...

# Econometric Methods for the Analysis of Dynamic General Equilibrium 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
- 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 $x_t$ :

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

- Suppose we have a model solution in hand:

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

**State Space-Observer Form ...**

- Consider example #3 in solution notes, in which

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

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:

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

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

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

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

- Semantics: $\xi_t$ is the *state* of the system (do not confuse with the economic state $(K_t, \varepsilon_t)$!).

- The state equation
  - Law of motion of the state, $\xi_t$

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

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

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

- Summary: State-Space, Observer System -

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

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

- Can be constructed from model parameters

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

so

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

- State space observer system very useful
  - Estimation of $\theta$ and forecasting $\xi_t$ 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:
$$P\left[\xi_t | Y_{t-1}^{data}, Y_{t-2}^{data}, ..., Y_1^{data}\right],\ t = 1, 2, ..., T.$$
    - Smoothing:
$$P\left[\xi_t | Y_T^{data}, ..., Y_1^{data}\right],\ t = 1, 2, ..., T.$$
    - Example: ‘real rate of interest’ and ‘output gap’ can be recovered from $\xi_t$ 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, $t = 0, 1/3, 2/3, 1, 4/3, 5/3, 2, ...$
  - suppose gdp is quarterly, $t = 0, 1, 2, 3, ....$

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

  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.

- 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, $t = 0, 1/3, 2/3, \ldots$

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

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

$$\begin{aligned} \xi_t &= F\xi_{t-1/3} + u_t,\ Eu_tu_t' = Q,\ u_t\tilde{\ } iid\; t = 0, 1/3, 2/3, \ldots \\ Y_t &= H\xi_t,\ Y_t = \begin{pmatrix} y_t \\ \pi_t \\ h_t \end{pmatrix}. \end{aligned}$$

- Note:

$$\overbrace{\xi_t = F^3\xi_{t-1} + u_t + Fu_{t-1/3} + F^2u_{t-2/3}}^{\text{first order vector autoregressive representation for quarterly state}},$$

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

- 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

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

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

**Mixed Monthly/Quarterly Observations ...**

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

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

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

- 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 -

$$\begin{aligned} u_t &= \begin{pmatrix} B \\ 0 \\ I \end{pmatrix} \epsilon_t, \\ \xi_t &= F\xi_{t-1} + u_t,\ \xi_0 = 0, \end{aligned}$$

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

- Choose model parameters, $\theta$, 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

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

  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.

- System (ignoring constant terms and measurement error):

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

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

- 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:

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

- If eigenvalues of $M$ are less than unity,

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

- Substituting into $(**)$

$$\begin{aligned}\epsilon_t \ &= \ \left(HD\right)^{-1}Y_t \\ &\quad - \left(HD\right)^{-1}HF\left[D\left(HD\right)^{-1}Y_{t-1} + MD\left(HD\right)^{-1}Y_{t-2} + M^2D\left(HD\right)^{-1}Y_{t-3} + ...\right]\end{aligned}$$

or,

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

where

$$\begin{aligned} u_t &= HD\epsilon_t \\ B_j &= HFM^{j-1}D\,(HD)^{-1}\,,\ j = 1, 2, ... \end{aligned}$$

- The latter is the VAR representation.
  - Note: $\epsilon_t$ 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:

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

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

- Parameters of system: $(F, Q, a_0, H, R)$. These are functions of model parameters, $\theta$.

- Formulas for computing likelihood

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

are standard (see Hamilton's textbook).

# Bayesian Maximum Likelihood

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

- General property of probabilities:

$$p\left(Y^{data},\theta\right)=\left\{\begin{array}{c} p\left(Y^{data}|\theta\right)\times p(\theta) \\ p\left(\theta|Y^{data}\right)\times p\left(Y^{data}\right) \end{array}\right. ,$$

which implies Bayes' rule:

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

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

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

- Computation of mode sometimes referred to as 'Basyesian maximum likelihood':

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

  maximum likelihood with a penalty function.

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

$$\theta\left(1\right), ..., \theta\left(N\right),$$

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

  – Marginal posterior distribution of any element of $\theta$ displayed as the histogram of the corresponding element $\{\theta\left(i\right), i = 1, .., N\}$

# Metropolis-Hastings Algorithm (MCMC)

- We have (except for a constant):

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

- We want the marginal posterior distribution of $\theta_i$ :

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

- MCMC algorithm can approximate $h\left(\theta_i|Y\right)$.

- Obtain ($V$ produced automatically by gradient-based maximization methods):

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

- Compute the sequence, $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(M)}$ ($M$ 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 $\theta^{(r)}$, $x$,

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

    * step 2: compute scalar, $\lambda$ :

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

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

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

**Metropolis-Hastings Algorithm (MCMC) ...**

- Approximating marginal posterior distribution, $h(\theta_i|Y)$, of $\theta_i$

  – compute and display the histogram of $\theta_i^{(1)}, \theta_i^{(2)}, ..., \theta_i^{(M)}, i = 1, ..., N.$

- Other objects of interest:

  – mean and variance of posterior distribution $\theta$ :

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

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

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

- Some intuition
  - Algorithm is more likely to select moves into high probability regions than into low probability regions.
  - Set, $\left\{\theta^{(1)}, \theta^{(2)}, ..., \theta^{(M)}\right\}$, populated relatively more by elements near mode of $f\left(\theta|Y\right)$.
  - Set, $\left\{\theta^{(1)}, \theta^{(2)}, ..., \theta^{(M)}\right\}$, also populated (though less so) by elements far from mode of $f\left(\theta|Y\right)$.

- 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

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

- 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

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

- Interior optimality implies:

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

- Then,

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

- 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 $\theta$

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

- So, posterior of $\theta_i$ (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.

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

- Property of Normal distribution:

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

- Then,

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

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

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

- 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, $\beta$, have the following property:

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

$$u_t\left(\beta\right) \sim \text{stationary stochastic process (and other conditions)}$$

  - $n = N$ : 'exactly identified'
  - $n < N$ : 'over identified'

– Example 1: mean

$$\beta = Ex_t,$$

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

– Example 2: mean and variance

$$\beta = \begin{bmatrix} \mu & \sigma \end{bmatrix},$$

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

then,

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

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

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

where

$$Ey_t\ =\ \mu_y,\ E\left(y_t-\mu_y\right)^2 = \sigma_y^2$$

$$Ex_t\ =\ \mu_x,\ E\left(x_t-\mu_x\right)^2 = \sigma_x^2$$

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

then

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

– Example 4: New Keynesian Phillips curve

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

or,

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

where,

$$\eta_{t+1} \; = \; 0.99\left(E_t \pi_{t+1} - \pi_{t+1}\right) \Longrightarrow E_t \eta_{t+1} = 0$$

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

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

- Inference about $\beta$
  - Estimator of $\beta$ in exactly identified case ($n = N$)
    - Choose $\hat{\beta}$ to mimick population property of true $\beta$,

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

    - Define:

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

    - Solve

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

– Example 1: mean

$$\beta = Ex_t,$$

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

Choose $\hat{\beta}$ so that

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

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

– Example 4 in exactly identified case

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

choose $\hat{\gamma}$ so that

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

or. standard instrumental variables estimator:

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

– 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.

– Estimating $\beta$ 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

$W_T$ ~ 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 $g_T\left(\hat{\beta}\right)$ :

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

- This choice of weighting matrix very sensible:
  - weight heavily those moment conditions (i.e., elements of $g_T\left(\hat{\beta}\right)$) that are precisely estimated
  - pay less attention to the others.

– Estimator of $W_T^{-1}$

$*$ Note:

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

$$= \frac{1}{T^2}E\left[u_1\left(\hat{\beta}\right)+u_2\left(\hat{\beta}\right)+...+u_T\left(\hat{\beta}\right)\right]\left[u_1\left(\hat{\beta}\right)+u_2\left(\hat{\beta}\right)+...+u_T\left(\hat{\beta}\right)\right]'$$

$$= \frac{1}{T}[\frac{T}{T}Eu_t\left(\hat{\beta}\right)u_t\left(\hat{\beta}\right)'+\frac{T-1}{T}Eu_t\left(\hat{\beta}\right)u_{t+1}\left(\hat{\beta}\right)'+...+\frac{1}{T}Eu_t\left(\hat{\beta}\right)u_{t+T-1}\left(\hat{\beta}\right)'$$

$$+\frac{T-1}{T}Eu_t\left(\hat{\beta}\right)u_{t-1}\left(\hat{\beta}\right)'+\frac{T-2}{T}Eu_t\left(\hat{\beta}\right)u_{t-2}\left(\hat{\beta}\right)'+..+\frac{1}{T}Eu_t\left(\hat{\beta}\right)u_{t-T+1}\left(\hat{\beta}\right)']$$

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

where

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

$*$ $W_T^{-1}$ is '$\frac{1}{T}\times$spectral density matrix at frequency zero, $S_0$, of $u_t\left(\hat{\beta}\right)$'

– Conclude:

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

– $W_T^{-1}$ estimated by

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

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...

- Sampling Uncertainty in $\hat{\beta}$.
  - The exactly identified case
  - By the Mean Value Theorem, $g_T\left(\hat{\beta}\right)$ can be expressed as follows:

$$g_T\left(\hat{\beta}\right) = g_T\left(\beta_0\right) + D\left(\hat{\beta} - \beta_0\right),$$

where $\beta_0$ is the true value of the parameters and

$$D = \frac{\partial g_T\left(\beta\right)}{\partial \beta'}|_{\beta=\beta^*},\text{ some } \beta^* \text{ between } \beta_0 \text{ and } \hat{\beta}.$$

  - Since $g_T\left(\hat{\beta}\right) = 0$ and $g_T\left(\beta_0\right) \overset{a}{\sim} N\left(0, S_0/T\right)$, it follows:

$$\hat{\beta} - \beta_0 = -D^{-1}g_T\left(\beta_0\right),$$

so

$$\hat{\beta} - \beta_0 \overset{a}{\sim} N\left(0, \frac{\left(D'S_0^{-1}D\right)^{-1}}{T}\right)$$

- 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.