ECON880, Computational Economics
Fall 2024, University of Wisconsin
Instructor: Dean Corbae

Problem Set #4- Due Date 10/9/24

But that date includes an additional indirect inference problem TBD

This problem set is based on Section 3 of Michaelides and Ng (2000, Journal of Econometrics) used to assess the properties of a Simulated Methods of Moments (SMM) estimator, as well as indirect inference and efficient method of moments estimators. A simple statement of the three estimators is given in Section 2 of that paper. As stated on p. 237, "With the SMM, the practitioner only needs to specify the empirical moments and is the easiest to implement". While that section provides an example estimating the $\ell = 1$ parameter of an $MA(1)$, which was presented in class, here we will consider estimating $\ell = 2$ parameters of an $AR(1)$.

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

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

where $\varepsilon_t \sim N(0, \sigma_0^2)$, $\rho_0 = 0.5,$, $\sigma_0 = 1$, $x_0 = 0$, 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 first order autocovariance using:

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

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

$$J_{TH}(b) \equiv [M_T(x) - M_{TH}(y(b))]' W [M_T(x) - M_{TH}(y(b))] \tag{3}$$

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 $\widehat{b}_{TH}$ is then obtained from:

$$\widehat{b}_{TH} = \arg\min_b J_{TH}(b) \tag{4}$$

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 $b = \hat{b}^1_{TH}$ using the estimator proposed by Newey and West (1987). More specifically,

$$\hat{S}_{y,TH} = \hat{\Gamma}_{0,TH} + \sum_{j=1}^{i(T)}\left(1 - \left[\frac{j}{i(T)+1}\right]\right)\left(\hat{\Gamma}_{j,TH} + \hat{\Gamma}'_{j,TH}\right)$$
$$\hat{S}_{TH} = \left(1 + \frac{1}{H}\right)\hat{S}_{y,TH}.$$

with

$$\hat{\Gamma}_{j,TH} = \frac{1}{TH}\sum_{h=1}^H\sum_{t=j+1}^T [m(y_t^h(\hat{b}^1_{TH})) - M_{TH}(y(\hat{b}^1_{TH}))][m(y_{t-j}^h(\hat{b}^1_{TH})) - M_{TH}(y(\hat{b}^1_{TH}))]'$$

where $i(T)$ is the lag length.

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

$$\frac{\partial g_T(\hat{b}^2_{TH})}{\partial \rho} \approx -\frac{M_{TH}(y(\hat{b}^2_{TH})) - M_{TH}(y(\hat{b}^2_{TH} - \mathbf{s}_\rho))}{s}. \tag{5}$$

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

$$\nabla_b g_T(\hat{b}^2_{TH}) = \left[\frac{\partial g_T(\hat{b}^2_{TH})}{\partial \rho} \quad \frac{\partial g_T(\hat{b}^2_{TH})}{\partial \sigma}\right]. \tag{6}$$

Note that if there was no sample error, then we could use the rank condition on (6) to understand Local Identification. Next, we can obtain the variance-covariance matrix of $\widehat{b}^2_{TH}$ by computing:

$$\frac{1}{T}\left[\nabla_b g_T(\hat{b}^2_{TH})'\hat{S}^{-1}_{TH}\nabla_b g_T(\hat{b}^2_{TH})\right]^{-1}. \tag{7}$$

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

$$\sqrt{diag\left(\frac{1}{T}\left[\nabla_b g_T(\hat{b}^2_{TH})' \hat{S}^{-1}_{TH} \nabla_b g_T(\hat{b}^2_{TH})\right]^{-1}\right)} \tag{8}$$

1. Derive the following asymptotic moments associated with $m_3(x)$: mean, variance, first order autocorrelation. Furthermore, compute $\nabla_b g(b_0)$ and compute the AGS statistics. 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 σe_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 part (1), 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 as well as the J test will tell us something. Let's see.

   (a) Set $W = I$ and graph in three dimensions, the objective function (3) over $\rho \in [0.35, 0.65]$ and $\sigma \in [0.8, 1.2]$. Obtain an estimate of b by using $W = I$ in (4) using fminsearch. Report $\widehat{b}^1_{TH}$.

   (b) Set $i(T) = 4$. Obtain an estimate of W^*. Using $\widehat{W^*}_{TH} = \widehat{S}^{-1}_{TH}$ in (4), obtain an estimate of $\hat{b}^2_{TH}$. Report $\hat{b}^2_{TH}$.

   (c) To obtain standard errors, compute numerically $\nabla_b g_T(\hat{b}^2_{TH})$ defined in (6). Report the values of $\nabla_b g_T(\hat{b}^2_{TH})$. Next, obtain the $\ell \times \ell$ variance-covariance matrix of $\widehat{b}^2_{TH}$ as in (7). Finally, what are the standard errors defined in (8)? Further compute the AGS statistics. How can we use the information on $\nabla_b g_T(\hat{b}^2_{TH})$ to think about local identification?

   (d) Since we are in the just identified case, the J test should be zero (on a computer this may be not be exact). However, given the identification issues in this particular case where we use mean and variance, the J test may not be zero. Compute the value of the J test:

$$T\frac{H}{1+H} \times J_{TH}(\hat{b}^2_{TH}) \rightarrow \chi^2$$

noting that in this just identified case $n-\ell = 0$ degrees of freedom recognizing that there really is not distribution.

5. Next we estimating 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 as well as the J test will tell us something. Let's see. For this case, perform steps (a)-(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)-(d) above. Furthermore, bootstrap the 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 $(\hat{b}^1_{TH}, \hat{b}^2_{TH})$ as described.

   ii Repeat (e) using another seed.

   Every time you do step (i), the seed needs to change (which is done automatically by matlab if you don't specify it). Otherwise you will keep getting the same estimators.

7. Recall a moving average model of order N (MA(N)) is defined as

$$x_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_N\varepsilon_{t-N}, \quad \varepsilon_t \overset{i.i.d.}{\sim} N(0,\sigma^2)$$

   Also, recall the fact that an AR(1) process can be represented as an MA(∞) model with coefficients $\theta_j = \rho^j$. Motivated by this fact, perform an indirect inference procedure to estimate the "structural" model of an AR(1) using an MA(N) model by doing the following:

   i. Write a procedure to estimate an MA(N) model by minimizing the squared prediction errors.

   ii. Estimate the MA(N) auxiliary model on the "true" data from the AR(1) model to get $\hat{\theta} = \{\hat{\theta}_j\}_{j=1}^N$. Also estimate the sample standard deviation $\hat{s} = \sqrt{\frac{1}{T-1}\sum_{i=1}^T (x_i = \bar{x})^2}$ in order to identify σ.

   iii. Given a guess of the structural parameters (ρ, σ), simulate data from the structural model (AR(1)).

iv. For each simulation, estimate the auxiliary MA(N) model and take averages across the H simulations to get $\overline{\theta}$ and the sample standard deviation $\overline{s}$.

v. Repeat steps iii. and iv. across (ρ, σ) to get an estimate by

$$\min_{\rho,\sigma}(\overline{\theta}(\rho,\sigma) - \hat{\theta})'(\overline{\theta}(\rho,\sigma) - \hat{\theta}) + (\overline{s}(\rho,\sigma) - \hat{s})^2$$

Be sure to keep your ε shocks fixed across guesses of (ρ, σ).

Repeat this procedure for $N = 1, 2, 3$. Report the data moments $(\hat{\theta}, \hat{s})$ for each of the $MA(N)$ parameters as well as the estimates of (ρ, σ) from the indirect inference. Is the MA(N) auxiliary model misspecified? How do the indirect inference estimates perform?