

...

Econometric Methods for the Analysis of Dynamic General Equilibrium Models

Overview

- Multiple Equation Methods

Overview

- Multiple Equation Methods
 - State space-observer form

Overview

- Multiple Equation Methods
 - State space-observer form
 - Three Examples of Versatility of state space-observer form:
 - * Smoothing and filtering (estimation of ‘output gap’, ‘real interest rate’)
 - * Handling mixed monthly/quarterly data.
 - * Connection between DSGE models and Vector Autoregressions.

Overview

- Multiple Equation Methods
 - State space-observer form
 - Three Examples of Versatility of state space-observer form:
 - * Smoothing and filtering (estimation of ‘output gap’, ‘real interest rate’)
 - * Handling mixed monthly/quarterly data.
 - * Connection between DSGE models and Vector Autoregressions.
 - ‘Limited information estimation’: impulse response function matching
 - * Impulse response functions
 - * Formal connection between VARs and DSGE models

Overview

- Multiple Equation Methods
 - State space-observer form
 - Three Examples of Versatility of state space-observer form:
 - * Smoothing and filtering (estimation of ‘output gap’, ‘real interest rate’)
 - * Handling mixed monthly/quarterly data.
 - * Connection between DSGE models and Vector Autoregressions.
 - ‘Limited information estimation’: impulse response function matching
 - * Impulse response functions
 - * Formal connection between VARs and DSGE models
 - ‘Full information estimation’
 - * Maximum likelihood
 - * Bayesian inference

Overview

- Multiple Equation Methods
 - State space-observer form
 - Three Examples of Versatility of state space-observer form:
 - * Smoothing and filtering (estimation of ‘output gap’, ‘real interest rate’)
 - * Handling mixed monthly/quarterly data.
 - * Connection between DSGE models and Vector Autoregressions.
 - ‘Limited information estimation’: impulse response function matching
 - * Impulse response functions
 - * Formal connection between VARs and DSGE models
 - ‘Full information estimation’
 - * Maximum likelihood
 - * Bayesian inference
- Single equation (‘limited information’) methods: Introduction to Generalized Method of Moments

State Space-Observer Form

- Compact summary of the model, and of the data used in the analysis.
- Typically, data are available in log form. So, the following is useful:

– If x is steady state of x_t :

$$\begin{aligned}\hat{x}_t &\equiv \frac{x_t - x}{x}, \\ \implies \frac{x_t}{x} &= 1 + \hat{x}_t \\ \implies \log\left(\frac{x_t}{x}\right) &= \log(1 + \hat{x}_t) \approx \hat{x}_t\end{aligned}$$

- Suppose we have a model solution in hand:

$$\begin{aligned}z_t &= Az_{t-1} + Bs_t \\ s_t &= Ps_{t-1} + \epsilon_t, \quad E\epsilon_t\epsilon_t' = D.\end{aligned}$$

State Space-Observer Form ...

- Consider example #3 in solution notes, in which

$$z_t = \begin{pmatrix} \hat{K}_{t+1} \\ \hat{N}_t \end{pmatrix}, \quad s_t = \hat{\varepsilon}_t, \quad \epsilon_t = e_t.$$

Data used in analysis may include variables in z_t and/or other variables.

- Suppose variables of interest include employment and *GDP*.
 - GDP, y_t :

$$y_t = \varepsilon_t K_t^\alpha N_t^{1-\alpha},$$

so that

$$\begin{aligned} \hat{y}_t &= \hat{\varepsilon}_t + \alpha \hat{K}_t + (1 - \alpha) \hat{N}_t \\ &= \begin{pmatrix} 0 & 1 - \alpha \end{pmatrix} z_t + \begin{pmatrix} \alpha & 0 \end{pmatrix} z_{t-1} + s_t \end{aligned}$$

– Then,

$$Y_t^{data} = \begin{pmatrix} \log y_t \\ \log N_t \end{pmatrix} = \begin{pmatrix} \log y \\ \log N \end{pmatrix} + \begin{pmatrix} \hat{y}_t \\ \hat{N}_t \end{pmatrix}$$

State Space-Observer Form ...

- Model prediction for data:

$$\begin{aligned} Y_t^{data} &= \begin{pmatrix} \log y \\ \log N \end{pmatrix} + \begin{pmatrix} \hat{y}_t \\ \hat{N}_t \end{pmatrix} \\ &= \begin{pmatrix} \log y \\ \log N \end{pmatrix} + \begin{bmatrix} 0 & 1 & -\alpha \\ 0 & & 1 \end{bmatrix} z_t + \begin{bmatrix} \alpha & 0 \\ 0 & 0 \end{bmatrix} z_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} s_t \\ &= a + H\xi_t \end{aligned}$$

$$\xi_t = \begin{pmatrix} z_t \\ z_{t-1} \\ \hat{\varepsilon}_t \end{pmatrix}, \quad a = \begin{bmatrix} \log y \\ \log N \end{bmatrix}, \quad H = \begin{bmatrix} 0 & 1 & -\alpha & \alpha & 0 & 1 \\ 0 & & 1 & 0 & 0 & 0 \end{bmatrix}$$

- The *Observer Equation* may include measurement error, w_t :

$$Y_t^{data} = a + H\xi_t + w_t, \quad Ew_t w_t' = R.$$

- Semantics: ξ_t is the *state* of the system (do not confuse with the economic state (K_t, ε_t) !).

State Space-Observer Form ...

- The state equation

- Law of motion of the state, ξ_t

$$\xi_t = F\xi_{t-1} + u_t, \quad Eu_tu_t' = Q$$

$$\begin{pmatrix} z_{t+1} \\ z_t \\ s_{t+1} \end{pmatrix} = \begin{bmatrix} A & 0 & BP \\ I & 0 & 0 \\ 0 & 0 & P \end{bmatrix} \begin{pmatrix} z_t \\ z_{t-1} \\ s_t \end{pmatrix} + \begin{pmatrix} B \\ 0 \\ I \end{pmatrix} \epsilon_{t+1},$$

$$u_t = \begin{pmatrix} B \\ 0 \\ I \end{pmatrix} \epsilon_t, \quad Q = \begin{bmatrix} BDB' & 0 & BD \\ 0 & 0 & 0 \\ DB' & & D \end{bmatrix}, \quad F = \begin{bmatrix} A & 0 & BP \\ I & 0 & 0 \\ 0 & 0 & P \end{bmatrix}.$$

State Space-Observer Form ...

- Summary: State-Space, Observer System -

$$\xi_t = F\xi_{t-1} + u_t, \quad Eu_tu_t' = Q,$$

$$Y_t^{data} = a + H\xi_t + w_t, \quad Ew_tw_t' = R.$$

- Can be constructed from model parameters

$$\theta = (\beta, \delta, \dots)$$

so

$$F = F(\theta), \quad Q = Q(\theta), \quad a = a(\theta), \quad H = H(\theta), \quad R = R(\theta).$$

State Space-Observer Form ...

- State space observer system very useful

- Estimation of θ and forecasting ξ_t and Y_t^{data}
- Can take into account situations in which data represent a mixture of quarterly, monthly, daily observations.
- Software readily available on web and elsewhere.

– Useful for solving the following forecasting problems:

* Filtering:

$$P [\xi_t | Y_{t-1}^{data}, Y_{t-2}^{data}, \dots, Y_1^{data}], \quad t = 1, 2, \dots, T.$$

* Smoothing:

$$P [\xi_t | Y_T^{data}, \dots, Y_1^{data}], \quad t = 1, 2, \dots, T.$$

* Example: ‘real rate of interest’ and ‘output gap’ can be recovered from ξ_t using example #5 in solution notes.

Mixed Monthly/Quarterly Observations

- Different data arrive at different frequencies: daily, monthly, quarterly, etc.
- This feature can be easily handled in state space-observer system.
- Example:
 - suppose inflation and hours are monthly, $t = 0, 1/3, 2/3, 1, 4/3, 5/3, 2, \dots$
 - suppose gdp is quarterly, $t = 0, 1, 2, 3, \dots$

$$Y_t^{data} = \begin{pmatrix} GDP_t \\ \text{monthly inflation}_t \\ \text{monthly inflation}_{t-1/3} \\ \text{monthly inflation}_{t-2/3} \\ \text{hours}_t \\ \text{hours}_{t-1/3} \\ \text{hours}_{t-2/3} \end{pmatrix}, t = 0, 1, 2, \dots$$

that is, we can think of our data set as actually being quarterly, with quarterly observations on the first month's inflation, quarterly observations on the second month's inflation, etc.

Mixed Monthly/Quarterly Observations ...

- Problem: find state-space observer system in which observed data are:

$$Y_t^{data} = \begin{pmatrix} GDP_t \\ \text{monthly inflation}_t \\ \text{monthly inflation}_{t-1/3} \\ \text{monthly inflation}_{t-2/3} \\ \text{hours}_t \\ \text{hours}_{t-1/3} \\ \text{hours}_{t-2/3} \end{pmatrix}, t = 0, 1, 2, \dots$$

- Solution: easy!

Mixed Monthly/Quarterly Observations ...

- Model: specified at a monthly level, with solution, $t = 0, 1/3, 2/3, \dots$

$$\begin{aligned} z_t &= Az_{t-1/3} + Bs_t, \\ s_t &= Ps_{t-1/3} + \epsilon_t, \quad E\epsilon_t\epsilon_t' = D. \end{aligned}$$

- Monthly state-space observer system, $t = 0, 1/3, 2/3, \dots$

$$\xi_t = F\xi_{t-1/3} + u_t, \quad Eu_tu_t' = Q, \quad u_t \sim iid \quad t = 0, 1/3, 2/3, \dots$$

$$Y_t = H\xi_t, \quad Y_t = \begin{pmatrix} y_t \\ \pi_t \\ h_t \end{pmatrix}.$$

- Note:

first order vector autoregressive representation for quarterly state

$$\xi_t = \overbrace{F^3\xi_{t-1} + u_t + Fu_{t-1/3} + F^2u_{t-2/3}} \quad ,$$

$$u_t + Fu_{t-1/3} + F^2u_{t-2/3} \sim \underline{iid \text{ for } t = 0, 1, 2, \dots!!}$$

Mixed Monthly/Quarterly Observations ...

- Consider the following system:

$$\begin{pmatrix} \xi_t \\ \xi_{t-1/3} \\ \xi_{t-2/3} \end{pmatrix} = \begin{bmatrix} F^3 & 0 & 0 \\ F^2 & 0 & 0 \\ F & 0 & 0 \end{bmatrix} \begin{pmatrix} \xi_{t-1} \\ \xi_{t-4/3} \\ \xi_{t-5/3} \end{pmatrix} + \begin{bmatrix} I & F & F^2 \\ 0 & I & F \\ 0 & 0 & I \end{bmatrix} \begin{pmatrix} u_t \\ u_{t-1/3} \\ u_{t-2/3} \end{pmatrix}.$$

- Define

$$\tilde{\xi}_t = \begin{pmatrix} \xi_t \\ \xi_{t-1/3} \\ \xi_{t-2/3} \end{pmatrix}, \tilde{F} = \begin{bmatrix} F^3 & 0 & 0 \\ F^2 & 0 & 0 \\ F & 0 & 0 \end{bmatrix}, \tilde{u}_t = \begin{bmatrix} I & F & F^2 \\ 0 & I & F \\ 0 & 0 & I \end{bmatrix} \begin{pmatrix} u_t \\ u_{t-1/3} \\ u_{t-2/3} \end{pmatrix},$$

so that

$$\tilde{\xi}_t = \tilde{F}\tilde{\xi}_{t-1} + \tilde{u}_t, \tilde{u}_t \sim iid \text{ in quarterly data, } t = 0, 1, 2, \dots$$

$$E\tilde{u}_t\tilde{u}_t' = \tilde{Q} = \begin{bmatrix} I & F & F^2 \\ 0 & I & F \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} D & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & D \end{bmatrix} \begin{bmatrix} I & F & F^2 \\ 0 & I & F \\ 0 & 0 & I \end{bmatrix}'$$

Mixed Monthly/Quarterly Observations ...

- Done!
- State space-observer system for mixed monthly/quarterly data, for $t = 0, 1, 2, \dots$

$$\tilde{\xi}_t = \tilde{F}\tilde{\xi}_{t-1} + \tilde{u}_t, \tilde{u}_t \sim iid, E\tilde{u}_t\tilde{u}_t' = \tilde{Q},$$

$$Y_t^{data} = \tilde{H}\tilde{\xi}_t + w_t, w_t \sim iid, Ew_t w_t' = R.$$

- Here, \tilde{H} selects elements of $\tilde{\xi}_t$ to construct Y_t^{data}
 - can easily handle distinction between whether quarterly data represent monthly averages (as in flow variables), or point-in-time observations on one month in the quarter (as in stock variables).
- Can use Kalman filter to forecast current quarter data based on first month's (day's, week's) observations.

Matching Impulse Response Functions

- Set $\epsilon_t = 1$ for $t = 1$, $\epsilon_t = 0$ otherwise
- Impulse response function: log deviation of data with shock from where data would have been in the absence of a shock -

$$u_t = \begin{pmatrix} B \\ 0 \\ I \end{pmatrix} \epsilon_t,$$

$$\xi_t = F\xi_{t-1} + u_t, \quad \xi_0 = 0,$$

impulse response function $\implies \tilde{Y}_t^{data} = H\xi_t$, for $t = 1, 2, \dots$

- Choose model parameters, θ , to match \tilde{Y}_t^{data} with corresponding estimate from VAR (more on this later).

Connection Between DSGE's and VAR's

- Fernandez-Villaverde, Rubio-Ramirez, Sargent Result
- Vector Autoregression

$$Y_t = B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + u_t,$$

where u_t is iid. ‘Matching impulse response functions’ strategy for building DSGE models fits VARs and assumes u_t are a rotation of economic shocks (for details, see later notes).

- Can use the state space, observer representation to assess this assumption from the perspective of a DSGE.

Connection Between DSGE's and VAR's ...

- System (ignoring constant terms and measurement error):

$$\text{('State equation')} \quad \xi_t = F\xi_{t-1} + D\epsilon_t, \quad D = \begin{pmatrix} B \\ 0 \\ I \end{pmatrix},$$

$$\text{('Observer equation')} \quad Y_t = H\xi_t.$$

- Substituting:

$$Y_t = HF\xi_{t-1} + HD\epsilon_t$$

- Suppose HD is square and invertible. Then

$$\epsilon_t = (HD)^{-1} Y_t - (HD)^{-1} HF\xi_{t-1} (**)$$

- Substitute latter into the state equation:

$$\begin{aligned} \xi_t &= F\xi_{t-1} + D(HD)^{-1} Y_t - D(HD)^{-1} HF\xi_{t-1} \\ &= \left[I - D(HD)^{-1} H \right] F\xi_{t-1} + D(HD)^{-1} Y_t. \end{aligned}$$

Connection Between DSGE's and VAR's ...

- We have:

$$\xi_t = M\xi_{t-1} + D(HD)^{-1}Y_t, \quad M = \left[I - D(HD)^{-1}H \right] F.$$

- If eigenvalues of M are less than unity,

$$\xi_t = D(HD)^{-1}Y_t + MD(HD)^{-1}Y_{t-1} + M^2D(HD)^{-1}Y_{t-2} + \dots$$

- Substituting into (**)

$$\epsilon_t = (HD)^{-1}Y_t - (HD)^{-1}HF \left[D(HD)^{-1}Y_{t-1} + MD(HD)^{-1}Y_{t-2} + M^2D(HD)^{-1}Y_{t-3} + \dots \right]$$

or,

Connection Between DSGE's and VAR's ...

$$Y_t = B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + u_t,$$

where

$$u_t = HD\epsilon_t$$
$$B_j = HFM^{j-1}D(HD)^{-1}, j = 1, 2, \dots$$

- The latter is the VAR representation.
 - Note: ϵ_t is ‘invertible’ because it lies in space of current and past Y_t 's.
 - Note: VAR is *infinite*-ordered.
 - Note: assumed system is ‘square’. Sims-Zha (Macroeconomic Dynamics) show that square-ness is not necessary.

Maximum Likelihood Estimation

- State space-observer system:

$$\xi_{t+1} = F\xi_t + u_{t+1}, \quad Eu_t u_t' = Q,$$

$$Y_t^{data} = a_0 + H\xi_t + w_t, \quad Ew_t w_t' = R$$

- Parameters of system: (F, Q, a_0, H, R) . These are functions of model parameters, θ .
- Formulas for computing likelihood

$$P(Y^{data}|\theta) = P(Y_1^{data}, \dots, Y_T^{data}|\theta).$$

are standard (see Hamilton's textbook).

Bayesian Maximum Likelihood

- Bayesians describe the mapping from prior beliefs about θ , summarized in $p(\theta)$, to new posterior beliefs in the light of observing the data, Y^{data} .
- General property of probabilities:

$$p(Y^{data}, \theta) = \begin{cases} p(Y^{data}|\theta) \times p(\theta) \\ p(\theta|Y^{data}) \times p(Y^{data}) \end{cases},$$

which implies Bayes' rule:

$$p(\theta|Y^{data}) = \frac{p(Y^{data}|\theta) p(\theta)}{p(Y^{data})},$$

mapping from prior to posterior induced by Y^{data} .

Bayesian Maximum Likelihood ...

- Properties of the posterior distribution, $p(\theta|Y^{data})$.
 - The value of θ that maximizes $p(\theta|Y^{data})$ ('mode' of posterior distribution).
 - Graphs that compare the marginal posterior distribution of individual elements of θ with the corresponding prior.
 - Probability intervals about the mode of θ ('Bayesian confidence intervals')
 - Other properties of $p(\theta|Y^{data})$ helpful for assessing model 'fit'.

Bayesian Maximum Likelihood ...

- Computation of mode sometimes referred to as ‘Bayesian maximum likelihood’:

$$\theta^{\text{mode}} = \arg \max_{\theta} \left\{ \log [p(Y^{\text{data}}|\theta)] + \sum_{i=1}^N \log [p_i(\theta_i)] \right\}$$

maximum likelihood with a penalty function.

- Shape of posterior distribution, $p(\theta|Y^{\text{data}})$, obtained by Metropolis-Hastings algorithm.
 - Algorithm computes

$$\theta(1), \dots, \theta(N),$$

which, as $N \rightarrow \infty$, has a density that approximates $p(\theta|Y^{\text{data}})$ well.

- Marginal posterior distribution of any element of θ displayed as the histogram of the corresponding element $\{\theta(i), i = 1, \dots, N\}$

Metropolis-Hastings Algorithm (MCMC)

- We have (except for a constant):

$$f \left(\underbrace{\theta}_{N \times 1} | Y \right) = \frac{f(Y|\theta) f(\theta)}{f(Y)}.$$

- We want the marginal posterior distribution of θ_i :

$$h(\theta_i | Y) = \int_{\theta_{j \neq i}} f(\theta | Y) d\theta_{j \neq i}, \quad i = 1, \dots, N.$$

- MCMC algorithm can approximate $h(\theta_i | Y)$.
- Obtain (V produced automatically by gradient-based maximization methods):

$$\theta^{\text{mode}} \equiv \theta^* = \arg \max_{\theta} f(Y|\theta) f(\theta), \quad V \equiv \left[-\frac{\partial^2 f(Y|\theta) f(\theta)}{\partial \theta \partial \theta'} \right]_{\theta=\theta^*}^{-1}.$$

Metropolis-Hastings Algorithm (MCMC) ...

- Compute the sequence, $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(M)}$ (M large) whose distribution turns out to have pdf $f(\theta|Y)$.

– $\theta^{(1)} = \theta^*$

– to compute $\theta^{(r)}$, for $r > 1$

* step 1: select candidate $\theta^{(r)}$, x ,

‘jump’ distribution

$$\text{draw } \underbrace{x}_{N \times 1} \text{ from } \theta^{(r-1)} + kN \left(\underbrace{0}_{N \times 1}, V \right), \text{ } k \text{ is a scalar}$$

* step 2: compute scalar, λ :

$$\lambda = \frac{f(Y|x) f(x)}{f(Y|\theta^{(r-1)}) f(\theta^{(r-1)})}$$

* step 3: compute $\theta^{(r)}$:

$$\theta^{(r)} = \begin{cases} \theta^{(r-1)} & \text{if } u > \lambda \\ x & \text{if } u < \lambda \end{cases}, \text{ } u \text{ is a realization from uniform } [0, 1]$$

Metropolis-Hastings Algorithm (MCMC) ...

- Approximating marginal posterior distribution, $h(\theta_i|Y)$, of θ_i
 - compute and display the histogram of $\theta_i^{(1)}, \theta_i^{(2)}, \dots, \theta_i^{(M)}$, $i = 1, \dots, N$.
- Other objects of interest:
 - mean and variance of posterior distribution θ :

$$E\theta \simeq \bar{\theta} \equiv \frac{1}{M} \sum_{j=1}^M \theta^{(j)}, \quad Var(\theta) \simeq \frac{1}{M} \sum_{j=1}^M [\theta^{(j)} - \bar{\theta}] [\theta^{(j)} - \bar{\theta}]'$$

- marginal density of Y (actually, Geweke's 'harmonic mean' works better):

$$f(Y) \simeq \frac{1}{M} \sum_{j=1}^M f(Y|\theta^{(j)}) f(\theta^{(j)})$$

Metropolis-Hastings Algorithm (MCMC) ...

- Some intuition

- Algorithm is more likely to select moves into high probability regions than into low probability regions.

- Set, $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(M)}\}$, populated relatively more by elements near mode of $f(\theta|Y)$.

- Set, $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(M)}\}$, also populated (though less so) by elements far from mode of $f(\theta|Y)$.

Metropolis-Hastings Algorithm (MCMC) ...

- Practical issues

- what value should you set k to?

- * set k so that you accept (i.e., $\theta^{(r)} = x$) in step 3 of MCMC algorithm are roughly 27 percent of time

- what value of M should you set?

- * a value so that if M is increased further, your results do not change

- in practice, $M = 10,000$ (a small value) up to $M = 1,000,000$.

- large M is time-consuming. Could use Laplace approximation (after checking its accuracy) in initial phases of research project.

Laplace Approximation to Posterior Distribution

- In practice, Metropolis-Hasting algorithm very time intensive. Do it last!
- In practice, Laplace approximation is quick, essentially free and very accurate.
- Let $\theta \in R^N$ denote the N -dimensional vector of parameters and

$$g(\theta) \equiv \log f(y|\theta) f(\theta),$$

$f(y|\theta)$ ~likelihood of data

$f(\theta)$ ~prior on parameters

θ^* ~maximum of $g(\theta)$ (i.e., mode)

Laplace Approximation to Posterior Distribution ...

- Second order Taylor series expansion about $\theta = \theta^*$:

$$g(\theta) \approx g(\theta^*) + g_\theta(\theta^*)(\theta - \theta^*) - \frac{1}{2}(\theta - \theta^*)' g_{\theta\theta}(\theta^*)(\theta - \theta^*),$$

where

$$g_{\theta\theta}(\theta^*) = -\frac{\partial^2 \log f(y|\theta) f(\theta)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta^*}$$

- Interior optimality implies:

$$g_\theta(\theta^*) = 0, \quad g_{\theta\theta}(\theta^*) \text{ positive definite}$$

- Then,

$$f(y|\theta) f(\theta) \simeq f(y|\theta^*) f(\theta^*) \exp \left\{ -\frac{1}{2}(\theta - \theta^*)' g_{\theta\theta}(\theta^*)(\theta - \theta^*) \right\}.$$

Laplace Approximation to Posterior Distribution ...

- Note

$$\frac{1}{(2\pi)^{\frac{N}{2}}} |g_{\theta\theta}(\theta^*)|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\theta - \theta^*)' g_{\theta\theta}(\theta^*) (\theta - \theta^*) \right\}$$

= multinormal density for N – dimensional random variable θ

with mean θ^* and variance $g_{\theta\theta}(\theta^*)^{-1}$.

- So, posterior of θ_i (i.e., $h(\theta_i|Y)$) is approximately

$$\theta_i \sim N \left(\theta_i^*, \left[g_{\theta\theta}(\theta^*)^{-1} \right]_{ii} \right).$$

- This formula for the posterior distribution is essentially free, because $g_{\theta\theta}$ is computed as part of gradient-based numerical optimization procedures.

Laplace Approximation to Posterior Distribution ...

- Marginal likelihood of data, y , is useful for model comparisons. Easy to compute using the Laplace approximation.
- Property of Normal distribution:

$$\int \frac{1}{(2\pi)^{\frac{N}{2}}} |g_{\theta\theta}(\theta^*)|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\theta - \theta^*)' g_{\theta\theta}(\theta^*) (\theta - \theta^*) \right\} d\theta = 1$$

- Then,

$$\begin{aligned} \int f(y|\theta) f(\theta) d\theta &\simeq \int f(y|\theta^*) f(\theta^*) \exp \left\{ -\frac{1}{2} (\theta - \theta^*)' g_{\theta\theta}(\theta^*) (\theta - \theta^*) \right\} d\theta \\ &= \frac{f(y|\theta^*) f(\theta^*)}{\frac{1}{(2\pi)^{\frac{N}{2}}} |g_{\theta\theta}(\theta^*)|^{\frac{1}{2}}} \int \frac{1}{(2\pi)^{\frac{N}{2}}} |g_{\theta\theta}(\theta^*)|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\theta - \theta^*)' g_{\theta\theta}(\theta^*) (\theta - \theta^*) \right\} d\theta \\ &= \frac{f(y|\theta^*) f(\theta^*)}{\frac{1}{(2\pi)^{\frac{N}{2}}} |g_{\theta\theta}(\theta^*)|^{\frac{1}{2}}}. \end{aligned}$$

Laplace Approximation to Posterior Distribution ...

- Formula for marginal likelihood based on Laplace approximation:

$$f(y) = \int f(y|\theta) f(\theta) d\theta \simeq (2\pi)^{\frac{N}{2}} \frac{f(y|\theta^*) f(\theta^*)}{|g_{\theta\theta}(\theta^*)|^{\frac{1}{2}}}.$$

- Suppose $f(y|Model\ 1) > f(y|Model\ 2)$. Then, posterior odds on Model 1 higher than Model 2.
- ‘Model 1 fits better than Model 2’
- Can use this to compare across two different models, or to evaluate contribution to fit of various model features: habit persistence, adjustment costs, etc.

Generalized Method of Moments

- Express your econometric estimator into Hansen's GMM framework and you get standard errors
 - Essentially, *any* estimation strategy fits (see Hamilton)
- Works when parameters of interest, β , have the following property:

$$E \underbrace{u_t}_{N \times 1} \left(\underbrace{\beta}_{n \times 1} \right) = 0, \beta \text{ true value of some parameter(s) of interest}$$

$u_t(\beta) \sim$ stationary stochastic process (and other conditions)

– $n = N$: 'exactly identified'

– $n < N$: 'over identified'

Generalized Method of Moments ...

– Example 1: mean

$$\beta = E x_t,$$

$$u_t(\beta) = \beta - x_t.$$

– Example 2: mean and variance

$$\beta = [\mu \ \sigma],$$

$$E x_t = \mu, E (x_t - \mu)^2 = \sigma^2.$$

then,

$$u_t(\beta) = \begin{bmatrix} \mu - x_t \\ (x_t - \mu)^2 - \sigma^2 \end{bmatrix}.$$

Generalized Method of Moments ...

- Example 3: mean, variance, correlation, relative standard deviation

$$\beta = [\mu_y \ \sigma_y \ \mu_x \ \sigma_x \ \rho_{xy} \ \lambda], \quad \lambda \equiv \sigma_x / \sigma_y,$$

where

$$E y_t = \mu_y, \quad E (y_t - \mu_y)^2 = \sigma_y^2$$

$$E x_t = \mu_x, \quad E (x_t - \mu_x)^2 = \sigma_x^2$$

$$\rho_{xy} = \frac{E (y_t - \mu_y) (x_t - \mu_x)}{\sigma_y \sigma_x}.$$

then

$$u_t(\beta) = \begin{bmatrix} \mu_x - x_t \\ (x_t - \mu_x)^2 - \sigma_x^2 \\ \mu_y - y_t \\ (y_t - \mu_y)^2 - \sigma_y^2 \\ \sigma_y \sigma_x \rho_{xy} - (y_t - \mu_y) (x_t - \mu_x) \\ \sigma_y \lambda - \sigma_x \end{bmatrix}.$$

Generalized Method of Moments ...

– Example 4: New Keynesian Phillips curve

$$\pi_t = 0.99E_t\pi_{t+1} + \gamma s_t,$$

or,

$$\pi_t - 0.99\pi_{t+1} - \gamma s_t = \eta_{t+1}$$

where,

$$\eta_{t+1} = 0.99(E_t\pi_{t+1} - \pi_{t+1}) \implies E_t\eta_{t+1} = 0$$

Under Rational Expectations : $\eta_{t+1} \perp$ time t information, z_t

$$u_t(\gamma) = [\pi_t - 0.99\pi_{t+1} - \gamma s_t] z_t$$

Generalized Method of Moments ...

- Inference about β

- Estimator of β in exactly identified case ($n = N$)

- * Choose $\hat{\beta}$ to mimick population property of true β ,

$$Eu_t(\beta) = 0.$$

- * Define:

$$g_T(\beta) = \frac{1}{T} \sum_{t=1}^T u_t(\beta).$$

- * Solve

$$\hat{\beta} : g_T \left(\underbrace{\hat{\beta}}_{N \times 1} \right) = \underbrace{0}_{N \times 1}.$$

Generalized Method of Moments ...

– Example 1: mean

$$\beta = E x_t,$$

$$u_t(\beta) = \beta - x_t.$$

Choose $\hat{\beta}$ so that

$$g_T(\hat{\beta}) = \frac{1}{T} \sum_{t=1}^T u_t(\hat{\beta}) = \hat{\beta} - \frac{1}{T} \sum_{t=1}^T x_t = 0$$

and $\hat{\beta}$ is simply sample mean.

Generalized Method of Moments ...

– Example 4 in exactly identified case

$$Eu_t(\gamma) = E[\pi_t - 0.99\pi_{t+1} - \gamma s_t] z_t, \quad z_t \sim \text{scalar}$$

choose $\hat{\gamma}$ so that

$$g_T(\hat{\beta}) = \frac{1}{T} \sum_{t=1}^T [\pi_t - 0.99\pi_{t+1} - \hat{\gamma} s_t] z_t = 0,$$

or. standard instrumental variables estimator:

$$\hat{\gamma} = \frac{\frac{1}{T} \sum_{t=1}^T [\pi_t - 0.99\pi_{t+1}] z_t}{\frac{1}{T} \sum_{t=1}^T s_t z_t}$$

Generalized Method of Moments ...

– Key message:

- * In exactly identified case, GMM does not deliver a new estimator you would not have thought of on your own
 - means, correlations, regression coefficients, exactly identified IV estimation, maximum likelihood.

- * GMM provides framework for deriving asymptotically valid formulas for estimating sampling uncertainty.

Generalized Method of Moments ...

- Estimating β in overidentified case ($N > n$)
 - * Cannot exactly implement sample analog of $Eu_t(\beta) = 0$:

$$g_T \left(\underbrace{\hat{\beta}}_{n \times 1} \right) = \underbrace{0}_{N \times 1}$$

- * Instead, ‘do the best you can’:

$$\hat{\beta} = \arg \min_{\beta} g_T(\beta)' W_T g_T(\beta),$$

where

$W_T \sim$ is a positive definite weighting matrix.

- * GMM works for any positive definite W_T , but is most efficient if W_T is inverse of estimator of variance-covariance matrix of $g_T(\hat{\beta})$:

$$(W_T)^{-1} = E g_T(\hat{\beta}) g_T(\hat{\beta})'.$$

Generalized Method of Moments ...

– This choice of weighting matrix very sensible:

- * weight heavily those moment conditions (i.e., elements of $g_T(\hat{\beta})$) that are precisely estimated
- * pay less attention to the others.

Generalized Method of Moments ...

– Estimator of W_T^{-1}

* Note:

$$E g_T \left(\hat{\beta} \right) g_T \left(\hat{\beta} \right)'$$

$$\begin{aligned} &= \frac{1}{T^2} E \left[u_1 \left(\hat{\beta} \right) + u_2 \left(\hat{\beta} \right) + \dots + u_T \left(\hat{\beta} \right) \right] \left[u_1 \left(\hat{\beta} \right) + u_2 \left(\hat{\beta} \right) + \dots + u_T \left(\hat{\beta} \right) \right]' \\ &= \frac{1}{T} \left[\frac{T}{T} E u_t \left(\hat{\beta} \right) u_t \left(\hat{\beta} \right)' + \frac{T-1}{T} E u_t \left(\hat{\beta} \right) u_{t+1} \left(\hat{\beta} \right)' + \dots + \frac{1}{T} E u_t \left(\hat{\beta} \right) u_{t+T-1} \left(\hat{\beta} \right)' \right. \\ &\quad \left. + \frac{T-1}{T} E u_t \left(\hat{\beta} \right) u_{t-1} \left(\hat{\beta} \right)' + \frac{T-2}{T} E u_t \left(\hat{\beta} \right) u_{t-2} \left(\hat{\beta} \right)' + \dots + \frac{1}{T} E u_t \left(\hat{\beta} \right) u_{t-T+1} \left(\hat{\beta} \right)' \right] \\ &= \frac{1}{T} \left[C(0) + \sum_{r=1}^{T-1} \frac{T-r}{T} (C(r) + C(r)') \right], \end{aligned}$$

where

$$C(r) = E u_t \left(\hat{\beta} \right) u_{t-r} \left(\hat{\beta} \right)'$$

* W_T^{-1} is $\frac{1}{T} \times$ spectral density matrix at frequency zero, S_0 , of $u_t \left(\hat{\beta} \right)$,

Generalized Method of Moments ...

– Conclude:

$$W_T^{-1} = E g_T(\hat{\beta}) g_T(\hat{\beta}) = \frac{1}{T} \left[C(0) + \sum_{r=1}^{T-1} \frac{T-r}{T} (C(r) + C(r)') \right] = \frac{S_0}{T}.$$

– W_T^{-1} estimated by

$$\widehat{W_T^{-1}} = \frac{1}{T} \left[\hat{C}(0) + \sum_{r=1}^{T-1} \frac{T-r}{T} (\hat{C}(r) + \hat{C}(r)') \right] = \frac{1}{T} \hat{S}_0,$$

imposing whatever restrictions are implied by the null hypothesis, i.e., (as in ex. 4)

$$C(r) = 0, r > R \text{ some } R.$$

– which is ‘Newey-West estimator of spectral density at frequency zero’

* Problem: need $\hat{\beta}$ to compute W_T^{-1} and need W_T^{-1} to compute $\hat{\beta}$!!

• Solution - first compute $\hat{\beta}$ using $W_T = I$, then iterate...

Generalized Method of Moments ...

- Sampling Uncertainty in $\hat{\beta}$.
 - The exactly identified case
 - The first order Taylor series expansion of $g_T(\beta)$, where β is the true value of the parameters, about $\hat{\beta}$, is:

$$\begin{aligned}g_T(\beta) &\approx g_T(\hat{\beta}) + D(\beta - \hat{\beta}) \\ &= D(\beta - \hat{\beta}),\end{aligned}$$

since $g_T(\hat{\beta}) = 0$. Here,

$$D = \left. \frac{\partial g_T(\beta)}{\partial \beta'} \right|_{\beta = \hat{\beta}}$$

- Since $g_T(\beta)$ is a sample average,

$$g_T(\beta) \stackrel{a}{\sim} N\left(0, \frac{S_0}{T}\right).$$

Generalized Method of Moments ...

– The overidentified case.

- * An extension of the ideas we have already discussed.
- * Can derive the results for yourself, using the ‘delta function method’ for deriving the sampling distribution of statistics.
- * Hamilton’s text book has a great review of GMM.