Forecasting

in Economics, Business, Finance and Beyond

CH 16: Multivariate: Vector Autoregression

Multivariate Forecasting

So far we have used only *univariate* models for forecasting

In many settings, the evolution of one variable is related to developments in others; and can improve forecasts- multivariate regression modeling.

2

Bivariate Regression model

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

where x helps determine (*cause*) y.

Terminology: dependent (endogenous) and independent (exogenous) variables.

3

Conditional versus unconditional forecasting models

A conditional forecasting model takes x as given and forecasts y conditional on the value of x

$$y_{T+h,T} \mid x^*_{T+h} = \beta_0 + \beta_1 x^*_{T+h}$$

Also called scenario analysis. Why?

Scenario analysis is a process of analyzing possible future events by considering alternative possible outcomes (sometimes called "alternative worlds"). Thus, scenario analysis, which is one of the main forms of projection, does not try to show one exact picture of the future. Instead, it presents several alternative future developments.

4

Conditional versus unconditional forecasting models

Unconditional forecasting models. we usually are interested in *unconditional* forecasts:

$$y_{T+h,T} = \beta_0 + \beta_1 x_{T+h,T}$$

We need to have an optimal forecast of x in order to form an optimal unconditional forecast of y.

5

Unconditional forecasts

Separately model x (say, as an autoregressive model, and insert x-hat into our y forecasting equation.

$$\hat{x}_{T+h,T} = \beta_{x0} + \beta_{x1} x_{T+h-1}$$

Usually better to estimate all parameters simultaneously,

$$y_t = \beta_0 + \delta x_{t-1} + \varepsilon_t$$

6

Distributed lag models

Capture cross-variable dynamics using a sequence of lags of the other variable

$$\begin{aligned} y_t &= \beta_0 + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \delta_3 x_{t-3} + \ldots + \varepsilon_t \\ &= \beta_0 + \sum_{i=1}^{N_x} \delta_i x_{t-i} + \varepsilon_t \end{aligned}$$

The deltas are the "*lag weights*" and their pattern is called the *lag distribution*

7

Estimation problems

- Problem is there may be a large number (N_x) of parameters to estimate.
- How can we overcome this problem?
  - polynomial distributed lags
  - rational distributed lags

8

Polynomial Distributed Lags

Polynomial distributed lag models are estimated by minimizing the sum of squared residuals in the usual way, subject to the constraint that the lag weights follow a low-order polynomial whose degree must be specified. For example, solve

$$\min_{\beta_0, \delta_i} \sum_{t=N_x+1}^{T} \left[y_t - \beta_0 - \sum_{i=1}^{N_x} \delta_i x_{t-i} \right]^2,$$

subject to $\delta_i = P(i) = a + bi + ci^2, \; i = 1, ..., N_x.$

9

Rational Distributed Lags

Polynomial distributed lags produce aesthetically appealing, but basically ad hoc, lag distributions. After all, why should the lag weights necessarily follow a low-order polynomial?

Rational distributed lags promote parsimony, and hence smoothness in the lag distribution, but they do so in a way that's potentially much less restrictive than requiring the lag weights to follow a low-order polynomial.

Rational distributed lags brings both lags of x and lags of y into the model. (lags of y, lags of x, and lags of errors (MA), or, in effect, an ARMA model for y with lagged x terms added)

10

Adding own-variable dynamics

lag dependent variables.

$$y_t = \beta_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \delta_1 x_{t-1} + \delta_2 x_{t-2} + ... + \varepsilon_t$$
$$= \beta_0 + \sum_{i=1}^{N_y} \alpha_i y_{t-i} + \sum_{i=1}^{N_x} \delta_i x_{t-i} + \varepsilon_t$$

We can write this more compactly as:

$$B(L) y_t = A(L) x_t + \varepsilon_t$$

11

Adding own-variable dynamics

Another way is to **include ARMA disturbances**

$$y_t = \beta_0 + \delta_1 x_{t-1} + \delta_2 x_{t-2} + ... + \varepsilon_t \quad = \beta_0 + \sum_{i=1}^{N_x} \delta_i x_{t-i} + \varepsilon_t$$

$$\varepsilon_t \sim \frac{\Theta(L)}{\Phi(L)} v_t$$

$$v_t \sim WN(0, \sigma^2)$$

12

Transfer function models

Or both: the **transfer function model** is the most general multivariate model and includes both types of influences

$$y_t = \frac{A(L)}{B(L)} x_t + \frac{C(L)}{D(L)} \varepsilon_t$$

13

Single or system equation models?

In some cases it may make sense to assume that right-hand side variables can be treated as exogenous for the purposes of modeling/forecasting a left-hand side variable

E.g. an individual firms revenues may depend on GDP, but not visa versa.

Formally what is needed for ***estimation*** is that the RHS variable is ***weakly*** exogenous with respect to the parameters we are trying to estimate, and what is needed for ***forecasting*** is that the RHS variable is ***strongly*** exogenous with respect to the parameters we are trying to estimate.

.

14

Exogeneity concepts

A variable is *weakly exogenous* for parameters of interest if the marginal process for the variable contains no information useful for estimating those parameters

The marginal process is very roughly speaking the distribution of the variable itself without regard to particular values of variables it may be correlated with.

This means we can estimate the parameters without having to worry about the random process behind the weakly exogenous right-hand-side variable.

15

Exogeneity concepts

A variable is *strongly exogenous* if it is weakly exogenous AND it is not affected by lagged values of the endogenous variable.

In this case, we can forecast without having to worry about how our left-hand-side variable might affect future values of our right-hand-side variable.

16

Single or system equation models?

Often, however, we want to allow for influences running potentially in both (all) directions

This leads us to system modeling approaches; i.e., vectors on the left-hand-side of the model, and matrix expressions on the right-hand-side.

These types of models are called Vector Auto-Regression models of order p, or VAR(p) models.

17

A VAR(1) model

A VAR(1) for a system of N=2 variables runs 2 equations where in each case 1 lag of the own and other variables are included.

$$y_{1,t} = \phi_{11} y_{1,t-1} + \phi_{12} y_{2,t-1} + \varepsilon_{1,t}$$
$$y_{2,t} = \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} + \varepsilon_{2,t}$$

where

$$\varepsilon_{1,t} \sim WN(0, \sigma_1^{\,2})$$
$$\varepsilon_{2,t} \sim WN(0, \sigma_2^{\,2})$$
$$\mathrm{cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) = \sigma_{12}$$

18

A VAR(1) model

$$y_{1,t} = \phi_{11} y_{1,t-1} + \phi_{12} y_{2,t-1} + \varepsilon_{1,t}$$
$$y_{2,t} = \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} + \varepsilon_{2,t}$$

So innovations can be correlated across regressions.

If exactly the same vars are on RHS (as in this case) then OLS (Ordinary Least Squares regression) on individual equations can be used; otherwise must use SUR (Seemingly Unrelated Regressions).

BIC and AICC for the complete system can be constructed.

19

Seemingly Unrelated Regressions

In econometrics, the **seemingly unrelated regressions** (**SUR**) or **seemingly unrelated regression equations** (**SURE**) model, proposed by Arnold Zellner in 1962, is a generalization of a linear regression model that consists of several regression equations, each having its own dependent variable and potentially different sets of exogenous explanatory variables.

Each equation is a valid linear regression on its own and can be estimated separately, which is why the system is called *seemingly unrelated*, although some authors suggest that the term *seemingly related* would be more appropriate, since the error terms are assumed to be correlated across the equations.

20

Seemingly Unrelated Regressions

The model can be estimated equation-by-equation using standard ordinary least squares (OLS). Such estimates are consistent, however generally not as efficient as the SUR method, which amounts to feasible generalized least squares with a specific form of the variance-covariance matrix.

Two important cases when SUR is in fact equivalent to OLS are when the error terms are in fact uncorrelated between the equations (so that they are truly unrelated) and when each equation contains exactly the same set of regressors on the right-hand-side.

21

A VAR(p) model

A VAR(p) for a system of N variables runs N equations where in each case p lags of the own and other variables are included.

Example: VAR(2) for a bivariate (N=2) system:

$$y_{1,t} = \phi_{11}^{1} y_{1,t-1} + \phi_{11}^{2} y_{1,t-2} + \phi_{12}^{1} y_{2,t-1} + \phi_{12}^{2} y_{2,t-2} + \varepsilon_{1,t}$$

$$y_{2,t} = \phi_{21}^{1} y_{1,t-1} + \phi_{21}^{2} y_{1,t-2} + \phi_{22}^{1} y_{2,t-1} + \phi_{22}^{2} y_{2,t-2} + \varepsilon_{2,t}$$

(Here, I am just using superscripts to keep track of the lags, not to indicate powers, so there are 8 distinct parameters in this model—maybe not great notation!)

22

Danger! Danger!

I am going to stop now before your heads blow up.

ITSM can fit multivariate time series models, but we would need another three weeks (at least) to build up the necessary concepts before it would make sense to show you an example.

So, I won't!

23

Before We Go….

Important Points to Take Away From This Class!

24

Before We Go….

What is forecasting?

Forecasting is a tool used for predicting future demand based on past demand information.

Before We Go….

Key issues in forecasting

1. A forecast is only as good as the information included in the forecast (past data)
2. History is not a perfect predictor of the future (i.e.: there is no such thing as a perfect forecast)

REMEMBER: Forecasting is based on the assumption that the past predicts the future! When forecasting, think carefully whether or not the past is strongly related to what you expect to see in the future…

Before We Go….

What's Forecasting All About?

From the March 10, 2006 WSJ:

Ahead of the Oscars, an economics professor, at the request of Weekend Journal, processed data about this year's films nominated for best picture through his statistical model and predicted with 97.4% certainty that "Brokeback Mountain" would win.

Oops.

Last year, the professor tuned his model until it correctly predicted 18 of the previous 20 best-picture awards; then it predicted that "The Aviator" would win; "Million Dollar Baby" won instead.

Sometimes models tuned to prior results don't have great predictive powers.

Quotes

"Predicting is very difficult, especially if it's about the future."
Nils Bohr

"Essentially, all models are wrong, but some are useful".
George Box, 1987

"The best material model of a cat is another, or preferably the same, cat."
Norbert Wiener, 1945

Quotes

Other “Wiener-isms”

Scientific discovery consists in the interpretation for our own convenience of a system of existence which has been made with no eye to our convenience at all.

One of the chief duties of a mathematician in acting as an advisor to scientists is to discourage them from expecting too much of mathematicians.

What most experimenters take for granted before they begin their experiments is infinitely more interesting than any results to which their experiments lead.

29

Before We Go….

30

THE END!

It’s been fun!

31