Nonlinear GMM

Eric Zivot

Winter, 2013

Nonlinear GMM estimation occurs when the K GMM moment conditions $\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})$ are nonlinear functions of the p model parameters $\boldsymbol{\theta}$.

- The moment conditions $\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})$ may be $K \geq p$ nonlinear functions satisfying

$$E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$$

- Alternatively, for a response variable y_t, L explanatory variables $\mathbf{z}_t$, and K instruments $\mathbf{x}_t$, the model may define a nonlinear error term ε_t

$$a(y_t, \mathbf{z}_t; \boldsymbol{\theta}_0) = \varepsilon_t$$

such that

$$E[\varepsilon_t] = E[a(y_t, \mathbf{z}_t; \boldsymbol{\theta}_0)] = 0$$

Note: if $\mathbf{z}_t$ is endogenous then cannot use nonlinear least squares to estimate θ.

Given $\mathbf{x}_t$ orthogonal to ε_t, define

$$\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0) = \mathbf{x}_t \varepsilon_t = \mathbf{x}_t a(y_t, \mathbf{z}_t; \boldsymbol{\theta}_0)$$

so that

$$E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)] = E[\mathbf{x}_t \varepsilon_t] = E[\mathbf{x}_t a(y_t, \mathbf{z}_t; \boldsymbol{\theta}_0)] = \mathbf{0}$$

defines the GMM orthogonality conditions.

In general, the GMM moment equations produce a system of K nonlinear equations in p unknowns.

Global identification of $\boldsymbol{\theta}_0$ requires that

$$\begin{aligned} E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)] &= \mathbf{0} \\ E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})] &\neq \mathbf{0} \text{ for } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{aligned}$$

Local Identification requires that the $K \times p$ matrix

$$\mathbf{G} = E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right]$$

has full column rank p.

Remark

Global identification does not require differentiability of $\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)$

Intuition about local identification

Consider a first order Taylor series expansion of $\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})$ about $\boldsymbol{\theta} = \boldsymbol{\theta}_0$:

$$\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}) = \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0) + \frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + error$$

Then

$$\begin{aligned} E\left[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})\right] &\approx E\left[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)\right] + E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right](\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ &= E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right](\boldsymbol{\theta} - \boldsymbol{\theta}_0) \end{aligned}$$

For $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ to be the unique solution to $E\left[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})\right] = \mathbf{0}$ it must be the case that

$$\text{rank}\left(E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right]\right) = p$$

The sample moment condition for an arbitrary $\boldsymbol{\theta}$ is

$$\mathbf{g}_n(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^{n} \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})$$

If $K = p$, then $\boldsymbol{\theta}_0$ is apparently just identified and the GMM objective function is

$$J(\boldsymbol{\theta}) = n\mathbf{g}_n(\boldsymbol{\theta})'\mathbf{g}_n(\boldsymbol{\theta})$$

which does not depend on a weight matrix.

The corresponding GMM estimator is then

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \; J(\boldsymbol{\theta})$$

and solves

$$\mathbf{g}_n(\hat{\boldsymbol{\theta}}) = \mathbf{0}$$

If $K > p$, then $\boldsymbol{\theta}_0$ is apparently overidentified.

Let $\hat{\mathbf{W}}$ denote a $K \times K$ symmetric and p.d. weight matrix, possibly dependent on the data, such that $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$ as $n \to \infty$ with $\mathbf{W}$ symmetric and p.d.

The GMM estimator of $\boldsymbol{\theta}_0$, denoted $\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})$, is defined as

$$\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}) = \arg\min_{\boldsymbol{\theta}} \; J(\boldsymbol{\theta}, \hat{\mathbf{W}}) = n\mathbf{g}_n(\boldsymbol{\theta})'\hat{\mathbf{W}}\mathbf{g}_n(\boldsymbol{\theta})$$

The first order conditions are

$$\begin{aligned} \frac{\partial J(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}), \hat{\mathbf{W}})}{\partial \boldsymbol{\theta}} &= 2\mathbf{G}_n(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}))'\hat{\mathbf{W}}\mathbf{g}_n(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})) = \mathbf{0} \\ \mathbf{G}_n(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})) &= \frac{\partial \mathbf{g}_n(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}))}{\partial \boldsymbol{\theta}'} \end{aligned}$$

The efficient GMM estimator uses $\hat{\mathbf{W}} = \hat{\mathbf{S}}^{-1}$ such that

$$\hat{\mathbf{S}} \xrightarrow{p} \mathbf{S} = \mathsf{avar}(\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0)).$$

If $\{\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)\}$ is an ergodic-stationary MDS then

$$\mathbf{S} = E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)']$$

If $\{\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)\}$ is a serially correlated linear process then

$$\begin{aligned}
\mathbf{S} &= \mathsf{LRV} = \boldsymbol{\Gamma}_0 + \sum_{j=1}^{\infty}(\boldsymbol{\Gamma}_j + \boldsymbol{\Gamma}_j') = \boldsymbol{\Psi}(1)\boldsymbol{\Sigma}\boldsymbol{\Psi}(1)' \\
\boldsymbol{\Gamma}_0 &= E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)\mathbf{g}_t'(\mathbf{w}_t, \boldsymbol{\theta}_0)],\ \boldsymbol{\Gamma}_j = E[\mathbf{g}(\mathbf{w}_\mathbf{t}, \boldsymbol{\theta}_0)\mathbf{g}(\mathbf{w}_{t-j}, \boldsymbol{\theta}_0)']
\end{aligned}$$

As with efficient GMM estimation of linear models, the efficient GMM estimator of nonlinear models may be computed using a two-step, iterated, or continuous updating estimator.

Computation

Since the GMM objective function is a quadratic form, the Gauss-Newton (GN) algorithm is well suited for finding the minimum.

The GN algorithm starts from a first order Taylor series approximation to $\mathbf{g}_n(\boldsymbol{\theta})$ at a starting value $\hat{\boldsymbol{\theta}}_1$

$$\begin{aligned}
\mathbf{g}_n(\boldsymbol{\theta}) &= \mathbf{g}_n(\hat{\boldsymbol{\theta}}_1) + \mathbf{G}_n(\hat{\boldsymbol{\theta}}_1)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_1) + error \\
&\approx \left[\mathbf{g}_n(\hat{\boldsymbol{\theta}}_1) - \mathbf{G}_n(\hat{\boldsymbol{\theta}}_1)\hat{\boldsymbol{\theta}}_1\right] - \left[-\mathbf{G}_n(\hat{\boldsymbol{\theta}}_1)\right]\boldsymbol{\theta} \\
&= \mathbf{v}_1 - \mathbf{G}_1\boldsymbol{\theta}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{g}_n(\hat{\boldsymbol{\theta}}_1) + \mathbf{G}_1\hat{\boldsymbol{\theta}}_1 \\
\mathbf{G}_1 &= -\mathbf{G}_n(\hat{\boldsymbol{\theta}}_1) = -\frac{\partial \mathbf{g}_n(\hat{\boldsymbol{\theta}}_1)}{\partial \boldsymbol{\theta}'}
\end{aligned}$$

Note: The linear approximation

$$\mathbf{g}_n(\boldsymbol{\theta}) = \mathbf{v}_1 - \mathbf{G}_1\boldsymbol{\theta}$$

is like the linear GMM moment condition

$$\mathbf{g}_n(\boldsymbol{\theta}) = \mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\theta}$$

where

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{g}_n(\hat{\boldsymbol{\theta}}_1) + \mathbf{G}_1\hat{\boldsymbol{\theta}}_1 = \mathbf{S}_{xy} \\
\mathbf{G}_1 &= -\mathbf{G}_n(\hat{\boldsymbol{\theta}}_1) = \mathbf{S}_{xz}
\end{aligned}$$

Now, do linear GMM using the approximate linear moment condition

$$\min_{\boldsymbol{\theta}} \tilde{J}(\boldsymbol{\theta}, \hat{\mathbf{W}}) = [\mathbf{v}_1 - \mathbf{G}_1\boldsymbol{\theta}]' \, \hat{\mathbf{W}} \, [\mathbf{v}_1 - \mathbf{G}_1\boldsymbol{\theta}]$$

The linear GMM estimator has the closed form solution

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_2 &= \left(\mathbf{G}_1'\hat{\mathbf{W}}\mathbf{G}_1\right)^{-1} \mathbf{G}_1'\hat{\mathbf{W}}\mathbf{v}_1 \\
&= \left(\mathbf{G}_1'\hat{\mathbf{W}}\mathbf{G}_1\right)^{-1} \mathbf{G}_1'\hat{\mathbf{W}} \left(\mathbf{g}_n(\hat{\boldsymbol{\theta}}_1) + \mathbf{G}_1\hat{\boldsymbol{\theta}}_1\right) \\
&= \hat{\boldsymbol{\theta}}_1 + \left(\mathbf{G}_1'\hat{\mathbf{W}}\mathbf{G}_1\right)^{-1} \mathbf{G}_1'\hat{\mathbf{W}}\mathbf{g}_n(\hat{\boldsymbol{\theta}}_1)
\end{aligned}$$

The GN interative algorithm is then

$$\hat{\boldsymbol{\theta}}_{j+1} = \hat{\boldsymbol{\theta}}_j + \left(\mathbf{G}_j'\hat{\mathbf{W}}\mathbf{G}_j\right)^{-1} \mathbf{G}_j'\hat{\mathbf{W}}\mathbf{g}_n(\hat{\boldsymbol{\theta}}_j)$$

Common Convergence Criteria

- Stop when

$$
\begin{aligned}
||\hat{\boldsymbol{\theta}}_{j+1} - \hat{\boldsymbol{\theta}}_j|| \; &< \varepsilon \approx 10^{-6} \\
||\mathbf{x}|| &= \left(x_1^2 + \cdots + x_n^2\right)^{1/2}
\end{aligned}
$$

This criteria is sensitive to the units of $\boldsymbol{\theta}$. Better to replace ε by

$$
\eta\left(\left\|\hat{\boldsymbol{\theta}}_j\right\| + \tau\right),\ \eta \approx 10^{-5},\ \tau \approx 10^{-3}
$$

Then stop when

$$
\frac{||\hat{\boldsymbol{\theta}}_{j+1} - \hat{\boldsymbol{\theta}}_j||}{\left(\left\|\hat{\boldsymbol{\theta}}_j\right\| + \tau\right)} < \eta \approx 10^{-5}
$$

- Stop when

$$\left\| \frac{\partial J(\hat{\boldsymbol{\theta}}_j, \hat{\mathbf{W}})}{\partial \boldsymbol{\theta}'} \right\| < \varepsilon$$

- Stop when

$$|J(\hat{\boldsymbol{\theta}}_{j+1}, \hat{\mathbf{W}}) - J(\hat{\boldsymbol{\theta}}_j, \hat{\mathbf{W}})| \ < \varepsilon$$

  This criteria is sensitive to the units of $J(\hat{\boldsymbol{\theta}}_j, \hat{\mathbf{W}})$.Better to replace ε by

$$\eta \left(J(\hat{\boldsymbol{\theta}}_j, \hat{\mathbf{W}}) + \tau \right), \ \eta \approx 10^{-5}, \ \tau \approx 10^{-3}$$

  Then stop when

$$\frac{|J(\hat{\boldsymbol{\theta}}_{j+1}, \hat{\mathbf{W}}) - J(\hat{\boldsymbol{\theta}}_j, \hat{\mathbf{W}})|}{\left(J(\hat{\boldsymbol{\theta}}_j, \hat{\mathbf{W}}) + \tau \right)} < \eta \approx 10^{-5}$$

Example: Student's-t Distribution (Hamilton, 1994)

Consider a random sample $y_1, \ldots, y_T$ from a centered Student's-t distribution with θ_0 degrees of freedom with pdf

$$\begin{aligned} f(y_t;\theta_0) &= \frac{\Gamma[(\theta_0+1)/2]}{(\pi\theta_0)^{1/2}\Gamma(\theta_0/2)}[1+(y_t^2/\theta_0)]^{-(\theta_0+1)/2} \\ \Gamma(\cdot) &= \text{gamma function} \end{aligned}$$

The goal is to estimate the degrees of freedom parameter θ_0 by GMM using the moment conditions

$$\begin{aligned} E[y_t^2] &= \frac{\theta_0}{\theta_0-2} \\ E[y_t^4] &= \frac{3\theta_0^2}{(\theta_0-2)(\theta_0-4)},\ \theta_0 > 4 \end{aligned}$$

Let $\mathbf{w}_t = (y_t^2, y_t^4)'$ and define

$$\mathbf{g}(\mathbf{w}_t,\theta) = \begin{pmatrix} y_t^2 - \theta/(\theta-2) \\ y_t^4 - 3\theta^2/(\theta-2)(\theta-4) \end{pmatrix}$$

Then $E[\mathbf{g}(\mathbf{w}_t,\theta_0)] = \mathbf{0}$ is the moment condition used for defining the GMM estimator for θ_0.

Here, $K = 2$ and $p = 1$ so θ_0 is apparently overidentified.

Since we assume random sampling, in this example, $\mathbf{g}(\mathbf{w}_t, \theta_0)$ is an iid process.

Using the sample moments

$$\begin{aligned}\mathbf{g}_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n}\mathbf{g}(\mathbf{w}_t,\theta) \\ &= \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n} y_t^2 - \theta/(\theta-2) \\ \frac{1}{n}\sum_{t=1}^{n} y_t^4 - 3\theta^2/(\theta-2)(\theta-4) \end{pmatrix}\end{aligned}$$

the GMM objective function has the form

$$J(\theta) = n\mathbf{g}_n(\theta)'\hat{\mathbf{W}}\mathbf{g}_n(\theta)$$

where $\hat{\mathbf{W}}$ is a 2×2 p.d. and symmetric weight matrix, possibly dependent on the data, such that $\hat{\mathbf{W}}\xrightarrow{p}\mathbf{W}$.

The efficient GMM estimator uses the weight matrix $\hat{\mathbf{S}}^{-1}$ such that

$$\hat{\mathbf{S}} \xrightarrow{p} \mathbf{S} = E[\mathbf{g}(\mathbf{w}_t, \theta_0)\mathbf{g}(\mathbf{w}_t, \theta_0)']$$

For example, use

$$\begin{aligned}\hat{\mathbf{S}} &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{g}(\mathbf{w}_t, \hat{\theta})\mathbf{g}(\mathbf{w}_t, \hat{\theta})' \\ &\hat{\theta} \xrightarrow{p} \theta\end{aligned}$$

Remarks

1. In estimating the model, the restriction $\theta > 4$ should be imposed. This may be done by reparameterization. Define

$$\theta = h(\gamma) = \exp(\gamma) + 4, \quad -\infty < \gamma < \infty$$

and then estimate γ freely. Given

$$\sqrt{n}(\hat{\gamma} - \gamma_0) \xrightarrow{d} N(0, V)$$

a consistent and asymptotically normal estimate for θ, by Slutsky's theorem and the delta method, is $\hat{\theta} = h(\hat{\gamma}) = \exp(\hat{\gamma}) + 4$ where

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, h'(\gamma_0)^2 \times V)$$

2. Since the pdf of the data is known, the most efficient estimator is the maximum likelihood estimator.

Example: MA(1) model

The MA(1) model has the form

$$y_t = \mu_0 + \varepsilon_t + \psi_0 \varepsilon_{t-1},\ t = 1, \ldots, n$$
$$\varepsilon_t \sim iid\ (0, \sigma_0^2),\ |\psi_0| < 1$$
$$\boldsymbol{\theta}_0 = (\mu_0, \psi_0, \sigma_0^2)'$$

Some population moment equations that can be used for GMM estimation are

$$\begin{aligned}
E[y_t] &= \mu_0 \\
E[y_t^2] &= \gamma_0 + \mu_0^2 = \sigma_0^2(1 + \psi_0^2) + \mu_0^2 \\
E[y_t y_{t-1}] &= \gamma_1 + \mu_0^2 = \sigma_0^2 \psi_0 + \mu_0^2 \\
E[y_t y_{t-2}] &= \gamma_2 + \mu_0^2 = \mu_0^2 \\
E[y_t y_{t-j}] &= \gamma_j + \mu_0^2 = \mu_0^2
\end{aligned}$$

Let $\mathbf{w}_t = (y_t, y_t^2, y_t y_{t-1}, y_t y_{t-2})'$ and define the 4×1 moment vector

$$\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}) = \begin{pmatrix} y_t - \mu \\ y_t^2 - \mu^2 - \sigma^2(1 + \psi^2) \\ y_t y_{t-1} - \mu^2 - \sigma^2 \psi \\ y_t y_{t-2} - \mu^2 \end{pmatrix}$$

Then

$$E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$$

is the population moment condition used for GMM estimation of the model parameters $\boldsymbol{\theta}_0$.

Here, $p = 3$ and $K = 4$ the model is apparently overidentified.

The process $\{\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)\}$ will be autocorrelated (at least at lag 1) since y_t follows an MA(1) process.

The sample moments are

$$
\begin{aligned}
\mathbf{g}_n(\boldsymbol{\theta}) &= \frac{1}{n-2}\sum_{t=3}^{n} g(\mathbf{w}_t,\boldsymbol{\theta}) \\
&= \begin{pmatrix} \frac{1}{n-2}\sum_{t=3}^{n} y_t - \mu \\ \frac{1}{n-2}\sum_{t=3}^{n} y_t^2 - \mu^2 - \sigma^2(1+\psi^2) \\ \frac{1}{n-2}\sum_{t=3}^{n} y_t y_{t-1} - \mu^2 - \sigma^2\psi \\ \frac{1}{n-2}\sum_{t=3}^{n} y_t y_{t-2} - \mu^2 \end{pmatrix}
\end{aligned}
$$

The efficient GMM objective function has the form

$$
J(\boldsymbol{\theta}) = (n-2)\cdot \mathbf{g}_n(\boldsymbol{\theta})'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\boldsymbol{\theta})
$$

where $\hat{\mathbf{S}}$ is a consistent estimate of $\mathbf{S} = \text{avar}(\bar{\mathbf{g}}(\boldsymbol{\theta}_0))$.

Note: Since $\{g(\mathbf{w}_t,\boldsymbol{\theta})\}$ is autocorrelated $\mathbf{S} = \text{LRV}$

1. The process $\{\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)\}$ will be autocorrelated (at least at lag 1) since y_t follows an MA(1) process. As a result, an HAC type estimator must be used to estimate $\mathbf{S}$:

$$\begin{aligned}
\hat{\mathbf{S}}_{\text{HAC}} &= \hat{\boldsymbol{\Gamma}}_0(\hat{\boldsymbol{\theta}}) + \sum_{j=1}^{n-1} k\left(\frac{j}{q(n)}\right)(\hat{\boldsymbol{\Gamma}}_j(\hat{\boldsymbol{\theta}}) + \hat{\boldsymbol{\Gamma}}_j'(\hat{\boldsymbol{\theta}})) \\
\hat{\boldsymbol{\Gamma}}_j(\hat{\boldsymbol{\theta}}) &= \frac{1}{n}\sum_{t=j+1}^{n} \mathbf{g}_t(\hat{\boldsymbol{\theta}})\mathbf{g}_{t-j}(\hat{\boldsymbol{\theta}})'
\end{aligned}$$

2. Suppose it is known that $0 < \psi < 1$ and $\sigma^2 > 0$. These restrictions may be imposed using the reparameterization

$$\begin{aligned}
\psi &= \frac{\exp(\gamma_0)}{1+\exp(\gamma_0)}, \ -\infty < \gamma_0 < \infty \\
\sigma^2 &= \exp(\gamma_1), \ -\infty < \gamma_1 < \infty
\end{aligned}$$

3. If $\varepsilon_t \sim$ iid $N(0, \sigma^2)$ then the MLE is the efficient estimator.

Example: log-normal stochastic volatility model

The simple log-normal stochastic volatility (SV) model, due to Taylor (1986), is given by

$$
\begin{gathered}
y_t = \sigma_t Z_t,\ t = 1, \ldots, n \\
\ln \sigma_t^2 = \omega_0 + \beta_0 \ln \sigma_{t-1}^2 + \sigma_{0,u} u_t \\
(Z_t, u_t)' \sim \text{iid } N(\mathbf{0}, \mathbf{I}_2) \\
\boldsymbol{\theta}_0 = (\omega_0, \beta_0, \sigma_{0,u})'
\end{gathered}
$$

For $0 < \beta_0 < 1$ and $\sigma_{0,u} \geq 0$, the series y_t is strictly stationary and ergodic, and unconditional moments of all orders exist.

In the SV model, the series y_t is serially uncorrelated but dependency in the higher-order moments is induced by the serially correlated stochastic volatility term $\ln \sigma_t^2$.

The GMM estimation of the SV model is surveyed in Andersen and Sorensen (1996).

They recommended using moment conditions for GMM estimation based on lower-order moments of y_t, since higher-order moments tend to exhibit erratic finite sample behavior.

They considered a GMM estimation based on (subsets) of 24 moments considered by Jacquier, Polson, and Rossi (1994). To describe these moment conditions, first define

$$\mu = \frac{\omega}{1-\beta},\ \sigma^2 = \frac{\sigma_u^2}{1-\beta^2}$$

The moment conditions, which follow from properties of the log-normal distribution and the Gaussian AR(1) model, are expressed as

$$
\begin{aligned}
E[|y_t|] &= (2/\pi)^{1/2} E[\sigma_t] \\
E[y_t^2] &= E[\sigma_t^2] \\
E[|y_t^3|] &= 2\sqrt{2/\pi} E[\sigma_t^3] \\
E[y_t^4] &= 3E[\sigma_t^4] \\
E[|y_t y_{t-j}|] &= (2/\pi) E[\sigma_t \sigma_{t-j}],\ j = 1, \ldots, 10 \\
E[y_t^2 y_{t-j}^2] &= E[\sigma_t^2 \sigma_{t-j}^2],\ j = 1, \ldots, 10
\end{aligned}
$$

where for any positive integer j and positive constants r and s,

$$
\begin{aligned}
E[\sigma_t^r] &= \exp\left(\frac{r\mu}{2} + \frac{r^2\sigma^2}{8}\right) \\
E[\sigma_t^r \sigma_{t-j}^s] &= E[\sigma_t^r] E[\sigma_t^s] \exp\left(\frac{rs\beta^j \sigma^2}{4}\right)
\end{aligned}
$$

Let

$$\mathbf{w}_t = (|y_t|, y_t^2, |y_t^3|, y_t^4, |y_t y_{t-1}|, \ldots, |y_t y_{t-10}|, y_t^2 y_{t-1}^2, \ldots, y_t^2 y_{t-10}^2)'$$

and define the 24×1 vector

$$g(\mathbf{w}_t, \boldsymbol{\theta}) = \begin{pmatrix} |y_t| - (2/\pi)^{1/2} \exp\left(\frac{\mu}{2} + \frac{\sigma^2}{8}\right) \\ y_t^2 - \exp\left(\mu + \frac{\sigma^2}{2}\right) \\ \vdots \\ y_t^2 y_{t-10}^2 - \exp\left(\mu + \frac{\sigma^2}{2}\right)^2 \exp\left(\beta^{10}\sigma^2\right) \end{pmatrix}$$

Then, $E[g(\mathbf{w}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$ is the population moment condition used for the GMM estimation of the model parameters $\boldsymbol{\theta}_0$.

Since the elements of $\mathbf{w}_t$ are serially correlated, the efficient weight matrix $\mathbf{S} = \text{avar}(\bar{\mathbf{g}})$ must be estimated using an HAC estimator.

Example: Euler Equation Asset Pricing Model

A representative agent is assumed to choose an optimal consumption path by maximizing the present discounted value of lifetime utility from consumption

$$\max \sum_{t=1}^{\infty} E\left[\beta_0^t U(C_t) | I_t\right]$$

subject to the budget constraint

$$C_t + P_t Q_t \leq V_t Q_{t-1} + W_t$$

where I_t denotes the information available at time t, C_t denotes real consumption at t, W_t denotes real labor income at t, P_t denotes the price of a pure discount bond maturing at time $t+1$ that pays V_{t+1}, Q_t represents the quantity of bonds held at t, and β_0 represents a time discount factor.

The first order condition for the maximization problem may be represented as the conditional moment equation (Euler equation)

$$
\begin{aligned}
E\left[(1+R_{t+1})\beta_0 \frac{U'(C_{t+1})}{U'(C_t)} | I_t\right] - 1 &= 0 \\
1 + R_{t+1} &= \frac{V_{t+1}}{V_t}
\end{aligned}
$$

Assume a power utility function

$$
\begin{aligned}
U(C) &= \frac{C^{1-\alpha_0}}{1-\alpha_0} \\
\alpha_0 &= \text{risk aversion parameter}
\end{aligned}
$$

Then

$$\frac{U'(C_{t+1})}{U'(C_t)} = \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha_0}$$

and the conditional moment equation becomes

$$E\left[(1+R_{t+1})\beta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\alpha_0}|I_t\right] - 1 = 0$$

Define the nonlinear error term as

$$
\begin{aligned}
\varepsilon_{t+1} &= a(R_{t+1}, C_{t+1}/C_t; \alpha_0, \beta_0) \\
&= (1 + R_{t+1})\beta_0 \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha_0} - 1 \\
&= a(\mathbf{z}_{t+1}, \boldsymbol{\theta}_0) \\
\mathbf{z}_{t+1} &= (R_{t+1}, C_{t+1}/C_t)',\ \boldsymbol{\theta}_0 = (\alpha_0, \beta_0)'
\end{aligned}
$$

Then the conditional moment equation may be represented as

$$
E[\varepsilon_{t+1}|I_t] = E[a(\mathbf{z}_{t+1}, \boldsymbol{\theta}_0)|I_t] = 0
$$

Since $\{\varepsilon_{t+1}, I_{t+1}\}$ is a MDS, potential instruments $\mathbf{x}_t$ include current and lagged values of the elements in $\mathbf{z}_t$ as well as a constant. For example, one could use

$$\mathbf{x}_t = (1, C_t/C_{t-1}, C_{t-1}/C_{t-2}, R_t, R_{t-1})'$$

Since $\mathbf{x}_t \subset I_t$, the conditional moment implies that

$$E[\mathbf{x}_t \varepsilon_{t+1} | I_t] = E[\mathbf{x}_t a(\mathbf{z}_{t+1}, \boldsymbol{\theta}_0) | I_t] = \mathbf{0}$$

and by the law of iterated expectations the conditional moment equation implies the unconditional moment equation

$$E[\mathbf{x}_t \varepsilon_{t+1}] = \mathbf{0}$$

For GMM estimation, define the nonlinear residual as

$$e_{t+1} = (1 + R_{t+1})\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha} - 1$$

and form the 5×1 vector of moments

$$g(\mathbf{w}_{t+1}, \boldsymbol{\theta}) = \mathbf{x}_t e_{t+1} = \mathbf{x}_t a(\mathbf{z}_{t+1}, \boldsymbol{\theta})$$

$$= \begin{pmatrix} (1 + R_{t+1})\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha} - 1 \\ (C_t/C_{t-1}) \left((1 + R_{t+1})\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha} - 1\right) \\ (C_{t-1}/C_{t-2}) \left((1 + R_{t+1})\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha} - 1\right) \\ R_t \left((1 + R_{t+1})\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha} - 1\right) \\ R_{t-1} \left((1 + R_{t+1})\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha} - 1\right) \end{pmatrix}$$

There are $K = 5$ moment conditions to identify $L = 2$ model parameters giving $K - L = 3$ overidentifying restrictions.

Note: $\{g(\mathbf{w}_{t+1}, \boldsymbol{\theta}_0)\}$ is an ergodic-stationary MDS by assumption so that

$$\mathbf{S} = E[g(\mathbf{w}_{t+1}, \boldsymbol{\theta}_0)g(\mathbf{w}_{t+1}, \boldsymbol{\theta}_0)']$$

The GMM objective function is

$$J(\boldsymbol{\theta}, \hat{\mathrm{S}}^{-1}) = (n-2) \cdot \mathbf{g}_n(\boldsymbol{\theta})'\hat{\mathrm{S}}^{-1}\mathbf{g}_n(\boldsymbol{\theta})$$

where $\hat{\mathbf{S}}$ is a consistent estimate of $\mathbf{S} = \text{avar}(\bar{\mathbf{g}})$.

Since $\{g(\mathbf{w}_{t+1}, \boldsymbol{\theta}_0)\}$ is a MDS, $\mathbf{S}$ may be estimated using

$$\begin{aligned} \hat{\mathbf{S}} &= \frac{1}{n-2}\sum_{t=3}^{n} \mathbf{g}(\mathbf{w}_{t+1}, \hat{\boldsymbol{\theta}})\mathbf{g}(\mathbf{w}_{t+1}, \hat{\boldsymbol{\theta}})' \\ & \hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta} \end{aligned}$$

An extension of the model allows the individual to invest in J risky assets with returns $R_{j,t+1}$ ($j = 1, \ldots, J$), as well as a risk-free asset with certain return $R_{f,t+1}$.

Assuming power utility and restricting attention to unconditional moments (i.e., using $\mathbf{x}_t = 1$), the Euler equations may be written as

$$
\begin{aligned}
E\left[(1 + R_{f,t+1})\beta_0 \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha_0}\right] - 1 &= 0 \\
E\left[(R_{j,t+1} - R_{f,t+1})\beta_0 \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha_0}\right] &= 0,\ j = 1, \ldots, J
\end{aligned}
$$

For the GMM estimation, one may use the $J+1$ vector of moments

$$g(\mathbf{w}_{t+1}, \boldsymbol{\theta}) = \begin{pmatrix} (1 + R_{f,t+1})\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha} - 1 \\ (R_{1,t+1} - R_{f,t+1})\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha} \\ \vdots \\ (R_{J,t+1} - R_{f,t+1})\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha} \end{pmatrix}$$

Example: Interest Rate Diffusion Model

Consider estimating the parameters of the continuous-time interest rate diffusion model

$$\begin{aligned} dr_t &= (\alpha_0 + \beta_0 r_t)dt + \sigma_0 r_t^{\gamma_0} dW_t \\ \boldsymbol{\theta}_0 &= (\alpha_0, \beta_0, \sigma_0, \gamma_0)' \\ W_t &= \text{Wiener process} \end{aligned}$$

Moment conditions for GMM estimation of $\boldsymbol{\theta}_0$ may be derived from the Euler discretization

$$r_{t+\Delta t} - r_t = (\alpha_0 + \beta_0 r_t)\Delta t + \sigma_0 r_t^{\gamma_0} \sqrt{\Delta t} z_{t+\Delta t}$$
$$z_{t+\Delta t} \sim N(0,1),\ E[z_{t+\Delta t}] = 0,\ E[z_{t+\Delta t}^2] = 1$$

Define the true model error as

$$\begin{aligned}
\varepsilon_{t+\Delta t} &= a(r_{t+\Delta t} - r_t, r_t; \alpha_0, \beta_0, \sigma_0, \gamma_0) \\
&= (r_{t+\Delta t} - r_t) - (\alpha_0 + \beta_0 r_t)\Delta t \\
&= \sigma_0 r_t^{\gamma_0} \sqrt{\Delta t} z_{t+\Delta t} = a(\mathbf{z}_{t+\Delta t}, \boldsymbol{\theta}_0) \\
\mathbf{z}_{t+\Delta t} &= (r_{t+\Delta t} - r_t, r_t)'
\end{aligned}$$

Letting I_t represent information available at time t, the true error satisfies $E[\varepsilon_{t+\Delta t}|I_t] = 0$.

Since $\{\varepsilon_{t+\Delta t}, I_{t+\Delta}\}$ is a MDS, potential instruments $\mathbf{x}_t$ include current and lagged values of the elements of $\mathbf{z}_t$ as well as a constant. Using $\mathbf{x}_t = (1, r_t)'$ as the instrument vector, the following four conditional moments may be deduced

$$\begin{aligned}
E[\varepsilon_{t+\Delta t}|I_t] &= 0,\ E[\varepsilon_{t+\Delta t}^2|I_t] = \sigma_0^2 r_t^{2\gamma_0} \Delta t \\
E[\varepsilon_{t+\Delta t} r_t|I_t] &= 0,\ E[\varepsilon_{t+\Delta t}^2 r_t|I_t] = \sigma_0^2 r_t^{2\gamma_0} \Delta t \cdot r_t
\end{aligned}$$

For given values of α and β define the nonlinear residual

$$e_{t+\Delta t} = (r_{t+\Delta t} - r_t) - (\alpha + \beta r_t)\Delta t$$

and, for $\mathbf{w}_{t+\Delta t} = (r_{t+\Delta t} - r_t, r_t, r_t^2)'$, define the 4×1 vector of moments

$$\mathbf{g}(\mathbf{w}_{t+\Delta t}, \boldsymbol{\theta}) = \begin{pmatrix} e_{t+\Delta t} \\ e_{t+\Delta t}^2 \end{pmatrix} \otimes \mathbf{x}_t = \begin{pmatrix} e_{t+\Delta t} \\ e_{t+\Delta t} r_t \\ e_{t+\Delta t}^2 - \sigma^2 r_t^{2\gamma} \Delta t \\ \left(e_{t+\Delta t}^2 - \sigma^2 r_t^{2\gamma} \Delta t\right) r_t \end{pmatrix}$$

Then $E[\mathbf{g}(\mathbf{w}_{t+\Delta t}, \boldsymbol{\theta}_0)] = \mathbf{0}$ gives the GMM estimating equation.

Even though $\{\varepsilon_t, I_t\}$ is a MDS, the moment vector $\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)$ is likely to be autocorrelated since it contains ε_t^2. However, since $K = L = 4$,the model is just identified and so the GMM objective function does not depend on a weight matrix:

$$J(\boldsymbol{\theta}) = n\mathbf{g}_n(\boldsymbol{\theta})'\mathbf{g}_n(\boldsymbol{\theta})$$