Economics of the Labor Force
HW I: Simulated Methods of Moments
Due Date: 1403/8/9

Suppose the true data generating process for a series $\{x_t\}_{t=1}^T$ is given by the following AR(1) model:

$$x_t = \rho_0 x_{t-1} + \epsilon_t \tag{1}$$

where $\epsilon_t \sim N(0, \sigma_0^2)$, $\rho_0 = 0.5$, $\sigma_0 = 1$, $x_0 = 0$, and $T = 200$. Let $b_0 = (\rho_0, \sigma_0^2)$.

We will take the model generation process to be:

$$y_t(b) = \rho y_{t-1}(b) + e_t, \quad e_t \overset{i.i.d.}{\sim} N(0, \sigma^2) \tag{2}$$

where $b = (\rho, \sigma^2)$.

For any series z_t (either true data or model generated pseudodata), we can define the n vector-valued function m_n used to construct moments. For instance, when $n = 3$ we will construct the mean, the variance, and the first-order autocovariance using:

$$m_3(z_t) = \begin{bmatrix} z_t \\ (z_t - \bar{z})^2 \\ (z_t - \bar{z})(z_{t-1} - \bar{z}) \end{bmatrix} \tag{3}$$

We are interested in estimating the $\ell = 2$ parameter vector b using SMM. Let the objective function be given by:

$$J_{TH}(b) \equiv [M_T(x) - M_{TH}(y(b))]' W [M_T(x) - M_{TH}(y(b))] \tag{4}$$

where $g_{TH}(b) \equiv M_T(x) - M_{TH}(y(b))$ is an n vector with the distance between data moments and model moments, W is a positive semi-definite weighting matrix, and the number of simulations is denoted H. In the last expression, $M_T(x) = \frac{1}{T}\sum_{t=1}^T m_n(x_t)$ is the vector of empirical moments based on the true data $\{x_t\}_{t=1}^T$, and $M_{TH}(y(b)) = \frac{1}{TH}\sum_{t=1}^T \sum_{h=1}^H m_n(y_t^h(b))$ is based on the simulated data. For example, when $m_1(x_t) = x_t$, we simply match the sample average.

The SMM estimate $\hat{b}_{TH}$ is then obtained from:

$$\hat{b}_{TH} = \arg\min_b J_{TH}(b) \tag{5}$$

To obtain a consistent estimate of b, we can use $W = I$. However, to find the efficient estimator, we need the optimal weighting matrix W_{TH}^*. One possible estimate of the asymptotic variance-covariance matrix S, $\hat{S}_{TH}$, can be obtained from the simulated data at $b = \hat{b}_{1TH}$ using the estimator[1]

[1] Note that even though this estimator is consistent, if there is serial correlation or heteroskedasticity in the moment conditions (which is common in time series models), we typically use an estimator proposed by Newey and West (1987)

$$\hat{S}_{TH} = \frac{1}{TH} \sum_{h=1}^{H} \sum_{t=1}^{T} [M_T(x) - M_{TH}(y(\hat{b}_{1TH}))][M_T(x) - M_{TH}(y(\hat{b}_{1TH}))]' \tag{6}$$

The estimate of the variance-covariance matrix of the estimator $\hat{b}_{2TH}$ can be obtained using $\hat{S}_{TH}$ and $\nabla_b g_T(\hat{b}_{2TH})$. In general, we cannot obtain this derivative analytically, so we will use an approximation for $\nabla_b g_T(\hat{b}_{2TH})$. In the current case, we have two parameters (ρ, σ), so we have to compute the derivative for both dimensions. First, we can compute $M_{TH}(y(\hat{b}_{1TH}))$, then compute $M_{TH}(y(\hat{b}_{2TH} - s_\rho))$ where $s_\rho = [s, 0]'$ and s is a small number. Then take the difference, and divide by the step size s to get the $n \times 1$ vector:

$$\frac{\partial g_T(\hat{b}_{2TH})}{\partial \rho} \approx -\frac{M_{TH}(y(\hat{b}_{2TH})) - M_{TH}(y(\hat{b}_{2TH} - s_\rho))}{s} \tag{7}$$

We can compute $\frac{\partial g_T(\hat{b}_{2TH})}{\partial \sigma}$ in a similar way, using $s_\sigma = [0, s]'$. Then using (5) and its analogue, we can form the $n \times \ell$ matrix as:

$$\nabla_b g_T(\hat{b}_{2TH}) = \begin{bmatrix} \frac{\partial g_T(\hat{b}_{2TH})}{\partial \rho} & \frac{\partial g_T(\hat{b}_{2TH})}{\partial \sigma} \end{bmatrix} \tag{8}$$

Next, we can obtain the variance-covariance matrix of $\hat{b}_{2TH}$ by computing:

$$\frac{1}{T} \left[\nabla_b g_T(\hat{b}_{2TH})' \hat{S}_{TH}^{-1} \nabla_b g_T(\hat{b}_{2TH}) \right]^{-1} \tag{9}$$

Finally, standard errors can be obtained from the square root of the diagonal of the variance-covariance matrix (an $\ell \times 1$ vector):

$$\sqrt{\text{diag}\left(\frac{1}{T} \left[\nabla_b g_T(\hat{b}_{2TH})' \hat{S}_{TH}^{-1} \nabla_b g_T(\hat{b}_{2TH}) \right]^{-1} \right)} \tag{10}$$

1. Derive the following asymptotic moments associated with $m_3(x)$: mean, variance, first-order autocorrelation. Furthermore, compute $\nabla_b g(b_0)$. Which moments are informative for estimating b?

2. Simulate a series of "true" data of length $T = 200$ using (1). We will use this to compute $M_T(x)$.

3. Set $H = 10$ and simulate H vectors of length $T = 200$ random variables e_t from $N(0, 1)$. We will use this to compute $M_{TH}(y(b))$. Store these vectors. You will use the same vector of random variables throughout the entire exercise. Since this exercise requires you to estimate σ^2, you want to change the variance of e_t during the estimation. You can simply use $\sigma(e_t)$ when the variance is σ^2.

4. We will start by estimating the $\ell = 2$ vector b for the just identified case where m_2 uses mean and variance. Given what you found in the first part, do you think there will be a problem? Of course, in general we would not know whether this case would be a problem, so hopefully the standard error of the estimate of b will tell us something. Let's see.

   (a) Set $W = I$ and graph in three dimensions, the objective function (4) over $\rho \in [0.35, 0.65]$ and $\sigma \in [0.8, 1.2]$. Obtain an estimate of b by using $W = I$ in (5) and utilizing an optimizer. Report $\hat{b}_{1TH}$.

for more efficiency.

(b) Experiment with different parameters for your optimizer. How do these changes influence the computing time and the outcomes? Which parameters have the greatest impact?

(c) Obtain an estimate of $\hat{W}^*_{TH} = \hat{S}^{-1}_{TH}$ using the estimate of $\hat{b}_{1TH}$, Calculate an estimate of $\hat{b}_{2TH}$. Report $\hat{b}_{2TH}$.

(d) To obtain standard errors, compute numerically $\nabla_b g_T(\hat{b}_{2TH})$ defined in (8). Report the values of $\nabla_b g_T(\hat{b}_{2TH})$. Next, obtain the $\ell \times \ell$ variance-covariance matrix of $\hat{b}_{2TH}$ as in (9). Finally, what are the standard errors defined in (10)? How can we use the information on $\nabla_b g_T(\hat{b}_{2TH})$ to think about local identification?

(e) Experiment with different methods for calculating numerical derivatives. How do these changes affect the computing time and results? Which method do you consider most suitable in this context?

5. Next, we estimate the $\ell = 2$ vector b for the just identified case where m_2 uses the variance and autocorrelation. Given what you found in part (1), do you now think there will be a problem? If not, hopefully the standard error of the estimate of b will tell us something. Let's see. For this case, perform steps (a), (c), and (d) above.

6. Next, we will consider the overidentified case where m_3 uses the mean, variance, and autocorrelation. Let's see. For this case, perform steps (a), (c), and (d) above. Furthermore, bootstrap[2] the finite sample distribution of the estimators using the following algorithm:

   i. Draw ϵ_t and e^h_t from $N(0,1)$ for $t = 1, 2, \ldots, T$ and $h = 1, 2, \ldots, H$. Compute $(\hat{b}_{1TH}, \hat{b}_{2TH})$ as described.

   ii. Repeat (i) using another seed.

   Every time you do step (i), the seed needs to change. Otherwise, you will keep getting the same estimators.

[2] Bootstrapping is a resampling technique used to estimate the distribution of a statistic by repeatedly drawing samples from the data with replacement. The goal of bootstrapping is to assess the variability of a statistic without making strong assumptions about the underlying population distribution or relying on asymptotic distributions which are less informative in small samples.