SMM Example

This version of the notes is based on Michaelides and Ng (2000). We will take the true data generation process to be a $k = 1$ MA(1) process

$$x_t = \varepsilon_t - b_0 \varepsilon_{t-1}, \varepsilon_t \overset{i.i.d}{\sim} N(0,1) \tag{1}$$

with $\ell = 1$ parameter $b_0 = 0.5$ and $\varepsilon_0 = 0$.

We will take the model generation process to be

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

with $\ell = 1$ parameter b and $e_0 = 0$. We do not know the true parameter value b_0 so will estimate it via simulated method of moments.

Let m denote the mapping from some $k \times 1$ vector z_t (which could be true data or simulated data) to an $n \times 1$ moment vector. Here we take $k = 1$ and consider $n = 4$ moments: mean, variance, first order autocorrelation, and second order autocorrelation given by:

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

1 Asymptotics

Note that we can write the population (unconditional) moment vector for the true data and the model using m as $\mu(x) = E[m(x)]$ and $\mu(y(b)) = E[m(y(b))]$.

For this particular $n = 4$ mapping we know the population data moments

$$\begin{aligned}
\mu(x) &= \begin{bmatrix} E\left[\varepsilon_t\right] - b_0 E\left[\varepsilon_{t-1}\right] \\ E\left[(\varepsilon_t - b_0\varepsilon_{t-1})^2\right] = E\left[\varepsilon_t^2\right] - 2b_0 E\left[\varepsilon_t \varepsilon_{t-1}\right] + b_0^2 E[\varepsilon_{t-1}^2] \\ E\left[(\varepsilon_t - b_0\varepsilon_{t-1})(\varepsilon_{t-1} - b_0\varepsilon_{t-2})\right] = E\left[\varepsilon_t\varepsilon_{t-1}\right] - b_0 E\left[\varepsilon_t\varepsilon_{t-2}\right] - b_0 E[\varepsilon_{t-1}^2] + b_0^2 E[\varepsilon_{t-1}\varepsilon_{t-2}] \\ E\left[(\varepsilon_t - b_0\varepsilon_{t-1})(\varepsilon_{t-2} - b_0\varepsilon_{t-3})\right] = E\left[\varepsilon_t\varepsilon_{t-2}\right] - b_0 E\left[\varepsilon_t\varepsilon_{t-3}\right] - b_0 E[\varepsilon_{t-1}\varepsilon_{t-2}] + b_0^2 E[\varepsilon_{t-1}\varepsilon_{t-3}] \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 1 + b_0^2 \\ -b_0 \\ 0 \end{bmatrix}
\end{aligned} \tag{4}$$

and the population model moments

$$\mu(y(b)) = \begin{bmatrix} 0 \\ 1 + b^2 \\ -b \\ 0 \end{bmatrix}. \tag{5}$$

The n moment conditions we are going to use in SMM is given by

$$u(x_t, b) = m(x_t) - \frac{1}{H} \sum_{h=1}^{H} m(y_t^h(b)) \tag{6}$$

where H is the number of simulations of the model. Then define an $nq \times 1$ vector (where $n = 4$ and $q = 1$) as in our previous notes

$$\begin{aligned}
g(b) \equiv E[u(x_t, b)] &= E\left[m(x_t)\right] - \frac{1}{H} \sum_{h=1}^{H} E\left[m(y_t^h(b))\right] \\
&= \mu(x) - \mu(y(b))
\end{aligned}$$

Note that **Global Identification** requires $g(b) = 0 \iff b = b_0$. In this example, the Global Identification condition is that there is a unique solution $b = b_0$ to the following equation

$$
\begin{aligned}
& g(b) = \mu(x) - \mu(y(b)) = 0 \\
\iff & \mu(x) = \begin{bmatrix} 0 \\ 1+b_0^2 \\ -b_0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1+b^2 \\ -b \\ 0 \end{bmatrix} = \mu(y(b)).
\end{aligned} \tag{7}
$$

In particular, in (7) the mean and second autorcorrelation do not identify b, the variance does not uniquely identify it since $b = \pm b_0$, while the autocorrelation uniquely identifies $b = b_0$.

In general we never actually have analytical expressions for $\mu(x)$ and $\mu(y(b))$ so cannot obtain an estimate as above. That's why we will use SMM to estimate b in a finite sample. Before moving to the finite sample case, it is instructive to state the asymptotic version of SMM since we can obtain an analytic expression for the variance covariance matrix S and hence the optimal weighting matrix $W^* = S^{-1}$. Let S denote the $n \times n$ asymptotic variance-covariance matrix of the n moment conditions u at the true parameter value $b = b_0$:

$$
S = \sum_{j=-\infty}^{\infty} E[u(x_t, b_0)u(x_{t-j}, b_0)'] \tag{8}
$$

If we make the assumption that the second moment of the true data and model are equal at the parameter $b = b_0$, then x_t and $y_t^h(b_0)$ have the same variance-covariance matrix in which case we can use either actual data or simulated data. In that case,

$$
S_x \equiv \sum_{j=-\infty}^{\infty} E[\{m(x_t) - \mu(x)\}\{m(x_{t-j}) - \mu(x)\}'] = \sum_{j=-\infty}^{\infty} E[\{m(y_t^h(b_0)) - \mu(y(b_0))\}\{m(y_{t-j}^h(b_0)) - \mu(y(b_0))\}'] \equiv S_y
$$

where S_x (S_y) is the asymptotic variance-covariance matrix of $m(x_t)$ (or $m(y_t^h(b_0))$). Note that data draws in x_t and simulation draws in $y_t^h(b_0)$ are uncorrelated and also there is no correlation between $\{y_t^h(b_0)\}_{h=1}^{H}$ across H. Hence, we can obtain (see appendix for the detailed derivation)

$$
S = \left(1 + \frac{1}{H}\right) S_x.
$$

Let $\Gamma_j \equiv E[\{m(x_t) - \mu(x)]\}\{m(x_{t-j}) - \mu(x)]\}']$ denote the j-th autocovariance of m. Then the asymptotic variance-covariance matrix of $m(x_t)$ is given by

$$
S_x = \Gamma_0 + \sum_{j=1}^{\infty}(\Gamma_j + \Gamma_j') \tag{9}
$$

Because we know the true DGP, we can compute Γ_js analytically:

$$
\Gamma_0 = \begin{bmatrix} \sigma_x^2 & 0 & 0 & 0 \\ 0 & 2\sigma_x^4 & -2b_0\sigma_x^2 & 0 \\ 0 & -2b_0\sigma_x^2 & \sigma_x^4 + b_0^2 & -b_0\sigma_x^2 \\ 0 & 0 & -b_0\sigma_x^2 & \sigma_x^4 \end{bmatrix}
$$

$$
\Gamma_1 = \begin{bmatrix} -b_0 & 0 & 0 & 0 \\ 0 & 2b_0^2 & 0 & 0 \\ 0 & -2b_0\sigma_x^2 & b_0^2 & 0 \\ 0 & 2b_0^2 & -b_0\sigma_x^2 & b_0^2 \end{bmatrix}
$$

$$
\Gamma_j = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \forall j \geq 2.
$$

where $\sigma_x^2 = 1 + b_0^2$ is the variance of x_t.[1] So the asymptotic variance-covariance matrix is:

$$\begin{aligned} S_x &= \Gamma_0 + \Gamma_1 + \Gamma_1' \\ &= \begin{bmatrix} (1-b_0)^2 & 0 & 0 & 0 \\ 0 & 2(1+4b_0^2+b_0^4) & -4b_0(1+b_0^2) & 2b_0^2 \\ 0 & -4b_0(1+b_0^2) & 1+5b_0^2+b_0^4 & -2b_0(1+b_0^2) \\ 0 & 2b_0^2 & -2b_0(1+b_0^2) & 1+4b_0^2+b_0^4 \end{bmatrix} \end{aligned}$$

At $b_0 = 0.5$,

$$S_x = \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0 & 4.125 & -2.5 & 0.5 \\ 0 & -2.5 & 2.3125 & -1.25 \\ 0 & 0.5 & -1.25 & 2.0625 \end{bmatrix}$$

The inverse of S is the optimal weighting matrix:

$$W^* = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 1.1115 & 1.5705 & 0.6823 \\ 0 & 1.5705 & 2.8621 & 1.3539 \\ 0 & 0.6823 & 1.3539 & 1.1400 \end{bmatrix} \tag{10}$$

In order to compute standard errors, we need the derivative of g (an $n \times \ell$ matrix): [2]

$$\begin{aligned} \nabla_b g(b_0) &= -\nabla_b \mu(y(b_0)) \\ &= -\begin{bmatrix} 0 \\ 2b_0 \\ -1 \\ 0 \end{bmatrix} \end{aligned} \tag{11}$$

This derivative is useful to see if the parameter is **Locally Identified**. To see this, take the first order

[1] Note first that since ε_t is $N(0,1)$, all odd higher order moments are zero. Then we can compute each cell using that fact. For example,

1. Consider cell $(1,1)$ in Γ_0. That is given by

$$\begin{aligned} & E[\{m(x_t) - \mu(x)\}\{m(x_t) - \mu(x)\}]' \\ =\; & E[\{(x_t - 0) - 0\}\{(x_t - 0) - 0\}] \\ =\; & E[x_t^2] = \sigma_x^2. \end{aligned}$$

2. Consider cell $(2,1)$ in Γ_0. That is given by

$$\begin{aligned} & E[\{(x_t - 0)^2 - (1+b_0^2)\}\{(x_t - 0) - 0\}] \\ =\; & E[x_t^3 - (1+b_0^2)x_t] \\ =\; & E[x_t^3] - (1+b_0^2)E[x_t] = 0 \end{aligned}$$

because the Normal distribution is zero as is its mean.

3. Consider cell $(2,2)$ in Γ_0. That is given by

$$\begin{aligned} & E[\{(x_t - 0)^2 - (1+b_0^2)\}\{(x_t - 0)^2 - (1+b_0^2)\}] \\ =\; & E[(x_t^2)^2 - 2(1+b_0^2)x_t^2 + (1+b_0^2)^2] \\ =\; & E[x_t^4] - 2(1+b_0^2)E[x_t^2] + (1+b_0^2)^2 \\ =\; & 3(\sigma_x^4) - 2\sigma_x^4 + \sigma_x^4 = 2\sigma_x^4 \end{aligned}$$

using the fact that the fourth moment of a normal is $3(\sigma_x^4)$.

[2] Recall, from Theorem 3.2 of Hansen:

$$\sqrt{T}(b_T - b_0) \to N(0, \left[\nabla_b g(b_0)' S^{-1} \nabla_b g(b_0)\right]^{-1})$$

approximation of $g(b)$ around b_0:

$$\begin{aligned} g(b) &\approx g(b_0) + \nabla_b g(b_0)(b - b_0) \\ &= \nabla_b g(b_0)(b - b_0) \end{aligned}$$

since $g(b_0) = 0$. For $b = b_0$ to be the unique solution to $\nabla_b g(b_0)(b - b_0) = 0$, it must be true that

$$\text{rank}(\nabla_b g(b_0)) = \ell$$

From (11), we can see that $\text{rank}(\nabla_b g(b_0)) = 1 = \ell$, so the parameter is locally identified. In contrast, if $\nabla_b g(b_0)$ in (11) was the zero vector (which has rank $0 < \ell$), then the objective would not respond to changes in the parameter.

Finally, once we know these analytical results, we can write down the population analogue of the SMM objective function as

$$J(b) = g(b)' W^* g(b)$$

The first order condition is

$$\begin{aligned} &\nabla_b \left(g(b)' W^* g(b) \right) = 0 \\ \iff &- \begin{bmatrix} 0 & 2b & -1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 1.1115 & 1.5705 & 0.6823 \\ 0 & 1.5705 & 2.8621 & 1.3539 \\ 0 & 0.6823 & 1.3539 & 1.1400 \end{bmatrix} g(b) = 0 \end{aligned}$$

where the second line results after some algebra.[3] If we evaluate $\nabla_b g(b)$ at $b = b_0$ and compute $\nabla_b g(b_0)' W^*$ then we obtain

$$\begin{bmatrix} 0 & 0.4590 & 1.2916 & 0.6715 \end{bmatrix} g(b) = 0$$

This means that the weight on the mean is zero. This is because the mean is not useful at all for the estimation of b. On the other hand, the second order autocovariance gets a positive weight, even though it is not useful itself. This is because even though the second order autocovariance doesn't depend on b, it is correlated with the variance and first order autocovariance, which is useful for the estimation of b. If we want to make the estimator efficient, we should take the information in the second autocovariance into account.

2 Small Samples

In general, we only have a finite sample of size T data, so we must construct the $n \times 1$ vector of data moments $M_T(x)$. Given the finite sample, in general $M_T(x) \neq \mu(x)$, but $M_T(x) \overset{a.s.}{\to} \mu(x)$ as $T \to \infty$.

2.1 Sample moments for the true data

We first generate a series of random sample $\{\varepsilon\}_{t=1}^T$ from $N(0, 1)$ and then construct a series of $\{x_t\}_{t=1}^T$ using the true DGP in (1) or

$$x_t = \varepsilon_t - b_0 \varepsilon_{t-1}$$

with $\varepsilon_0 = 0, b_0 = 0.5$, and $T = 200$ The 'true' data is plotted in Figure 1.

Using this data, we can compute the $m \times 1$ data moment vector by

$$M_T(x) = \frac{1}{T} \sum_{t=1}^{T} m(x_t).$$

For our case, the $m = 4$ data moment vector obtained from this simulation is

$$M_T(x) = \begin{bmatrix} -0.0153 \\ 1.1874 \\ -0.4269 \\ -0.0868 \end{bmatrix}. \tag{12}$$

[3] To see this, go to the Appendix.

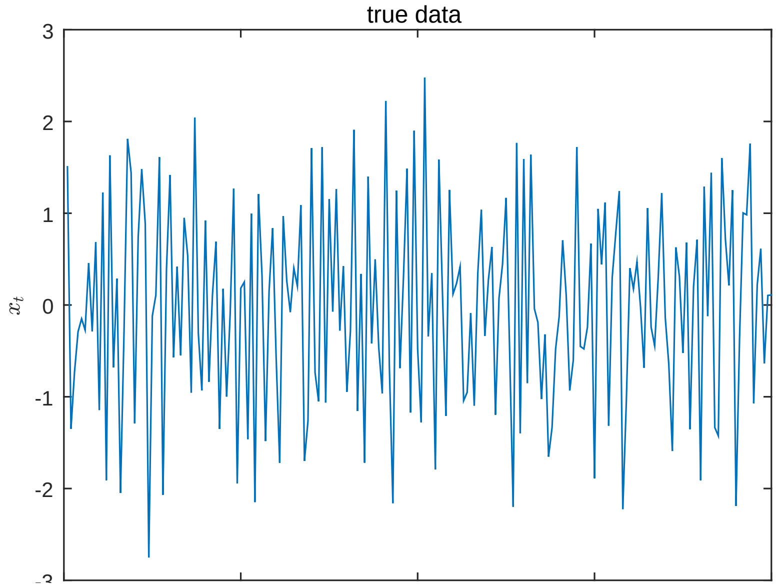

Figure 1: Simulated 'true' data

We can use this data to estimate the long run variance-covariance matrix. Unlike the handout, where it was assumed that there was no autocorrelation in $u(x_t, b)$, here we have autocorrelation so apply the Newey-West correction to what was presented in (9). In particular, letting

$$\widehat{\Gamma}_{T,j} \equiv \frac{1}{T}\sum_{t=j+1}^{T}\left[m(x_t) - M_T(x)\right]\left[m(x_{t-j}) - M_T(x)\right]'$$

denote the j-th autocovariance of m. Then the estimated sample variance-covariance matrix of $m(x_t)$ is given by

$$\widehat{S}_{x,T} = \widehat{\Gamma}_{T,0} + \sum_{j=1}^{\infty}\left(1 - \frac{j}{i(T)+1}\right)(\widehat{\Gamma}_{T,j} + \widehat{\Gamma}'_{T,j})$$

where $i(T)$ is the key to the Newey-West correction (here taken to be 4). The sample variance covariance matrix is given by

$$\hat{S}_{x,T} = \begin{bmatrix} 0.4147 & 0.0058 & -0.0895 & -0.0244 \\ 0.0058 & 1.8946 & -0.8869 & -0.1872 \\ -0.0895 & -0.8869 & 1.2988 & -0.6078 \\ -0.0244 & -0.1872 & -0.6078 & 1.5729 \end{bmatrix} \tag{13}$$

2.2 SMM Estimation

We first draw a series of random sample $\{\{e_t^h\}_{t=1}^T\}_{h=1}^H$. We will use the same draw in the whole estimation process. Given parameter value b, we can compute $\{\{y_t^i(b)\}_{t=1}^T\}_{h=1}^H$ in (2) or

$$y_t^h = e_t^h - be_{t-1}^h$$

where $e_0^h = 0$, $T = 200$, and $H = 10$. Then given b, we can compute the simulated moment

$$M_{TH}(y(b)) = \frac{1}{H}\sum_{i=1}^{H}\frac{1}{T}\sum_{t=1}^{T}m(y_t^h(b)).$$

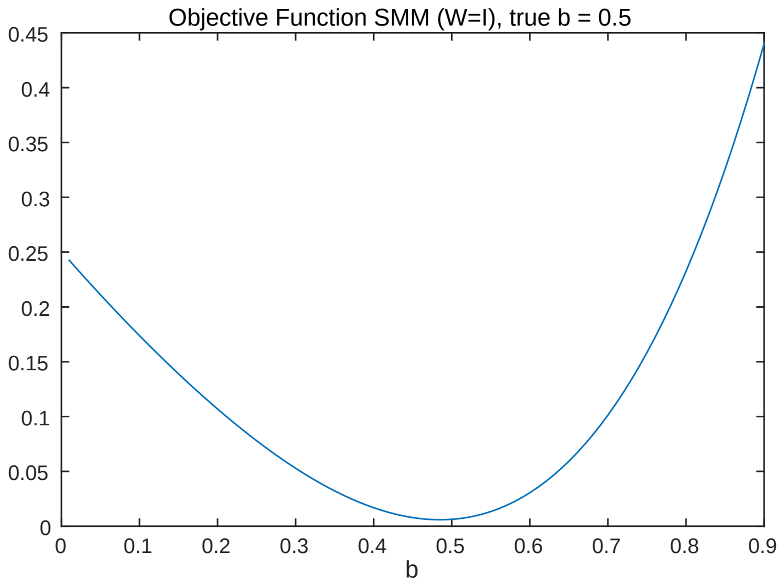

Figure 2: $J_{TH}(b)$ when $W = I$

Our objective is to choose b so that the weighted sum of squared residuals between the model moments $M_{TH}(y(b))$ and data moments $M_T(x)$ is minimized. In the first stage, we will use the $n \times n$ identity matrix I as a weighting matrix $W = I$ and obtain a consistent estimate of the parameter b which solves

$$\hat{b}^1_{TH} = \arg\min_b J_{TH}(b)$$
$$\text{where} \quad J_{TH}(b) \equiv [M_T(x) - M_{TH}(y(b))]'[M_T(x) - M_{TH}(y(b))].$$

Figure 2 plots the objective function $J_{TH}(b)$ weighted by the identity matrix.

The consistent value of the estimator in this case is

$$\hat{b}^1_{TH} = 0.4850.$$

In the second stage, we will use the inverse of the long-run variance-covariance matrix S as a weighting matrix to obtain the efficient SMM estimator. There are two ways to implement this.

2.2.1 Optimal Weighting Matrix from Data

We have already computed the variance covariance matrix in (13). Then, the weighting matrix is the inverse of (13), and it is given by

$$W^*_T = \begin{bmatrix} 2.5146 & 0.2155 & 0.4282 & 0.2301 \\ 0.2155 & 1.0110 & 0.9316 & 0.4836 \\ 0.4282 & 0.9316 & 1.8197 & 0.8206 \\ 0.2301 & 0.4836 & 0.8206 & 1.0140 \end{bmatrix}$$

Figure 3 plots the objective function $J_{TH}(b)$ when W^*_T using the optimal weighting matrix derived from the true data.

By minimizing this function, we can obtain the efficient SMM estimator[4]

$$\hat{b}^2_{TH,data} = 0.4993.$$

[4] If we use the actual data, we don't need $\hat{b}^1_{TH}$ to estimate the variance-covariance matrix because we only use the actual data. Hence this is actually not the second step estimator.

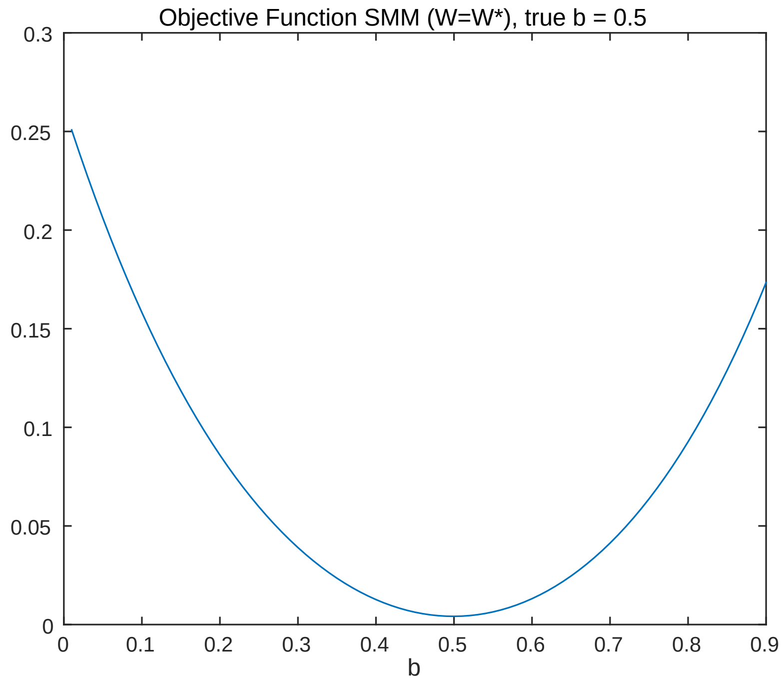

Figure 3: $J_{TH}(b)$ when $W = W_T^*$

2.2.2 Optimal Weighting Matrix from Simulation

Instead of the true data, we can use the simulated data under $b = \hat{b}^1_{TH}$ (which is consistent) to estimate the variance-covariance matrix. The variance-covariance matrix in this case is

$$\hat{S}_{y,TH} = \begin{bmatrix} 0.4472 & -0.0600 & 0.0459 & -0.0348 \\ -0.0600 & 3.4277 & -1.9441 & 0.3156 \\ 0.0459 & -1.9441 & 1.8748 & -0.8975 \\ -0.0348 & 0.3156 & -0.8975 & 1.7306 \end{bmatrix}.$$

Then the weighting matrix is

$$W^*_{TH} = \begin{bmatrix} 2.2441 & 0.0369 & 0.0023 & 0.0395 \\ 0.0369 & 0.8928 & 1.1272 & 0.4225 \\ 0.0023 & 1.1272 & 2.1335 & 0.9009 \\ 0.0395 & 0.4225 & 0.9009 & 0.9688 \end{bmatrix}.$$

If we use this as the weighting matrix, we obtain the second stage estimate[5]

$$\hat{b}^2_{TH,sim} = 0.4970.$$

2.2.3 Standard Errors and J Test

Once we have computed the estimator, we want to compute the standard errors of the estimator. We have

$$\sqrt{T}(\widehat{b}^2_{TH} - b_0) \to N(0, (1 + 1/H) \left[\nabla_b g_T(\widehat{b}^2_{TH})' \widehat{S}^{-1}_{y,TH} \nabla_b g_T(\widehat{b}^2_{TH})\right]^{-1})$$

[5]While there may be differences in the objective function when we use the optimal weighting matrix derived from the true and simulated data, in this case we did not find a large enough change so we don't plot it.

where

$$g_T(b) \equiv \frac{1}{T}\sum_{t=1}^{T} u(x_t, b) = \frac{1}{T}\sum_{t=1}^{T} m(x_t) - \frac{1}{H}\sum_{h=1}^{H}\frac{1}{T}\sum_{t=1}^{T} m(y_t^h(b))$$
$$= M_T(x) - M_{TH}(y(b))$$

So the derivative of g_T is given by

$$\nabla_b g_T(\hat{b}^2_{TH}) = -\nabla_b M_{TH}(y(\hat{b}^2_{TH}))$$
$$= -\frac{1}{TH}\sum_{h=1}^{H}\sum_{t=1}^{H}\frac{\partial m(y_t^h(\hat{b}^2_{TH}))}{\partial b}$$

In general we don't have an analytical formula for this derivative, so we will use the numerical derivative. Once can compute $M_{TH}(y(\hat{b}^2_{TH}))$, then compute $M_{TH}(y(\hat{b}^2_{TH} - s))$, take the difference, and divide by the step size s. The result is

$$\frac{\Delta M_{TH}(\hat{b}^2_{TH})}{\Delta b} = \begin{bmatrix} -0.0104 \\ 0.9342 \\ -0.9330 \\ -0.0234 \end{bmatrix}$$

Again, since there is a small sample error, this is broadly consistent with the theoretical result computed in (11) evaluated at $b_0 = 0.5$ given by $[0\ 1\ \ -1\ 0]'$.

The standard error of the estimator is

$$\text{Std}(\hat{b}^2_{TH}) = \frac{1}{T}\left[\nabla_b g_T(\hat{b}^2_{TH})' \left\{\left(1 + \frac{1}{H}\right)\hat{S}_{y,TH}\right\}^{-1} \nabla_b g_T(\hat{b}^2_{TH})\right]^{-1} = 0.089.$$

Once we have estimated the parameter, we can also test if the moment condition is true or not.

$$T\frac{H}{1+H} \times [M_T(x) - M_{TH}(y(\hat{b}^2_{TH}))]' W^*_{TH} [M_T(x) - M_{TH}(y(\hat{b}^2_{TH}))] = 0.7588.$$

The asymptotic distribution of this test statistics is $\chi(n-k)$, where n is the number of moments ($= 4$) and k is the number of parameters ($= 1$). The p value is 0.14, so we cannot reject the hypothesis that the model is true.

3 Bootstrap

In order to see the finite sample distribution of the estimators, we can use the bootstrap method. The algorithm is as follows.

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

2. Repeat 1 using another seed.

Every time you do step 1, 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.

The histogram of the estimator is plotted in figure 4. As theory predicts, $\hat{b}^2_{TH,data}$, which is the efficient estimator, has a smaller variance than $\hat{b}^1_{TH}$. To make it easier for us to compare the distributions, figure 5 plots the density function of the estimators, obtained by Kernel density estimation

$$f(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{cI}\sum_{i=1}^{I}\exp\left[-\frac{1}{2}\left(\frac{x - x_i}{c}\right)^2\right]$$

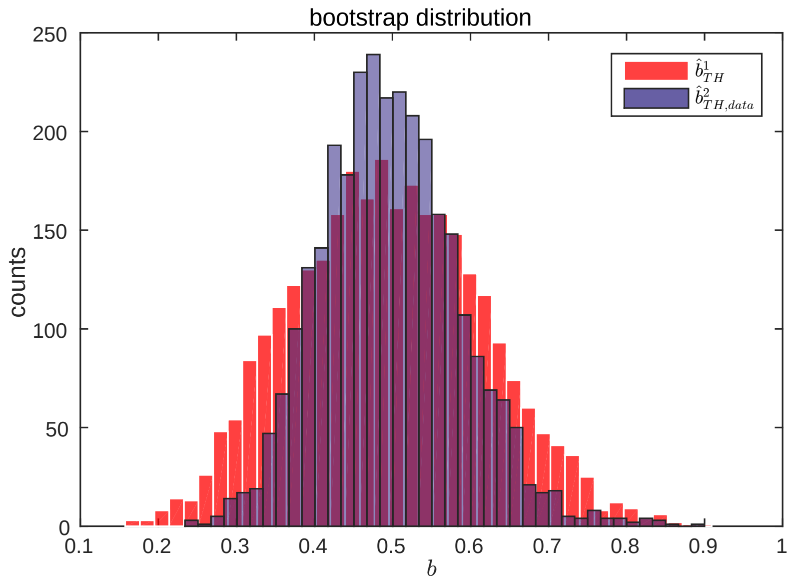

Figure 4: Bootstrap distributions: histogram

where I is the number of data and c is the bandwidth. We can see that the distribution of $b^2_{TH,data}$ looks very similar to that of $b^2_{TH,sim}$. This is because the model nests the true DGP (in the sense that it is the true DGP at b_0), so even if we use the simulated data to estimate the variance-covariance matrix, we can obtain the efficient estimator.

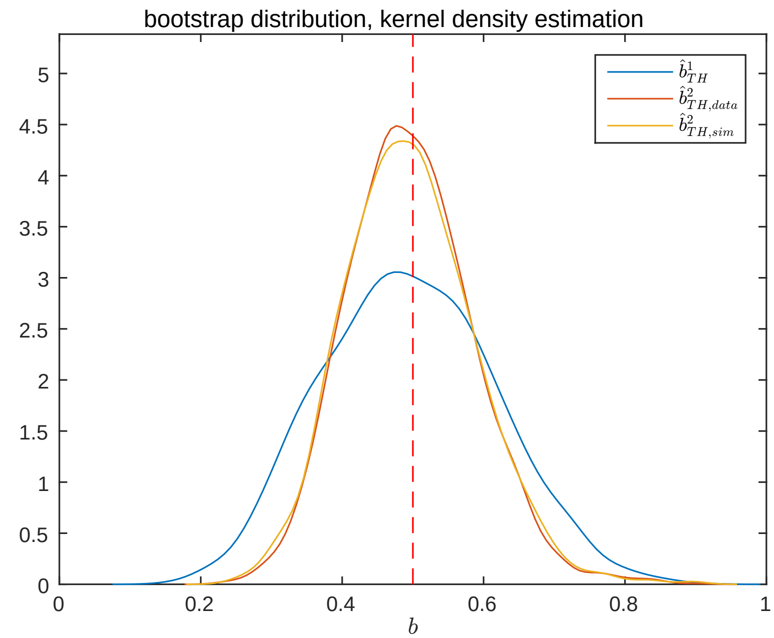

Figure 5: Bootstrap distributions, approximated by the kernel density estimation.

Appendix

3.1 Variance Covariance Matrix

From (6), we know

$$S = \sum_{j=-\infty}^{\infty} E\left\{\left[m(x_t) - \frac{1}{H}\sum_{h=1}^{H} m(y_t^h(b_0))\right]\left[m(x_{t-j}) - \frac{1}{H}\sum_{h=1}^{H} m(y_{t-j}^h(b_0))\right]'\right\} \tag{14}$$

Note that x_t and $y_t^h(b_0)$ are uncorrelated, and also there is no correlation between $\{y_t^h(b_0)\}_{h=1}^H$ across H. Then the terms inside the summation can be written as

$$E\left\{\left[\{m(x_t)-\mu(x)\} - \frac{1}{H}\sum_{h=1}^{H}\{m(y_t^h(b_0)) - \mu(y(b_0))\}\right]\left[\{m(x_{t-j})-\mu(x)\} - \frac{1}{H}\sum_{h=1}^{H}\{m(y_{t-j}^h(b_0)) - \mu(y(b_0))\}\right]'\right\}$$

$$= E[\{m(x_t)-\mu(x)\}\{m(x_{t-j})-\mu(x)\}'] + \frac{1}{H^2}E\left\{\left[\sum_{h=1}^{H}\{m(y_t^h(b_0)) - \mu(y(b_0))\}\right]\left[\sum_{h=1}^{H}\{m(y_{t-j}^h(b)) - \mu(y(b_0))\}\right]'\right\}$$

$$= E[\{m(x_t)-\mu(x)\}\{m(x_{t-j})-\mu(x)\}'] + \frac{1}{H}E[\{m(y_t^h(b_0)) - \mu(y(b_0))\}\{m(y_{t-j}^h(b_0)) - \mu(y(b_0))\}'] \tag{15}$$

In that case, (14) can be written

$$
\begin{aligned}
S &= \sum_{j=-\infty}^{\infty} \left\{ E[\{m(x_t) - \mu(x)\}\{m(x_{t-j}) - \mu(x)\}'] + \frac{1}{H} E[\{m(y_t^h(b_0)) - \mu(y(b_0))\}\{m(y_{t-j}^h(b_0)) - \mu(y(b_0))\}'] \right\} \\
&= \sum_{j=-\infty}^{\infty} E[\{m(x_t) - \mu(x)\}\{m(x_{t-j}) - \mu(x)\}'] + \frac{1}{H} \sum_{j=-\infty}^{\infty} E[\{m(y_t^h(b_0)) - \mu(y(b_0))\}\{m(y_{t-j}^h(b_0)) - \mu(y(b_0))\}'] \\
&= \left(1 + \frac{1}{H}\right) S_x. \qquad (16)
\end{aligned}
$$

3.2 First Order Condition

The first order condition is

$$
\begin{aligned}
0 &= \frac{dJ(b)}{db} \\
&= \nabla_b g(b)' W^* g(b) + g(b)' W^* \nabla_b g(b)
\end{aligned}
$$

Since $g(b)' W^* \nabla_b g(b)$ is a scalar,

$$
\begin{aligned}
g(b)' W^* \nabla_b g(b) &= (g(b)' W^* \nabla_b g(b))' \\
&= \nabla_b g(b)' W^* g(b)
\end{aligned}
$$

So

$$
\begin{aligned}
0 &= \frac{dJ(b)}{db} \\
&= \nabla_b g(b)' W^* g(b) + \nabla_b g(b)' W^* g(b) \\
&= 2\nabla_b g(b)' W^* g(b) \\
\iff 0 &= \nabla_b g(b)' W^* g(b)
\end{aligned}
$$