Structural Equation Models and Granger Causality Tests

Jamie Monogan

University of Georgia

February 20, 2018
Objectives

By the end of this meeting, participants should be able to:

- Specify and estimate a structural equation model using 2SLS or 3SLS.
- Describe the variance decomposition model.
- Explain why $F$-ratios are useful in instances of multicollinear predictors.
- Conduct a direct Granger test between two variables.
Competing Views on Multivariate Time Series Analysis

- **View 1:** Theory should guide model specification.
  - With theory, equations can be reduced to a manageable specification.
  - Tools like 2SLS and 3SLS can be used to estimate equations.
  - Empirical approach: structural equation modeling.

- **View 2:** Multiple, competing theories are hard to untangle.
  - Imposing too many restrictions may miss true causal relationships.
  - Rely on data to sort out causal relationships.
  - Empirical approach: Granger causality tests and vector autoregression.
Structural Equations: Using Theory for Specification

- Four variables in study of British political economy, May 1997 to September 2006:
  - XR: Pound-dollar exchange rate.
  - IR: Short-term interest rate.
  - CPI: Consumer price index.
  - PM: Prime ministerial approval.
- Suppose our theory told us that:
  1. Interest rate and consumer price index have a reciprocal relationship.
  2. The current and lagged exchange rate is the only exogenous predictor of CPI.
  3. The current and lagged Prime Minister’s approval is the only exogenous predictor of IR.
- Do you buy this model?
  \[
  CPI_t = \alpha_0 + \alpha_1 IR_t + \alpha_2 XR_t + \alpha_3 XR_{t-1} + \epsilon_{1t} \\
  IR_t = \beta_0 + \beta_1 CPI_t + \beta_2 PM_t + \beta_3 PM_{t-1} + \epsilon_{2t}
  \]
Estimating Structural Equations

- How do we deal with the simultaneity problem in our system?
  \[ CPI_t = \alpha_0 + \alpha_1 IR_t + \alpha_2 XR_t + \alpha_3 XR_{t-1} + \epsilon_{1t} \]
  \[ IR_t = \beta_0 + \beta_1 CPI_t + \beta_2 PM_t + \beta_3 PM_{t-1} + \epsilon_{2t} \]

- Two-stage least squares (2SLS) proceeds in two steps.
  1. First stage is to estimate \( CPI_t \) and \( IR_t \) as a function of all exogenous predictors:
     \[ CPI_t = \gamma_0 + \gamma_1 XR_t + \gamma_2 XR_{t-1} + \gamma_3 PM_t + \gamma_4 PM_{t-1} + \delta_{1t} \]
     \[ IR_t = \zeta_0 + \zeta_1 XR_t + \zeta_2 XR_{t-1} + \zeta_3 PM_t + \zeta_4 PM_{t-1} + \delta_{2t} \]
  2. Second stage is to use the predicted values from the first stage as substitute predictors in our theoretical system:
     \[ CPI_t = \alpha_0 + \alpha_1 \hat{IR}_t + \alpha_2 XR_t + \alpha_3 XR_{t-1} + \epsilon_{1t} \]
     \[ IR_t = \beta_0 + \beta_1 \hat{CPI}_t + \beta_2 PM_t + \beta_3 PM_{t-1} + \epsilon_{2t} \]

- Three-stage least squares (3SLS) combines the 2SLS procedure with seemingly unrelated regression (SUR). SUR models allow errors across equations to correlate. For example, it could be that \( Cor(\epsilon_{1t}, \epsilon_{2t}) > 0 \).
Variance Decomposition Example

- Imagine that we wish to test the notion that prosperity leads to incumbent re-election success.
- Prosperity is an unobserved variable $\pi$, and we have five indicators:
  1. Unemployment
  2. Real growth in GDP
  3. Consumer confidence
  4. Wage growth
  5. Stock prices

The Common Factor Variance Decomposition

- The variance of any indicator $x$ of the latent concept $\pi$ consists of:
  1. Common variance, indicating $\pi$.
  2. Unique variance, due to the indicator itself.
  3. Error variance.

\[ \sigma^2_{\text{total}} = \sigma^2_{\text{common}} + \sigma^2_{\text{unique}} + \sigma^2_{\text{error}} \]

- This is the theory behind measurement models such as exploratory or confirmatory factor analysis.
Variance Decomposition Model
The Effect

- If we estimate a regression of \( y \) as a function of \( x_1, x_2, x_3, x_4, \) & \( x_5 \) to test the effect of \( \pi \) on \( y \), we are likely to see that:
  - One \( x \) term gets explanatory credit for \( \pi \), while the others may be nonsignificant, wrong signed, or even wrong signed and significant.
  - The problem is that after the common variance is modelled in one \( x \), the others are picking up only the screwy relationships between the atheoretical unique variances and \( y \).

The Error in Inference

- In this case it is wrong to assert blindly that one \( x \) influences \( y \) and the others do not. If they tap the same concept, you cannot know which represents \( \pi \) and which represents meaningless unique variances.
  - The correct approach is not to try to separate effects at all if they are inseparable and to ask the theoretically correct question about \( \pi \rightarrow y \), not the individual \( x \) terms.
  - How? Block exclusion (F) test.
Nested Model Test (Dropping q Parameters)

- Produce $R^2$ values for an unrestricted model (UR) and a restricted version (R) that drops several parameters.
- Then a test of the restriction is:

$$F_{q, N-k} = \frac{(R^2_{UR} - R^2_{R})/q}{(1 - R^2_{UR})/(N - k)}$$

- Where $q$ is the number of dropped parameters, $N$ is the number of observations, and $k$ is the number of parameters in the unrestricted model.
- Note: For the special case of 1 parameter, $F = t^2$ for that parameter and $p(F) = p(t)$ and thus we do not need a block exclusion test.
Software

**R**

- Estimate the restricted and unrestricted models as separate “lm” objects (perhaps “model.ur” and “model.r”).
- Then, “anova(model.ur, model.r)” will offer the $F$-test on the restrictions.

**Stata**

- In Stata, simply estimate the unrestricted model and then ask for an $F$-test on several variables in the model.
- It will still yield a comparison of $R^2$ with and without particular coefficient restrictions.
- To test the exclusion of d, e, and f:
  1. `reg y a b c d e f`
  2. `test d e f`
We are often in the position of having about equally good theories and research programs that posit both $x \rightarrow y$ and $y \rightarrow x$.

So we specify our model according to theory and get contrary results.

Researchers on both sides, according to Granger—and later more emphatically Sims—are arrogant about theory and data.

And since they get contrary results, at least one of them is wrong.

So we turn to causality testing in humility, admitting that we do not know the right model and just asking the data to speak to us.
A Substantive Example: The Macro Polity Model

- It entertains a causal ordering of Economic Outcomes (EO), Economic Perceptions (EP), Presidential Approval (A), Macropartisanship (M), Election Outcomes (E) (and more)
  - EO $\rightarrow$ EP $\rightarrow$ A $\rightarrow$ M $\rightarrow$ E
- For many of these linkages micro theory could justify causality in either direction.
- Example: Economic Perceptions $\rightarrow$ Approval
  - It is plausible that those who think the economy is strong are more likely to approve the president.
  - But also those who approve the president might be likely to distort their economic perceptions in the direction of finding more prosperity than actually is the case.
- This is a case (for the macro indicators) where we would rather let the data speak about cause and exogeneity that to settle it—maybe falsely—by assumption.
Granger Causality

- Intuition: if $x \rightarrow y$ then perturbing $x$ ($\Delta x$) leads to later changes in $y$ ($\Delta y$)
- Asymmetry: if $x \rightarrow y$ then perturbing $y$ ($\Delta y$) has no effect on future values of $x$
- Definition: A series $x$ may be said to cause a series $y$ if and only if the expectation of $y$ given the history of $x$ is different from the unconditional expectation of $Y$.
  - $E(y|y_{t-k},x_{t-k}) \neq E(y|y_{t-k})$
The Direct Granger Test

- For $k$ appropriate lags, we model:
  \[ y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_k y_{t-k} + \epsilon_t \]
- Then we ask whether adding similar information about $x$ will improve our ability to predict $y$. Thus:
  \[ y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_k y_{t-k} + \gamma_1 x_{t-1} + \ldots + \gamma_k x_{t-k} + \nu_t \]
- The $\hat{\beta}$s are uninformative.
- If the $\hat{\gamma}$s are jointly significant, we have established cause.

Exogeneity: The Mirror Image

- For $k$ appropriate lags, we model:
  \[ x_t = \alpha_0 + \alpha_1 x_{t-1} + \ldots + \alpha_k x_{t-k} + \delta_t \]
- Then we ask whether adding similar information about $y$ will improve our ability to predict $x$. Thus:
  \[ x_t = \alpha_0 + \alpha_1 x_{t-1} + \ldots + \alpha_k x_{t-k} + \zeta_1 y_{t-1} + \ldots + \zeta_k y_{t-k} + \eta_t \]
- The $\hat{\alpha}$s are uninformative.
- If the $\hat{\zeta}$s are jointly zero, then $x$ is exogenous with respect to $y$. 
The Direct Granger Test

Software

**R**
- Create a “tsunion” data set with several lags of y & x.
- Run two “lm” objects with y as the DV: one with lagged x’s (y.with.x) and one without (y.without.x).
- `anova(y.with.x, y.without.x)`, significant F implies x $\rightarrow$ y.
- Then the reverse: Run two “lm” objects with x as the DV: one with lagged y’s (x.with.y) and one without (x.without.y).
- `anova(x.with.y, x.without.y)`, nonsignificant F implies y $\not\rightarrow$ x.

**Stata**
- `reg y l(1/4).y l(1/4).x`
- `test l(1/4).x`, significant F implies x $\rightarrow$ y.
- Then the reverse: `reg x l(1/4).x l(1/4).y`
- `test l(1/4).y`, nonsignificant F implies y $\not\rightarrow$ x.
See Freeman (1983).

Both approaches are similar in that they first control for expected patterns of autodependence in the dependent variable, and then claim causation when the cleaned series are associated with the independent variable.

The difference is that Box-Jenkins modeling uses an empirical technique to identify the best simple model of autodependence whereas Granger modeling is more brute force, including serveral lags of $y$ so that dependence may be modeled out even without knowing exactly what it is.
The Direct Granger test violates the correct specification assumption of OLS, since the correct specification is assumed to be unknown.

It relies on overfitting to make sure that all the autodependence process is removed from the data.

Overfitting is harmless as regards unbiasedness, but causes estimator inefficiency.

Consequently the Granger coefficients are not optimal and should not be used. That is, the direct Granger test should be used as a significance test only, not as a source of structural coefficients.

Another concern: What did you think of Sheehan & Grieves’s (1982) attempt to “save” William Stanley Jevons’s finding?
For the Coming Week

- Thursday, February 22, 3:30-5:00, Pinnacle Room: Parthemos Lecture, Janet Box-Steffensmeier
- **February 27: Midterm Exams Due**
- Also for February 27, please read:
  - *Political Analysis Using R*, Section 9.3.