Multivariate Models Summary

Noé Debrois

June 16, 2024

Simultaneous Equations Models

All the models we have looked at thus far have been single equations models of the form:

outside the model

$$(\text{ARMAX}) \qquad y\phi(L) = X\beta + \theta(L)u \tag{1}$$

inside the model

where X contains exogenous variables, and y is an endogenous variable. In a simultaneous equations model, we consider systems of equations such as the demand and supply of a good:

$$\begin{cases} Q_d = \alpha_1 + \beta_1 P + \gamma_1 S + u_d & (2) \\ Q_s = \alpha_2 + \beta_2 P + \gamma_2 T + u_s & (3) \\ Q_d = Q_s & (4) \end{cases}$$

price "state of technology"

Here, P and Q are endogenous, while S and T are exogenous.

Structural Form and Reduced Form

The structural form of the simultaneous equations model is given by:

Assuming equilibrium of the market:

$$\begin{cases} Q_d = \alpha_1 + \beta_1 P + \gamma_1 S + u_d & (5) \\ Q_s = \alpha_2 + \beta_2 P + \gamma_2 T + u_s & (6) \end{cases}$$

By solving these equations for the endogenous variables, we obtain the reduced form:

$$\begin{cases} P = \pi_{10} + \pi_{11} S + \pi_{12} T + v_1 & (7) \\ Q = \pi_{20} + \pi_{21} S + \pi_{22} T + v_2 & (8) \end{cases}$$

Simultaneous Equations Bias

Estimating the structural form using OLS will lead to biased and inconsistent coefficient estimates due to the endogeneity of P. This is known as simultaneous equations bias.

both equations depend on P.

Not possible to use OLS directly on structural form.

Identification of Simultaneous Equations

To identify an equation, it must satisfy certain conditions:

necessary but not sufficient. "naive"

- **Order Condition:** An equation is exactly identified if the number of excluded variables from that equation is equal to the number of endogenous variables minus one. Over-id. if more than G-1 are absent. Under-if less.

In finance: often over identified.

- **Rank Condition:** A necessary and sufficient condition for identification, involving the rank of a matrix of coefficients.

Tests for Exogeneity

The Hausman test can be used to determine whether variables should be treated as endogenous. This involves estimating reduced form equations, obtaining fitted values and residuals, and testing the inclusion of these residuals in the original equation.

Recursive Systems

In recursive systems, equations can be estimated individually using OLS if the error terms are uncorrelated and the equations are ordered such that each endogenous variable is determined only by previously determined endogenous variables and exogenous variables.

RECURSIVE or TRIANGULAR SYSTEM.

not many systems will be recursive.

Estimation Techniques

Indirect Least Squares (ILS)

ILS involves estimating reduced form equations using OLS and substituting back to obtain structural parameters. This method is valid but not commonly used due to its complexity.

Two-Stage Least Squares (2SLS)

2SLS is a more practical approach, involving two stages:

- **Stage 1:** Estimate the reduced form equations using OLS and obtain fitted values.
- **Stage 2:** Replace every RHS endogenous variables in the structural equations with their fitted values from Stage 1 and estimate the structural equations using OLS.

Recall that OLS directly on the struct. eq. IS IMPOSSIBLE because the endogeneous var. are correlated with the errors.

Instrumental Variables (IV)

IV methods involve using instruments that are correlated with endogenous variables but uncorrelated with the error terms. These "instruments" replace the endogenous variables in the structural equations.

⇝ we don't use "directly" the instruments in the structural equation: but run regressions of the form: $\begin{cases} Y_2 = \lambda_1 + \lambda_2 z_2 + \varepsilon_1 \\ Y_3 = \lambda_3 + \lambda_4 z_3 + \varepsilon_2 \end{cases}$ z_i : instruments for Z_i.

Example of 2SLS

Consider the following system:

$$\begin{cases} Y_1 = \alpha_0 + \alpha_1 Y_2 + \alpha_2 Y_3 + \alpha_3 X_1 + \alpha_4 X_2 + u_1 & (9) \\ Y_2 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_2 & (10) \\ Y_3 = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + u_3 & (11) \end{cases}$$

- **Stage 1:** Estimate the reduced form equations for Y_2 and Y_3 using OLS to obtain fitted values $\hat{Y}_2$ and $\hat{Y}_3$.
- **Stage 2:** Replace Y_2 and Y_3 in the first equation with $\hat{Y}_2$ and $\hat{Y}_3$ and estimate using OLS.

Other Estimation Techniques

- **Three-Stage Least Squares (3SLS):** Allows for non-zero covariances between error terms.
- **Limited Information Maximum Likelihood (LIML):** Estimates reduced form equations by maximum likelihood.
- **Full Information Maximum Likelihood (FIML):** Estimates all equations simultaneously using maximum likelihood.

→ see article of George & Longstaff (1993): modelling the bid-ask spread & volume for options.
↳ 2SLS

Simultaneous models → NOT USED ANYMORE.

What do Fi. Econometrics are using when they deal w/ multivariate time series?

- Vector Autoregressive Models: generalization of autoreg. models (~1980).
  - ↳ more than 1 dependent variable.
  - ↳ $\begin{cases} y_{1t} = \dots \\ y_{2t} = \dots \end{cases}$ y_{1t} depends NOT ONLY on its previous values BUT ALSO on the previous values of the 2nd time series.
  - ↳ Can be extended to a VAR(g) model: g variables & g eq.

can be extend to the case with k lags.

Compactness of notation.

$$\underset{g\times1}{y_t} = \underset{g\times1}{\beta_0} + \underset{g\times g}{\beta_1}\,\underset{g\times1}{y_{t-1}} + \underset{g\times1}{u_t}$$

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad \begin{cases} y_{1t} = \beta_{10} + \beta_{11} y_{1t-1} + \alpha_{11} y_{2t-1} + u_{1t} \\ y_{2t} = \beta_{20} + \beta_{21} y_{2t-1} + \alpha_{21} y_{1t-1} + u_{2t} \end{cases}$$

Advantages of VAR:
- no need to specify which variables are endo: all are endogenous!
- more general than ARMA.
- can use OLS.
- Forecasts are often better than "traditional structural" models: you can put together "key" time-series to get better predictions.

Problems:
- a-theoretical.
- how to decide the appropriate lag? ⇝ I.C again (to be min.).
- Many params.
- how to interpret the coefficients?

⟶ Standard form of VAR: you can put contemporary terms on the left side & invert the matrix...: $y_t = A^{-1}\beta_0 + A^{-1}\beta_1 y_{t-1} + A^{-1}u_t$.

⟶ see article Brooks & Tsolacos (1999): use VAR methodology to investigate the interaction between the UK property mkt & various macroeconomic variables.

Method: how to work w/ multivariate stat. time-series?
1. Decide which endogenous variables you are interested in. Test for exogeneity w/ Hausman.
2. Decide simultaneous equations VS VAR.
3. Estimate the sim. system (2SLS) + stat. test.
4. Use VAR or VARX:
   a) find opti. lag with I.C.
   b) estimate the selected model.
   c) statistical analysis ...