Multivariate Time Series: Part 3

Identification and Estimation of Simultaneous Equations Systems

Gerald P. Dwyer

Clemson University

March 2016

Outline

Nonunique Representation

- Even more obviously than in univariate time series, there are various observationally equivalent representations of time series
- Normalization in the context of a vector moving-average representation

$$\mathbf{y}_t = \mathbf{B}(\mathrm{L})\,\boldsymbol{\varepsilon}_t$$

- Normalization issue in terms of vector autoregressions (VARs)

$$\mathbf{A}_0\mathbf{y}_t = \mathbf{A}_1(\mathrm{L})\,\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

  - Recursive dynamic systems, for example,

$$\begin{aligned} y_{1,t} &= a_1^* y_{2,t} + a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + \varepsilon_{1,t} \\ y_{2,t} &= a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + \varepsilon_{2,t} \end{aligned}$$

Systems of behavioral equations

- Relationship of VAR to economic system

$$
\begin{aligned}
q_t &= \alpha p_t + \beta y_t + \varepsilon_t^d, \qquad \alpha < 0, \beta > 0 \\
q_t &= \gamma p_t + \delta w_t + \varepsilon_t^s, \qquad \gamma, \delta > 0
\end{aligned}
$$

Structural equations

- Simultaneous system for supply and demand

$$q_t = \alpha p_t + \beta y_t + \varepsilon_t^d, \; \alpha < 0, \beta > 0$$
$$q_t = \gamma p_t + \delta w_t + \varepsilon_t^s, \;\; \gamma, \delta > 0$$

- Common usage in macroeconomics: An equation is *structural* relative to an intervention if it is invariant to that intervention
- Example of intervention
  - Suppose a sales tax is imposed on purchases and the demand equation is changed to include $(1 + \tau)\, p$ where τ is the sales tax
  - Then these equations are structural relative to the intervention of a change in τ

Model, structure and observational equivalence

- Simultaneous system for supply and demand

$$q_t = \alpha p_t + \beta y_t + \varepsilon_t^d,\ \varepsilon_t^d \sim N\left(0, \sigma_d^2\right)$$
$$q_t = \gamma p_t + \delta w_t + \varepsilon_t^s,\ \ \varepsilon_t^s \sim N\left(0, \sigma_s^2\right)$$

- This is a **model**
- A **structure** based on this model includes the model and a set of parameter values, including the variances of the innovations
- A structure is **identified** if a structure generates a one-to-one mapping from parameter values to the probability distribution of the observables
- Two structures are **observationally equivalent** if they generate the same probability distribution of the observables
- A model is identified if all possible (or at least interesting) structures are identifed
- If no structure is identified, then the model is under-identified or not identified

Example of model, structure and observational equivalence

- Model

$$\alpha_0 q_t = \alpha p_t + \beta y_t + \gamma_0 \varepsilon_t^d,\ \varepsilon_t^d \sim N\left(0, \sigma_d^2\right)$$

- A structure is

$$2q_t = 4p_t + 3y_t + 9\varepsilon_t^d,\ \ \varepsilon_t^d \sim N\left(0, 5\right)$$

- This structure is not identified because the identical set of pdfs of the variables is generated by

$$4q_t = 8p_t + 6y_t + 18\varepsilon_t^d,\ \varepsilon_t^d \sim N\left(0, 5\right)$$

- Identify it by
  - Normalizing on a dependent variable, e.g. $\alpha_0 = 1$ or $\alpha = -1$
    - ★ “For example” because could normalize to some other nonzero value
  - And normalizing the coefficient of the error term γ_0 to $\gamma_0 = 1$ or $\gamma_0 = -1$

Identification of simultaneous-equation system

- In the simultaneous system for supply and demand, we solved for the reduced form

$$q = (\gamma - \alpha)^{-1} \left[\gamma\beta y - \alpha\delta w + \gamma\varepsilon_t^d - \alpha\varepsilon_t^s \right]$$
$$p = (\gamma - \alpha)^{-1} \left[\beta y - \delta w + \varepsilon_t^d - \varepsilon_t^s \right]$$

  - which also can be written

$$\begin{aligned} q &= \pi_{11} y + \pi_{12} w + u_1 \\ p &= \pi_{21} y + \pi_{22} w + u_2 \end{aligned} \tag{1}$$

- Can solve for the parameters in the simultaneous equations uniquely from the reduced form
- Reduced form can be estimated consistently by OLS and is identified
- Therefore the simultaneous equations are identified

Identification of simultaneous-equation system

- Simultaneous system for supply and demand

$$q_t = \alpha p_t + \beta y_t + \beta_2 w_t + \varepsilon_t^d$$
$$q_t = \gamma p_t + \delta w_t + \delta_2 y_t + \varepsilon_t^s$$

- This set of equations is an example of a model that is not identified
  - You can see this by solving for reduced form and trying to solve back for structural parameters from reduced form (Do yourself)
  - There are four reduced-form coefficients and six structural coefficients
- In general, how do we identify coefficients?

Restrictions to identify simultaneous-equation system

- Normalizations
  - Example: coefficients on some variables equal to one
- Use linear identities in model
  - Examples

$$c + i + g + nx = y$$
$$R = r + \pi$$
$$m = M - P$$

- Exclusion restrictions
  - Example: coefficient of w_t is zero in demand equation and coefficient of y_t is zero in supply equation

Restrictions to identify simultaneous-equation system

- Restrictions on coefficients
  - Value of parameters known or restrictions on linear functions of coefficients
  - Example: sum of coefficients equal to one
- Restrictions on structural variance-covariance matrix
  - Example: uncorrelated errors across equations
- Nonlinear functions
  - Example: one variable is a nonlinear function of another

Exclusion restrictions

- Order condition: The number of exogenous variables excluded from an equation is at least as large as the number of jointly dependent variables included in the equation minus one
  - Counts all jointly dependent variables in the equation
  - Over-identified if the number of exogenous variables excluded is greater than the number of jointly dependent variables included in the equation minus one
- Rank condition: A restriction on a sub-matrix of the reduced form coefficients

Illustration of identification

- Shifts in the demand curve identify supply curve
- Shifts in the supply curve identify demand curve
  - Famous papers by E. J. Working (1927) and T. C. Koopmans (1949)

Indirect least squares

- We can estimate the parameters by estimating the reduced form by OLS and then inferring the structural parameters
- This is called **indirect least squares**
  - Estimate reduced form by OLS and solve for structural parameters
  - With simple equations here, can estimate reduced form consistently by ordinary least squares
  - Given the estimated reduced form, solve for the structural coefficients
- Works if structural equation is exactly identified
  - If under-identified, cannot solve uniquely for structural parameters
  - If over-identified, there is more than one implied value of one or more parameters
    - How choose which one?

OLS and estimation of structural equations

- Why can't we just estimate the structural equations by OLS?
- We know from last time that is inconsistent
- Peope often use OLS to estimate a structural equation to get a feel for the correlations in the data
  - What are the effects of the complicated estimation strategy?
  - Not a substitute for estimating the parameters in a consistent way!

Instrumental variables

- Instrumental variables here is no different than instrumental variables in other contexts
- Want variables not included in the equation that are correlated with the variable of interest and not with the error term in the equation
- Probably most common estimation strategy used

Instrumental variables and exogenous variables

- Why not use exogenous variables in system as instruments?
- By assumption, uncorrelated with error term
- By assumption, correlated with jointly dependent variables

Two-stage least squares

- Two-stage least squares (2SLS) uses all the exogenous variables in the system of equations as instruments
- By assumption these variables are uncorrelated with the error term in the equation being estimated
- Regress right-hand-side jointly dependent variables on all the exogenous variables
- Put predicted values in equation
- Run OLS to estimate coefficients with predicted values
    - Not to estimate standard errors of coefficients!
- Assumes homoskedastic errors in equations

Generalized method of moments

- Generalized method of moments (GMM) is similar to IV and 2SLS
- Can allow for heteroskedastic errors
- Easier to include covariance restrictions
- Will go through GMM in more detail later in lecture

Limited-information maximum likelihood

- Limited-information maximum likelihood (LIML) maximizes the likelihood for one equation
- Assume likelihood for errors (generally normal) and maximize likelihood
- Same asymptotic distribution as 2SLS if normal distribution assumption is satisfied
- Advantage over 2SLS
  - 2SLS generally produces different estimates with different normalizations of the equation
  - LIML estimates are unaffected by the normalization of the equation

System methods

- Three-stage least squares (3SLS) and Full-information maximum likelihood (FIML)
- 3SLS – Similar to Seemingly Unrelated Regressions with 2SLS estimates
- FIML – Just maximize likelihood function for system

Single equation versus system

- In system methods, a specification error in one equation is propogated throughout the entire set of equations

Method of Moments

- Generalized method of moments is a generalization of the Method of Moments
- Method of moments
  - Use moments of data to estimate parameters
  - Suppose want to estimate k parameters characterizing a distribution
  - Suppose there exists a well-defined mapping from n parameters θ_i to the first n moments μ_i of the distribution of a variable y
  - This can be written as

$$\begin{aligned} \mu_1 &= f(\theta_1, \theta_2, ..., \theta_n) \\ \mu_2 &= f(\theta_1, \theta_2, ..., \theta_n) \\ &... \\ \mu_n &= f(\theta_1, \theta_2, ..., \theta_n) \end{aligned}$$

  - Method of moments estimators substitute sample moments for population moments and estimate parameters by solving the equations

Method of Moments

- Simple examples of Method of Moments
- Want to compute expected value $\mathrm{E}\, y_t$
  - Use mean for expected value

$$\mathrm{E}\, y_t = \mu_1$$

- Use first and second moment to compute expected value and variance

$$\mathrm{E}\, y_t = \mu_1$$
$$\mathrm{Var}[y_t] = \mathrm{E}\left[y_t - \mathrm{E}\, y_t\right]^2 = \mu_2 - \mu_1^2$$

or

$$\mu_1 = \mathrm{E}\, y_t$$
$$\mu_2 = \mathrm{Var}[y_t] + \mu_1^2$$

- MOM involves substitution of the sample moments for the population moments in these equations

Method of Moments

- MOM involves substitution of the sample moments for the population moments
- Under fairly general conditions, MOM is a consistent estimator of the parameters of interest

Generalized Method of Moments I

- Can think of moments more generally
- Moments (uncentered) in general are

$$
\begin{aligned}
\mathrm{E}\, y_t &= \mu_1 \\
\mathrm{E}\, y_t^2 &= \mu_2 \\
&\dots \\
\mathrm{E}\, y_t^n &= \mu_n
\end{aligned}
$$

- Think of conditional expectations and other functions of the data
- Regression example

$$y_t = \beta x_t + \varepsilon_t$$

  - Impose restriction that

$$\mathrm{E}\left[\varepsilon_t | x_t\right] = 0$$

  - Limit analysis to linear projections

Generalized Method of Moments II

- For linear projections,

$$\mathrm{E}\left[\varepsilon_t | x_t\right] = \mathrm{Cov}\left[\varepsilon_t, x_t\right]$$

- Given zero means

$$\mathrm{E}\left[\varepsilon_t | x_t\right] = \mathrm{E}\, \varepsilon_t x_t$$

- Proceed by imposing restriction in the sample equation

$$y_t = b x_t + e_t$$

$$\mathrm{E}\, e_t x_t = 0$$

which is the same as imposing that the covariance of e_t and x_t is zero

GMM and Ordinary Least Squares I

- The covariance restriction is

$$\mathrm{E}\left[\varepsilon_t x_t\right] = 0$$

- This implies

$$\begin{aligned}
&\sum_{t=1}^{T} e_t x_t = 0 \\
&= \sum_{t=1}^{T} \left(y_t - b x_t\right) x_t \\
&= \sum_{t=1}^{T} y_t x_t - \sum_{t=1}^{T} b x_t x_t \\
&= \sum_{t=1}^{T} y_t x_t - b \sum_{t=1}^{T} x_t^2 = 0
\end{aligned}$$

GMM and Ordinary Least Squares II

- The last line implies

$$b = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2}$$

  if $\sum_{t=1}^{T} x_t^2 \neq 0$

- This is just the least squares estimator
- The GMM estimator with this covariance constraint on the estimated parameter b is the same as OLS

GMM and Instrumental Variables I

- Have an equation

$$y_t = \beta x_t + \varepsilon_t$$

but

$$\mathrm{E}\left[\varepsilon_t | x_t\right] \neq 0$$

Suppose there is another variable z_t such that

$$\mathrm{E}\left[\varepsilon | z_t\right] = 0$$

- In the estimated equation

$$y_t = b x_t + e_t$$

Impose the restriction

$$\mathrm{E}\left[e_t | z_t\right] = 0$$

GMM and Instrumental Variables II

- Again, treat conditional expectation as equivalent to linear projection, implying

$$\mathrm{Cov}\left[e_t z_t\right] = 0$$

- Note that $\mathrm{Cov}\left[e_t z_t\right] = 0$ imposed in the sample implies

$$\begin{aligned}
&\sum_{t=1}^{T} e_t z_t \\
&= \sum_{t=1}^{T} \left(y_t - b x_t\right) z_t \\
&= \sum_{t=1}^{T} y_t z_t - b \sum_{t=1}^{T} x_t z_t = 0
\end{aligned}$$

GMM and Instrumental Variables III

- This is just the instrumental variable estimator

$$b = \frac{\sum_{t=1}^{T} y_t z_t}{\sum_{t=1}^{T} x_t z_t}$$

  if $\sum_{t=1}^{T} x_t z_t \neq 0$
- This works for the identified case
- What if have two instruments for the variable x_t?

GMM when Overidentified I

- If in

$$y_t = \beta x_t + \varepsilon_t$$

there are more restrictions than parameters and

$$\mathrm{E}\left[\varepsilon_t | z_{i,t}\right] = 0, \quad i = 1, 2, ..., \ell$$

cannot necessarily all be satisfied exactly

- Use a minimum distance estimator
- Suppose a model involves a set of k parameters

$$\boldsymbol{\theta}' = \left[\theta_1, \theta_2, ..., \theta_k\right]$$

- Suppose it also involves a set of $\ell > k$ moment conditions

$$\mathrm{E}\, m_i\left(y_t, \boldsymbol{x_t}, \boldsymbol{z_t}, \boldsymbol{\theta}\right) \equiv m_{i,t}\left(\boldsymbol{\theta}\right), \quad i = 1, 2, ..., \ell$$

GMM when Overidentified II

- Then they cannot all be satisfied exactly if the equations are functionally independent
- Let the sample moment conditions be

$$\overline{m}_i = T^{-1} \sum_{t=1}^{T} m_{i,t}\left(\widehat{\boldsymbol{\theta}}\right), \quad i = 1, 2, ..., \ell$$

- These moment conditions are set up so that the correct expected value is zero
- The moment conditions might be, for example, for $i = 1, 2$

$$m_1 = \mathrm{E}\left[e_t | z_{1,t}\right] = 0$$
$$m_2 = \mathrm{E}\left[e_t | z_{2,t}\right] = 0$$

GMM when Overidentified III

- One solution would be to use a minimum distance estimator to satisfy the criterion

$$q = \min_{\boldsymbol{\theta}} \sum_{i=1}^{\ell} \overline{m}_i^2$$

or in matrix form

$$q = \min_{\boldsymbol{\theta}} \overline{\mathbf{m}}'\overline{\mathbf{m}}$$

with

$$\overline{\mathbf{m}}' = [\overline{m}_1, \overline{m}_2, ..., \overline{m}_\ell]$$

- Hansen (1982) shows that this estimator is consistent so that

$$\operatorname{plim} \overline{m}_i = \operatorname{E} \overline{m}_i = 0, \quad i = 1, 2, ..., \ell$$

GMM when Overidentified IV

- He also suggests, and people invariably use, a generalization similar to weighted least squares

$$Q = \min_{\boldsymbol{\theta}} \overline{\mathbf{m}}' \mathbf{W}_\ell \overline{\mathbf{m}}$$

  where $\mathbf{W}_\ell$ is any positive-definite ℓ by ℓ matrix which can depend on the data but not on $\boldsymbol{\theta}$
- An additional assumption is that $\operatorname{plim} \mathbf{W}_\ell = \mathbf{W}$
- The inverses of the individual variances on the diagonal are an obvious criterion which puts more weight on moments with less variance from the restricted value of zero
- This suggests

$$\mathbf{W} = \left[\text{Asym. Var.}\sqrt{T}\overline{\mathbf{m}}\right]^{-1}$$

  which can be defined as

$$\Phi^{-1} = \mathbf{W} = \left[\text{Asym. Var.}\sqrt{T}\overline{\mathbf{m}}\right]^{-1}$$

GMM when Overidentified V

- This is the optimal weighting matrix to use
- The asymptotic variance of this GMM estimator is

$$\mathbf{V}_{GMM} = \frac{1}{T}\left[\boldsymbol{\Gamma}'\mathbf{W}\boldsymbol{\Gamma}\right]^{-1} = \frac{1}{T}\left[\boldsymbol{\Gamma}'\boldsymbol{\Phi}^{-1}\boldsymbol{\Gamma}\right]^{-1}$$

where $\boldsymbol{\Gamma}$ is the matrix of derivatives with i'th row equal to

$$\boldsymbol{\Gamma}_i = \text{plim}\,\frac{\partial \overline{m}_i\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}'}$$

and

$$\boldsymbol{\Phi} = \left[\text{Asym. Var. } T^{\frac{1}{2}}\overline{m}\right]$$

- Asymptotically normally distributed

Method of Moments and Generalized Method of Moments

- Generally speaking, refer to the estimator as
  - Method of Moments when the number of moment conditions equals the number of moments
  - Generalized Method of Moments when the number of moment conditions is greater than the number of moments

Summary I

- Identification of structural equations can proceed in various ways
- Most common identification scheme for an equation is the use of exclusion restrictions
- The number of jointly endogenous variables in the equation minus one is less than or equal to the number of exogenous variables excluded from the equation
- There are a variety of estimation schemes, with Instrumental Variables (IV) and Two stage least squares probably being the most common
- Full-information maximum likelihood is increasingly feasible
- GMM is very general
  - Easy to impose covariance constraints
  - IV and 2SLS are special cases, as is OLS
  - Can be one equation or many
  - GMM is commonly used, especially in asset pricing, in Finance
  - GMM is commonly used when estimating first-order conditions for households and firms in Economics