I}_{d<0} \tag{4}$$

$$d\left(k,k',z\right) = \Pi(k,z) - AC\left(z,k,k'\right) \tag{5}$$

$$\log z = \rho \log z_{-1} + \sigma\varepsilon, \quad \varepsilon \sim N(0,1) \tag{6}$$

where $\Pi(k,z)$ are the firm profits net of the cost of purchasing investment goods, $AC\left(z,k,k'\right)$ are the adjustment costs, $d\left(k,k',z\right)$ are the distributions to the shareholders. When positive it means that firms are paying positive dividends to the shareholders while, when negative it corresponds to new equity issuance. $\eta(d\left(k,k',z\right))$ are the equity cost issuance which occur when d is negative.

(a) Above $\alpha, p, \delta, r, p, \sigma, \eta_0, \eta_1, \gamma_c,$ and γ_f are constants.Provide an economic interpretation for each, along with reasonable restrictions on their values. Study the 3 cases: (i) $\gamma_c = 0$; (ii) $\gamma_f = 0$; (iii) both equal to 0. What happen to the model? Do the same for η_0 and η_1.

(b) Calibrate the model as follows: $\alpha = 0.7$, $p = 1.2$, $\delta = 0.1$, $r = 0.04$, $\rho = 0.8$, $\sigma = 0.2$, $\gamma_f = 0$, $\gamma_c = 0.05$, $\eta_0 = 0.08$, $\eta_1 = 0.028$, what is this case of study?

(c) Convert the continuous exogenous process for z above into the discretized Markov chain using Tauchen (1986). This function is gonna give a discretization for log(z) when in the model profits depend on z. Choose the number of grid points, $N_z = 11$ We can use my function. *tauchen.m* to get this work done or you can code the program for yourself. In any case, the output is a vector of grid points and a transition probability matrix of size 11x11.

(d) Discretize the continuous endogenous state variable k into the discrete grid K. Choose $N_k = 100$. Then, choose a minimum value $\underline{k}$ and a maximum value $\bar{k}$ as grid boundary. Allocate the other points using a log-linear spacing grid, where $k_{j+1} = k_j(1-\delta)$ for all $j = 2, ..., N_k$.

Once we have setup the grids, we can use (VFI) to find a solution.

(e) Solve the Bellman equation using value function iteration (VFI) following the steps below.

i) Set s = 1 and error tolerance ϵ^{tol} = 1e-05. Start with an initial guess for the value function $V(0) = 0$. The value function is gonna be a matrix of size N_z x N_k.

(ii) Given V (s), we construct the RHS of the Bellman equation for each combination k_i, z_j. Because the capital tomorrow is defined over the grid of capital, we compute a vector of length N_k of utility from the full set of candidate policies in K. We then, set $V^{s+1}(z,k)$ equal to the maximum value of this vector and $k^{*(s+1)}$ (z , k) equal to the arg max of this vector.

iii) We compute the error of the s-th iteration of VFI, as the maximum absolute difference $\varepsilon^{(s)} = \max_{i,j} \left| \tilde{V}^{(s+1)}(z_i, k_j) - \tilde{V}^{(s)}(z_i, k_j) \right|$.

(iv) If $\varepsilon^{(s)} < \varepsilon_{tol}$ then exit. Otherwise, we set s = s + 1 and return to the top.

The output of this solution algorithm will be a policy function k^* as well as a value function V , both N_z x N_k matrices, such that the Bellman equation holds approximately.
Note: remember to save the grid indexes for capital and firm value when find the optimal policy function for next period capital. This is going to be important for simulating the model.

(f) Plot in two separate graphs the value function and the policy function given that the technology is at the steady state value and check that the policy function for capital intersect a 45 degree line.

4 The ergodic distribution of the model

This model features an ergodic or stationary or steady state distribution $\mu(z,k)$ of firms over the idiosyncratic states. In discretized form $\mu(z,k)$ is a N_z x N_k matrix storing weight at each grid point. We solve for the ergodic distribution $\tilde{\mu}$ in the discretized model, using the algorithm in Young (2010) called "non-stochastic simulation". Instead of simulating individual firms, this approach pushes mass forward across "periods" s repeatedly using only the policy functions and exogenous transitions, until the observed distribution converges.

g) We solve for the ergodic distribution $\tilde{\mu}$ following these steps:

i) Set an error tolerance $\epsilon_{tol} = 10^{-7}$ as well as a guess $\tilde{\mu}(0)$, which is a matrix of distributional weights which all add up to 1, i.e.

$$\sum_{i,j} \tilde{\mu}^{(0)}(z_i, k_j) = 1$$

ii) Initialize $\tilde{\mu}^{(s+1)} = 0$. For each state value (z,k) with $\mu^{(s)}(z,k) > 0$, extract the value of capital next period, $k^*(z,k)$ from the policy function obtained via VFI. Then, set

$$\tilde{\mu}^{(s+1)}(z_l, k^*(z_i, k_j)) := \tilde{\mu}^{(s+1)}(z_l, k^*(z_i, k_j)) + \tilde{\mu}^{(s)}(z_i, k_j) * \Pi^z_{i,l}, \quad l = 1, \ldots, N_z$$

In plain language, take the mass from the grid point today, and push it forward into tomorrow's distribution given the policy function and Markov transition matrix.

iii) Compute the distributional error $\epsilon = \max ||\tilde{\mu}^{(s)}(z,k) - \tilde{\mu}^{(s+1)}(z,k)||$. If $\epsilon < \epsilon_{tol}$, then exit. Otherwise, set $s = s + 1$ and return to the top.

The output $\tilde{\mu}$ of point g) is an approximation of the ergodic distribution of the model up to an arbitrary small error. With the policy function k^* and the ergodic distribution $\tilde{\mu}$ in hand, lets compute some cross-sectional moments.

(h) Plot the distribution $\tilde{\mu}$ over capital k at the steady state and ± 1 standard deviation values of idiosyncratic productivity z.

(i) Compute and report in a table the following model moments: (i) average investment rate $E(\frac{I}{K})$; (ii) volatility of investment $\text{Var}\,(I)$; (iii) average capital return $E(\frac{D}{K})$.

5 Estimation

In this part you have to simulate the model and use SMM to estimate the parameters.

(a) Simulate the economy for 5000 periods using the discrete grids. In order to have the same result, set the random number generator to 1, *rng(1)* in Matlab.

(b) Generate a vector of 5000 draws from the grid of technology using the transition probabilities matrix to dictate the probability of moving from one state to the other. To initialize the code, assume that at time 0 the technology is at its steady state value.

(c) Given the draws for technology, obtain the optimal path for capital, investment and profits.

Point (a)-(c) obtains a unique draw for the observable that we can use to compute moments. We need many of these simulations.

(d) Repeat (a)-(c) 3000 times to have a set of draws from the model. For each simulation compute the moments to match. We ends up with a matrix 3000 x nmom. Compute the average vector moments across simulations $\hat{m}(\tilde{x} \mid \theta)$.

(e) Create a function that takes the parameters to estimate and return the average moments for each simulation.

(f) Estimate the parameters $(\gamma_c, \alpha, \delta)$ that minimize the distance between the model moments computed in (g) and the equivalent data moments that you compute on Compustat.

$$\hat{\theta}_{SMM} = \theta : \quad \min_{\theta} \Big(\frac{\hat{m}(\tilde{x} \mid \theta) - m(x)}{m(x)}\Big) W \Big(\frac{\hat{m}(\tilde{x} \mid \theta) - m(x)}{m(x)}\Big)'$$

Assume for simplicity that the weighting matrix W is equal to the identity matrix I.

The value of θ that solve the problem in f) is the vector of estimated parameters using SMM.