Method of Simulated Moments

September 27, 2022

Table of Contents

Parametric Conditional Models

- The development of econometrics techniques (Gourieroux & Monfort, 1996)
  - – 1960s: models with analytical expression of estimators.
  - 1970s – 1980s: tractable criteria function, but estimators based on numerical optimization.
  - 1990s – : criteria function without simple analytical expression (simulation methods).
- Parametric conditional models
  - y_t: endogenous variables, $\underline{y_t} = \left(\underline{y_0}, y_1, \ldots, y_t\right)$
  - z_t: strongly exogenous variables, $\underline{z_t} = \left(\underline{z_0}, z_1, \ldots, z_t\right)$
  - (y_t, z_t) is stationary.
  - We are interested in $f_0\left(y_1, \ldots, y_T / z_1, \ldots, z_T, \underline{y_0}\right) = \prod_{t=1}^{T} f_0\left(y_t / x_t\right)$, where $x_t = \left(\underline{y_{t-1}}, \underline{z_t}\right)$, since the conditional distribution and $f_0(z_1, ..., z_T, \underline{y_0})$ determines the distribution of all the observations.
  - The model $M = \left\{f\left(y_t / x_t; \theta\right), \theta \in \Theta\right\}$ is:
    1. **well-specified** if $f_0(y_t / x_t)$ belongs to M.
    2. **identifiable** if $f_0(y_t / x_t) = f(y_t / x_t; \theta) \iff \theta = \theta_0$
  - Estimating $f_0(y_t / x_t)$ is equivalent to estimating θ_0.

Optimization Estimators

- Ψ_T: a criterion depending on the observations $y_t, z_t, t = 1, \ldots, T$.
- $\widehat{\theta}_T \equiv \arg\max_\theta \Psi_T\left(y_1, \ldots, y_T, z_1, \ldots, z_T; \theta\right) = \arg\max_\theta \Psi_T(\theta)$
- Use numerical methods (e.g. Newton-Raphson) to solve the system of F.O.C.s.
- Conditions for Consistency ($\widehat{\theta}_T \to \theta_0$):
  (i) The well normalized criterion function **uniformly converges** to some limit function:
  $$\lim_{T\to\infty} \frac{1}{T}\Psi_T(\theta) = \Psi_\infty(\theta).$$
  (ii) $\theta_0^\infty = \arg\max \Psi_\infty(\theta)$ is unique.
  (iii) $\theta_0^\infty = \theta_0$.
- Asymptotic Normality: If
  $\lim_{T\to\infty}\left[-\frac{1}{T}\frac{\partial^2\Psi_T}{\partial\theta\partial\theta'}\left(\theta_0\right)\right] = J\left(\theta_0\right)$, where $J\left(\theta_0\right)$ is invertible;
  $\frac{1}{\sqrt{T}}\frac{\partial\Psi_T}{\partial\theta}\left(\theta_0\right) \xrightarrow{d} N\left[0, I\left(\theta_0\right)\right]$, where $I\left(\theta_0\right)$ is invertible.
  Then $\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left[0, J\left(\theta_0\right)^{-1} I\left(\theta_0\right) J\left(\theta_0\right)^{-1}\right]$. ▸ proof

- i.i.d. observations $(y_i, z_i), i = 1, \ldots, n$
- We assume the conditional expectation of function $K(y_i, z_i)$ of size q given z_i has a well-specified form:

$$E_0 \left[K \left(y_i, z_i \right) / z_i \right] = k \left(z_i; \theta_0 \right),$$

  where E_0 is the expectation for the true distribution of (y, z), and θ_0 is the true value of the parameter whose size is p.
- Now let Z_i be a matrix function of z_i with size (K, q), where $K \geq p$. The elements of Z_i may be seen as instrumental variables:

$$E_0 Z_i \left[K \left(y_i, z_i \right) - k \left(z_i, \theta_0 \right) \right] = 0.$$

- The GMM estimators are based on the empirical counterpart of the above orthogonality conditions. If Ω is a (K, K) symmetric positive semi-definite matrix, the estimator is defined by:

$$\hat{\theta}_n(\Omega) = \arg\min_{\theta} \left(\sum_{i=1}^{n} Z_i \left[K \left(y_i, z_i \right) - k \left(z_i; \theta \right) \right] \right)' \Omega \left(\sum_{i=1}^{n} Z_i \left[K \left(y_i, z_i \right) - k \left(z_i; \theta \right) \right] \right).$$

GMM
Static Case

- Under regularity conditions (Newey & McFadden, 1994), GMM has the properties of optimization estimators:
  (i) $\hat{\theta}_n(\Omega)$ is a consistent estimator of the true value θ_0;
  (ii) The GMM estimator is asymptotically normal:
$$\sqrt{n}\left(\hat{\theta}_n(\Omega) - \theta_0\right) \xrightarrow{d} N\left(0, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\right),$$
  where:
$$\begin{aligned}
\Sigma_1 &= D'\Omega D \\
\Sigma_2 &= D'\Omega V_0\left\{Z\left[K(y,z) - k\left(z,\theta_0\right)\right]\right\}\Omega D, \\
D &= E_0\left[Z\frac{\partial k}{\partial \theta'}\left(z;\theta_0\right)\right].
\end{aligned}$$
- An optimal choice of the matrix $\Omega^* = \left(V_0\left\{Z\left[K(y,z) - k\left(z;\theta_0\right)\right]\right\}\right)^{-1}$
  Then $\Sigma_1 = \Sigma_2$, and $V_{as}\left[\sqrt{n}\left(\hat{\theta}_n\left(\Omega^*\right) - \theta_0\right)\right] = \left[D'\Omega^* D\right]^{-1}$.
- We can replace θ_0 with a consistent estimator $\tilde{\theta}_n$ to get a feasible estimator of Ω^*.

- GMM based on dynamic conditional moments
  - lagged endogenous variables are included in raw IVs:

$$E_0\left[K\left(y_t, x_t\right)/x_t\right] = E_0\left[K\left(y_t, \underline{y_{t-1}}, \underline{z_t}\right)/\underline{y_{t-1}}, \underline{z_t}\right] = k\left(x_t; \theta_0\right),$$
$$Z(x_t) = Z(\underline{y_{t-1}}, \underline{z_t})$$

  - GMM estimator:

$$\hat{\theta}_T(\Omega) = \arg\min_{\theta}\left\{\sum_{t=1}^{T} Z_t\left[K\left(y_t, x_t\right) - k\left(x_t; \theta\right)\right]\right\}' \Omega \left\{\sum_{t=1}^{T} Z_t\left[K\left(y_t, x_t\right) - k\left(x_t, \theta\right)\right]\right\}$$

  - The estimator has the same asymptotic properties as in the static case after replacing z_i by x_t.

- GMM based on static conditional moments
  - Raw IVs only contains exogenous variables

$$
\begin{aligned}
&E_0\left[K\left(y_t, \underline{z_t}\right)/\underline{z_t}\right] = k\left(\underline{z_t}; \theta_0\right). \\
&Z_t = Z(\underline{z_t})
\end{aligned}
$$

  - GMM estimator:

$$
\begin{aligned}
&\hat{\theta}_T(\Omega) \\
&= \arg\min_{\theta} \left\{ \sum_{t=1}^{T} Z_t \left[K\left(y_t, \underline{z_t}\right) - k\left(\underline{z_t}; \theta\right) \right] \right\}' \Omega \left\{ \sum_{t=1}^{T} Z_t \left[K\left(y_t, \underline{z_t}\right) - k\left(\underline{z_t}, \theta\right) \right] \right\}
\end{aligned}
$$

  - $\Sigma_2 = \lim_T V_0 \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t \left[K\left(y_t, \underline{z_t}\right) - k\left(\underline{z_t}; \theta_0\right) \right] \right\} = \Gamma_0 + \sum_{h=1}^{\infty} \left[\Gamma_h + \Gamma_h' \right],$
    where

$$
\Gamma_h = \mathrm{cov}_0 \left[Z_t \left(K\left(y_t, \underline{z_t}\right) - k\left(\underline{z_t}; \theta_0\right) \right), Z_{t-h} \left(K\left(y_{t-h}, \underline{z_{t-h}}\right) - k\left(\underline{z_{t-h}}; \theta_0\right) \right) \right].
$$

(use Newey-West estimator)

Table of Contents

MSM Estimators

Static Case

- GMM is calibrating the theoretical moments to the observed empirical moments.
- But the theoretical moments may be difficult to compute.
- MSM circumvents this issue by calibrating the simulated empirical moments to the observed empirical moments.
- **unbiased simulator**: $\tilde{k}(z_i, u_i; \theta)$, where u_i has a known distribution, and $E(\tilde{k}(z_i, u_i; \theta) \mid z_i) = k(z_i; \theta)$.
- MSM estimator: $\hat{\theta}_{Sn}(\Omega) = \arg\min_\theta \psi_{Sn}(\theta)$, where:

$$\psi_{Sn}(\theta) = \left\{ \sum_{i=1}^{n} Z_i \left[K(y_i, z_i) - \frac{1}{S} \sum_{s=1}^{S} \tilde{k}(z_i, u_i^s; \theta) \right] \right\}' \Omega \times \left\{ \sum_{i=1}^{n} Z_i \left[K(y_i, z_i) - \frac{1}{S} \sum_{s=1}^{S} \tilde{k}(z_i, u_i^s; \theta) \right] \right\}.$$

- When S tends to infinity, MSM estimator becomes GMM estimator.

MSM Estimators

Dynamic Case (dynamic moment conditions)

- MSM based on dynamic conditional moment
  - dynamic moment conditions

$$E_0\left[K\left(y_t,x_t\right)/x_t\right]=E_0\left[K\left(y_t,\underline{y_{t-1}},\underline{z_t}\right)/\underline{y_{t-1}},\underline{z_t}\right]=k\left(x_t;\theta_0\right),$$

  - This unbiased simulator $\tilde{k}\left(x_t,u;\theta\right)$ s.t. $E\left[\tilde{k}\left(x_t,u;\theta\right)/x_t\right]=k\left(x_t;\theta\right)$, where **the distribution of** u **given** x_t is known.
  - The simulated moment estimator is defined by:

$$\hat{\theta}_{ST}(\Omega)=\arg\min_{\theta}\psi_{ST}(\theta),$$

where:

$$\psi_{ST}(\theta)=\left\{\sum_{t=1}^{T}Z\left(x_t\right)\left[K\left(y_t,x_t\right)-\frac{1}{S}\sum_{s=1}^{S}\tilde{k}\left(x_t,u_t^s;\theta\right)\right]\right\}'\Omega$$
$$\times\left\{\sum_{t=1}^{T}Z\left(x_t\right)\left[K\left(y_t,x_t\right)-\frac{1}{S}\sum_{s=1}^{S}\tilde{k}\left(x_t,u_t^s;\theta\right)\right]\right\},$$

and u_t^s is drawn in the known **conditional distribution** of u given x_t.

MSM Estimators

Dynamic Case (static moment conditions)

- MSM based on static conditional moments
  - static moment condition: $E_0\left[K\left(y_t, \underline{z_t}\right)/\underline{z_t}\right] = k\left(\underline{z_t};\theta_0\right)$,
  - an unbiased simulator $\tilde{k}$ s.t. $E_0\left[\tilde{k}\left(\underline{z_t}, u;\theta\right)/\underline{z_i}\right] = k\left(\underline{z_i};\theta_0\right)$,
  - the simulated moment estimator is defined by:

$$\hat{\theta}_{ST}(\Omega) = \arg\min_{\theta} \psi_{ST}(\theta),$$

$$\psi_{ST}(\theta) = \left\{\sum_{t=1}^{T} Z\left(\underline{z_t}\right)\left[K\left(y_t, \underline{z_t}\right) - \frac{1}{S}\sum_{s=1}^{S}\tilde{k}\left(\underline{z_t}, u_t^s;\theta\right)\right]\right\}' \Omega$$
$$\times \left\{\sum_{t=1}^{T} Z\left(\underline{z_t}\right)\left[K\left(y_t, z_{\underline{t}}\right) - \frac{1}{S}\sum_{s=1}^{S}\tilde{k}\left(\underline{z_t}, u_t^s;\theta\right)\right]\right\},$$

  and u_t^s is drawn in the conditional distribution of u given $\underline{z_t}$.

  - We need path simulations for constructing $\tilde{k}$.
  - If the model cannot be converted to reduced form such as $y_t = r(\underline{y_{t-1}}, \underline{z_t}, u;\theta)$, we can only use static conditional moments.

Table of Contents

Asymptotic properties of MSM

- We first consider the static case.
- When n tends to infinity and S is fixed, (i) $\hat{\theta}_{Sn}(\Omega)$ is strongly consistent; (ii) $\sqrt{n}\left[\hat{\theta}_{Sn}(\Omega)-\theta_0\right] \underset{n\to\infty}{\xrightarrow{d}} N\left[0; Q_S(\Omega)\right]$, where:

$$\begin{aligned} Q_S(\Omega) &= \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} + \frac{1}{S}\Sigma_1^{-1}D'\Omega E_0 V(Z\tilde{k}/z)\Omega D\Sigma_1^{-1} \\ &= \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} + \frac{1}{S}\Sigma_1^{-1}D'\Omega V_0[Z(\tilde{k}-k)]\Omega D\Sigma_1^{-1} \\ &= \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} + \frac{1}{S}\Sigma_1^{-1}D'\Omega E_0\left[ZV(\tilde{k}/z)Z'\right]\Omega D\Sigma_1^{-1}, \end{aligned}$$

with:

$$\begin{aligned} \Sigma_2 &= D'\Omega V_0[Z(K-k)]\Omega D, \\ D &= E_0\left[Z\frac{\partial k}{\partial\theta'}\right], \\ \Sigma_1 &= D'\Omega D, \end{aligned}$$

and where $\tilde{k}$, k and K are simplified notations for $\tilde{k}\left(z,u;\theta_0\right)$, $k(z;\theta_0)$ and $K(y,z)$ respectively, $k=E(K(y,z)/z)$. (See Gourieroux and Monfort (1996) for proof.)
- The first part of MSM covariance matrix $Q_S(\Omega)$ is just GMM covariance matrix, and the second part is the effect of simulations.
- When S $\to\infty$, MSM estimators coincide with GMM estimators.

Asymptotic properties of MSM

- We can increase the efficiency of simulation by decease the number of random terms in simulator, more specifically, let us assume that $y = r(z, \varepsilon; \theta)$.
  (i) If u is a subvector of ε, and the simulator $\tilde{k}(u, z : \theta) = E_\theta[K(y, z)/z, u]$, then:
  $$Q_S(\Omega) = \left(1 + \frac{1}{S}\right) \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1} - \frac{1}{S} \Sigma_1^{-1} D' \Omega V_0[Z(K - \tilde{k})] \Omega D' \Sigma_1^{-1}.$$
  (ii) In particular, $Q_S(\Omega) \ll \left(1 + \frac{1}{S}\right) \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}$, and the upper bound is reached for the simulator $\tilde{k}(z, \varepsilon; \theta) = K[r(z, \varepsilon; \theta), z]$ (**frequency simulator**) corresponding to $u = \varepsilon$.
- If $S = 1$, the confidence intervals increase at most by 41% compared to GMM. If $S = 10$, the confidence intervals increase at most by only 5%!
- For **dynamic MSM** based on dynamic conditional moments, the covariance matrix is the same as the static case. If we use static (or unconditional) moments and simulate the whole path of y_t, then the covariance matrix is $(1 + 1/S)$ times the GMM covariance matrix.

Table of Contents

Optimal MSM

optimal weighting matrix

- The asymptotic variance-covariance matrix of the MSM estimator is:

$$
\begin{aligned}
Q_S(\Omega) &= \Sigma_1^{-1} D' \Omega \left\{ V_0[Z(K-k)] + \frac{1}{S} V_0[Z(\tilde{k}-k)] \right\} \Omega D \Sigma_1^{-1} \\
&= \left(D'\Omega D\right)^{-1} D'\Omega \left\{ V_0[Z(K-k)] + \frac{1}{S} V_0[Z(\tilde{k}-k)] \right\} \Omega D \left(D'\Omega D\right)^{-1}.
\end{aligned}
$$

- the optimal choice of the matrix is:

$$
\Omega^* = \left\{ V_0[Z(K-k)] + \frac{1}{S} V_0[Z(\tilde{k}-k)] \right\}^{-1}.
$$

The asymptotic variance-covariance matrix corresponding to this choice is:

$$
Q_S\left(\Omega^*\right) = \left(D'\Omega^* D\right)^{-1},
$$

where

$$
D = E_0\left(Z \frac{\partial k}{\partial \theta'} \right).
$$

Optimal MSM

optimal weighting matrix

- To get a good approximation of k, it is necessary to have **a large number of** replications S_2. Let us denote by $u^s_{i,2}, s = 1, \ldots S_2$, some other simulated values of the random term with known distribution, the matrix

$$
\begin{aligned}
\hat{\Omega}^* = \Bigg\{ & \frac{1}{n} \sum_{i=1}^{n} Z_i \left[K\left(y_i, z_i\right) - \frac{1}{S_2} \sum_{s=1}^{S_2} \tilde{k}\left(z_i, u^s_{i,2}; \tilde{\theta}_n\right) \right] \\
& \times \left[K\left(y_i, z_i\right) - \frac{1}{S_2} \sum_{s=1}^{S_2} \tilde{k}\left(z_i, u^s_{i,2}; \tilde{\theta}_n\right) \right]' Z_i' \\
& + \frac{1}{S} \frac{1}{n} \sum_{i=1}^{n} Z_i \left[\tilde{k}\left(z_i, u^{s_1}_i; \tilde{\theta}_n\right) - \frac{1}{S_2} \sum_{s=1}^{S_2} \tilde{k}\left(z_i, u^s_{i,2}; \tilde{\theta}_n\right) \right] \\
& \times \left[\tilde{k}\left(z_i, u^{s_1}_i; \tilde{\theta}_n\right) - \frac{1}{S_2} \sum_{s=1}^{S_2} \tilde{k}\left(z_i, u^s_{i,2}; \tilde{\theta}_n\right) \right]' Z_i' \Bigg\}^{-1},
\end{aligned}
$$

where $u^{s_1}_i$ is a simulated value of u, is a consistent estimator of the optimal matrix Ω^*, when n and S_2 tend to infinity.

Optimal MSM

optimal weighting matrix

- If we use frequency simulator ($K = \tilde{k}$), the optimal weighting matrix is the same with GMM.
- $\Omega^* = \{V_0[Z(K - k)]\}^{-1}$.
- $Q^* = (1 + 1/S)\,(D'\Omega^* D)^{-1}$.

Optimal MSM
optimal instruments

- When the optimal weighting matrix is retained, the asymptotic variance-covariance matrix is:

$$\begin{aligned} Q_S(\Omega^*) &= \left\{ E_0\left(\frac{\partial k'}{\partial\theta} Z'\right) \left\{ V_0[Z(K-k)] + \frac{1}{S} V_0[Z(\tilde{k}-k)] \right\}^{-1} E_0\left(Z \frac{\partial k}{\partial\theta'}\right) \right\}^{-1} \\ &= \left\{ E_0\left(\frac{\partial k'}{\partial\theta} Z'\right) \left\{ E_0\left(Z\left[V_0(K/z) + \frac{1}{S} V(\tilde{k}/z)\right] Z'\right) \right\}^{-1} E_0\left(Z \frac{\partial k}{\partial\theta}\right) \right\}^{-1}. \end{aligned}$$

- If A and C are random matrices of suitable dimensions, are functions of z, and are such that C is square and positive definite, then the matrix

$$E_0(A'Z')\left[E_0(ZCZ')\right]^{-1} E_0(ZA)$$

is maximized for $Z = A'C^{-1}$ and the maximum is $E_0\left(A'C^{-1}A\right)$.

- The optimal instruments are:

$$Z_S^* = \frac{\partial k'}{\partial \theta}\left[V_0(K/z) + \frac{1}{S}V(\tilde{k}/z)\right]^{-1},$$

where the different functions $k, \tilde{k}$ are evaluated at the true value θ_0. With this choice the asymptotic covariance matrix is:

$$Q_S^* = \left\{E_0\left[\frac{\partial k'}{\partial \theta}\left(V_0(K/z) + \frac{1}{S}V(\tilde{k}/z)\right)^{-1}\frac{\partial k}{\partial \theta'}\right]\right\}^{-1}.$$

- For frequency simulator, the optimal instruments for MSM coincide with which of GMM:

$$Z_S^* = \frac{\partial k'}{\partial \theta}\left[V_0(K/z)\right]^{-1},$$

$$Q_S^* = \left\{E_0\left[\frac{\partial k'}{\partial \theta}\left(V_0(K/z)\right)^{-1}\frac{\partial k}{\partial \theta'}\right]\right\}^{-1}.$$

Proofs

- Asymptotic normality of optimization estimators:

  Whenever the estimator is consistent, we may expand the first order conditions ($\frac{\partial \Psi_T}{\partial \theta}\left(\widehat{\theta}_T\right) = 0$) around the true-value θ_0. We get:

$$\frac{\partial \Psi_T}{\partial \theta}(\theta_0) + \frac{\partial^2 \Psi_T}{\partial \theta \partial \theta'}(\theta_0)\left(\widehat{\theta}_T - \theta_0\right) \simeq 0$$
$$\Longleftrightarrow \widehat{\theta}_T - \theta_0 \simeq \left[-\frac{\partial^2 \Psi_T}{\partial \theta \partial \theta'}(\theta_0)\right]^{-1} \frac{\partial \Psi_T}{\partial \theta}(\theta_0).$$

  Since we assume

$$\lim_{T \to \infty}\left[-\frac{1}{T}\frac{\partial^2 \Psi_T}{\partial \theta \partial \theta'}(\theta_0)\right] = J(\theta_0), \text{ where } J(\theta_0) \text{ is invertible;}$$
$$\frac{1}{\sqrt{T}}\frac{\partial \Psi_T}{\partial \theta}(\theta_0) \xrightarrow{d} N[0, I(\theta_0)], \text{ where } I(\theta_0) \text{ is invertible.}$$

Proofs

We have

$$
\begin{aligned}
\sqrt{T}\left(\widehat{\theta}_{T-}\theta_0\right) &\simeq \left[-\frac{1}{T}\frac{\partial^2\Psi_T}{\partial\theta\partial\theta'}\left(\theta_0\right)\right]^{-1}\frac{1}{\sqrt{T}}\frac{\partial\Psi_T}{\partial\theta}\left(\theta_0\right)\\
&\simeq J\left(\theta_0\right)^{-1}\frac{1}{\sqrt{T}}\frac{\partial\Psi_T}{\partial\theta}\left(\theta_0\right)\\
&\xrightarrow{d} J\left(\theta_0\right)^{-1}N[0, I(\theta_0)]. \quad \square
\end{aligned}
$$

back

References

Gourieroux, C., & Monfort, A. (1996). *Simulation-based econometric methods*. Oxford university press.

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, *4*, 2111–2245.