Eco 4306 Economic and Business Forecasting

Chapter 11: Forecasting with a System of Equations: Vector Autoregression

Motivation

- so far, we discussed the modeling and forecasting for a single stochastic process
- univariate ARIMA model: use the time dependence of the stochastic process to develop a model and produce a forecast
- we successfully applied this approach to various time series such as real GDP, construction spending, house prices, per capita personal income, interest rates, labor force participation rate, earnings per share, . . .

Motivation

- economists develop theoretical economic models - simplified representations of the economy - with many economic variables interacting with each other
  - consumption depending on the level of income
  - investment decisions as a function of interest rates
  - money supply influencing inflation
  - exchange rates linking production of two or more countries
- we will now explore how these kind of interactions can be modeled from an econometric and forecasting perspective

Motivation

- we will thus develop **multivariate forecasting models** to *jointly* forecast several time series (e.g. consumption and income)
- we will analyze not only the time dependence in each series, but also the interdependence between them over time
- we will thus create a system of equations, one equation for each variable, with much richer dynamics than those contained in each univariate process
- each equation will contain information not just about its dependent variable's history but also the history of the other variables in the system
- motivation: with a multivariate information set (e.g. past consumption and past income) we may be able to construct a better forecast for variables in the system

Motivation

- we will intorduce some new concepts: vector autoregression model (VAR), Granger-causality, impulse-response functions (IRF)
- vector autoregression models are commonly used to forecast systems of interrelated time series and to analyze the dynamic impact of random disturbances on the variables in the system

11.1 What Is Vector Autoregression (VAR)?

- suppose that we have two processes $\{Y_t\}$ and $\{X_t\}$
- a **vector autoregression of order p**, VAR(p) for $\{Y_t\}$ and $\{X_t\}$, is then defined as a system of two equations where the regressors are the lagged values of $\{Y_t\}$, $\{X_t\}$

$$Y_t = c_1 + \alpha_{11} Y_{t-1} + \ldots + \alpha_{1p} Y_{t-p} + \beta_{11} X_{t-1} + \ldots + \beta_{1p} X_{t-p} + \varepsilon_{1t}$$
$$X_t = c_2 + \alpha_{21} Y_{t-1} + \ldots + \alpha_{2p} Y_{t-p} + \beta_{21} X_{t-1} + \ldots + \beta_{2p} X_{t-p} + \varepsilon_{2t}$$

- error terms ε_{1t} and ε_{2t} are assumed to be normal random variables with $\varepsilon_{1t} \sim N(0, \sigma_1^2)$ and $\varepsilon_{2t} \sim N(0, \sigma_2^2)$
- errors can be contemporaneously correlated, that is, $cov(\varepsilon_{1t}, \varepsilon_{2t}) \neq 0$
- note: $\{Y_t\}$ and $\{X_t\}$ should be second order weakly stationary or trend stationary (we can include a deterministic trend in VAR)

11.1 What Is Vector Autoregression (VAR)?

characteristics of a VAR system

1. only lagged values of $\{Y_t\}$ and $\{X_t\}$ appear in the right-hand side of equations, we do not include moving average (MA) terms (lagged values of ε_{1t} and ε_{2t})
2. all equations in VAR contain same regressors, $\{Y_{t-1}, \ldots, Y_{t-p}, X_{t-1}, \ldots, X_{t-p}\}$, past history of $\{X_t\}$ is allowed to affect the present value of $\{Y_t\}$, and vice versa, past history of $\{Y_t\}$ may affect the present value of $\{X_t\}$
3. order of the system is the largest number of lag p, it is common to all equations

11.1 What Is Vector Autoregression (VAR)?

4. in general VAR may contain more than two variables - for example, in case of three processes $\{Y_t\}$, $\{X_t\}$, $\{Z_t\}$, we will have three equations, each with regressors: $\{Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}, X_{t-1}, X_{t-2}, \dots, X_{t-p}, Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}\}$
5. number of parameters increases very quickly with the number of lags and the number of variables in the system
   - in a VAR with 2 variables increasing order from p to $p+1$ adds 4 parameters
   - in a VAR with 3 variables, increasing order adds 9 parameters (3 parameters per equation)
   - in general, with n variables the number of parameters increases by n^2

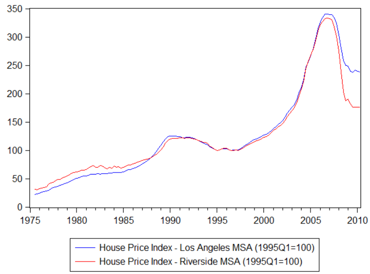
   VAR system can thus become overparameterized quickly, a large sample size is needed to be able to estimate such a large number of parameters

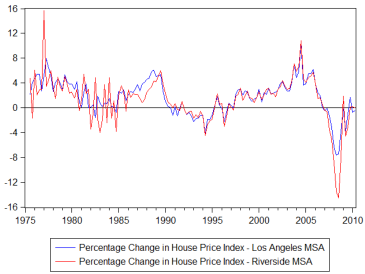
11.2 Estimation of VAR

- a common approach to choose the number of lags is to use the information criteria
- that is, estimate a VAR(p) for each value of $p \leq p_{max}$ and select the VAR(p^*) such as p^* minimizes the AIC or SIC
- each equation in a VAR can be estimated separately by an OLS

Example: House Price Index in Los Angeles and Riverside MSAs

- time series of the quarterly house price index for two Metropolitan Statistical Areas in California - Los Angeles and Riverside - about 60 miles apart





Example: House Price Index in Los Angeles and Riverside MSAs

why consider both locations jointly and developing a VAR is the right approach

- economies of Los Angeles and Riverside are linked, thousands of people who commute daily in both directions
- Los Angeles attracts many businesses in the manufacturing, entertainment, health, education, and services industries
- Riverside has a smaller economy, benefits greatly from the economic activity in Los Angeles area
- increased demand for housing in Los Angeles bids up real estate prices, causing more people move to Riverside where real estate is cheaper which increases prices there too

Example: House Price Index in Los Angeles and Riverside MSAs

- the two time series for house price index and also for percentage changes in house price index tend to move together
- time series for house price indexes are not second order weakly stationary, so we will consider their percentage changes instead
- correlation coefficient for the two series in percentage changes is 0.84
- the collapse of the housing market bubble has been more dramatic in Riverside (14.6% decline in 2008Q3, overall drop from peak 330 to 175) than in Los Angeles (7.5% decline in 2008Q3, overall drop from peak 340 to 240)

Example: House Price Index in Los Angeles and Riverside MSAs

- we thus estimate a VAR(1) for two endogenous variables - house price growth in Los Angeles `ghp_LA` and in Riverside `ghp_RI`

$$Y_t = c_1 + \alpha_{11} Y_{t-1} + \beta_{11} X_{t-1} + \varepsilon_{1t}$$
$$X_t = c_2 + \alpha_{21} Y_{t-1} + \beta_{21} X_{t-1} + \varepsilon_{2t}$$

- to estimate a VAR in EViews select **Object** → **New Object** → **VAR** and in the VAR specification dialog select "Unrestricted VAR", enter estimation sample, provide names of the variables in "Endogenous Variables", and the number of lags in "Lag Intervals for Endogenous"
- in our example we choose 1975Q2 to 2009Q2 as estimation sample, leaving 2009Q3 to 2010Q2 as prediction sample for forecasting evaluation
- we also first consider the case with one lag only and so enter "1 1" in "Lag Intervals for Endogenous"

Example: House Price Index in Los Angeles and Riverside MSAs

Vector Autoregression Estimates
Date: 04/05/18 Time: 19:06
Sample (adjusted): 1975Q4 2009Q2
Included observations: 135 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	GHP_LA	GHP_RI
GHP_LA(-1)	0.772775	0.889737
	(0.08730)	(0.13775)
	[8.85216]	[6.45918]
GHP_RI(-1)	0.074079	0.061802
	(0.06823)	(0.10766)
	[1.08575]	[0.57406]
C	0.270471	-0.355668
	(0.16604)	(0.26199)
	[1.62899]	[-1.35757]
R-squared	0.709407	0.551940
Adj. R-squared	0.705004	0.545151
Sum sq. resids	328.5469	818.0097
S.E. equation	1.577654	2.489386
F-statistic	161.1217	81.30165
Log likelihood	-251.5915	-313.1647
Akaike AIC	3.771726	4.683921
Schwarz SC	3.836288	4.748483
Mean dependent	1.812181	1.385227
S.D. dependent	2.904716	3.691124
Determinant resid covariance (dof adj.)		9.788695
Determinant resid covariance		9.358475
Log likelihood		-534.0625
Akaike information criterion		8.000925
Schwarz criterion		8.130049

Example: House Price Index in Los Angeles and Riverside MSAs

- in the estimation output each column in the table corresponds to an equation in the VAR
- our hypothesis that the real estate market in Riverside is very dependent on the Los Angeles market is confirmed: since $\alpha_{21} = 0.8805$ a 1 percentage point increase in house prices in Los Angeles translates into a 0.88 percentage point growth of house prices in Riverside
- on the contrary, the Los Angeles market does not seem to be affected by the Riverside market: since $\beta_{11} = 0.0719$ and its t-statistics is smaller than 2 an increase in the house prices in Riverside does not have a statistically significant effect on house prices in Los Angeles

Example: House Price Index in Los Angeles and Riverside MSAs

- we still need to verify whether one lag is sufficient or whether more lags are needed to model the dynamics of the system
- SIC and AIC are helpful here - we choose the lag length by minimizing these criteria as in the univariate case
- after estimating a VAR, select **View** → **Lag Structure** → **Lag Length Criteria**
- the optimal lag is denoted by an asterisk; in our example AIC and SIC agree on choosing one lag as the optimal length of the VAR

VAR Lag Order Selection Criteria
Endogenous variables: GHP_LA GHP_RI
Exogenous variables: C
Date: 04/05/18 Time: 19:06
Sample: 1975Q3 2009Q2
Included observations: 128

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-560.5819	NA	22.52417	8.790343	8.834906	8.808449
1	-478.4724	160.3702*	6.646854*	7.569881*	7.703570*	7.624199*
2	-477.1624	2.517645	6.932617	7.611912	7.834727	7.702443
3	-472.8949	8.068099	6.904669	7.607733	7.919674	7.734477
4	-469.9419	5.490750	7.020282	7.624093	8.025159	7.787048
5	-469.1486	1.450400	7.383826	7.674196	8.164389	7.873364
6	-464.1411	8.997823	7.272540	7.658454	8.237773	7.893834
7	-462.5510	2.807539	7.557332	7.696109	8.364553	7.967701
8	-461.9808	0.988865	7.981313	7.749700	8.507271	8.057505

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

11.3 Granger Causality

- the main idea behind VARs is that a multivariate information set could be more helpful than a univariate set in forecasting the variables of interest
- we could ask which among the variables in the information set and the VAR are most useful to forecast others
- for example, suppose that for a process of interest $\{Y_t\}$ we have a multivariate information set

$$I_t = \{y_0, y_1, \ldots, y_t, x_0, x_1, \ldots x_t, z_0, z_1, \ldots z_t\}$$

- we are interested in whether the information provided by time series $\{X_t\}$ and $\{Z_t\}$ is helpful to forecast future values of $\{Y_t\}$

11.3 Granger Causality

- **Granger-causality**: if some variable, for example, $\{X_t\}$ does not help to predict $\{Y_t\}$, we say that $\{X_t\}$ does not Granger-cause $\{Y_t\}$
- a test of $\{X_t\}$ not Granger causing $\{Y_t\}$ is performed by considering the equation in VAR for variable Y_t

  $$Y_t = c_1 + \alpha_{11} Y_{t-1} + \ldots + \alpha_{1p} Y_{t-p} + \beta_{11} X_{t-1} + \ldots + \beta_{1p} X_{t-p} + \gamma_{11} Z_{t-1} + \ldots + \gamma_{1p} Z_{t-p} + \varepsilon_{1t}$$

  and testing a joint hypothesis $H_0 : \beta_{11} = \ldots = \beta_{1p} = 0$ against the alternative $H_1 : \beta_{1i} \neq 0$ for some $i \in \{1, 2, \ldots, p\}$
- if the test statistic exceeds the critical value (or equivalently if the associate p-value is small) we reject the null hypothesis
- Granger-causality can be also tested for a group of variables: for example, we can test that $\{X_t\}$ and $\{Z_t\}$ do not help to predict $\{Y_t\}$ by performing the test with joint hypothesis $H_0 : \beta_{11} = \ldots = \beta_{1p} = \gamma_{11} = \ldots = \gamma_{1p} = 0$ against $H_1 : \beta_{1i} \neq 0$ for some $i \in \{1, 2, \ldots, p\}$ or $\gamma_{1i} \neq 0$ for some $i \in \{1, 2, \ldots, p\}$

Example: House Price Index in Los Angeles and Riverside MSAs

- going back to our data on house prices in Riverside and Los Angeles, we could ask two questions:

1. Is the Riverside market Granger-causing the Los Angeles real estate market?

   i.e. test whether $\{X_t\}$ does not Granger-cause $\{Y_t\}$, and the null hypotheses is $H_0 : \beta_{11} = 0$

2. Is the Los Angeles market Granger-causing the Riverside market?

   i.e. test whether $\{Y_t\}$ does not Granger-cause $\{X_t\}$, and the null hypotheses is $H_0 : \alpha_{21} = 0$

- to perform these tests in EViews after estimating a VAR, select **View** $\rightarrow$ **Lag Structure** $\rightarrow$ **Granger Causality/Block Exogeneity Tests**

Example: House Price Index in Los Angeles and Riverside MSAs

VAR Granger Causality/Block Exogeneity Wald Tests
Date: 04/08/19 Time: 01:46
Sample: 1975Q3 2009Q2
Included observations: 135

Dependent variable: GHP_LA

Excluded	Chi-sq	df	Prob.
GHP_RI	1.178856	1	0.2776
All	1.178856	1	0.2776

Dependent variable: GHP_RI

Excluded	Chi-sq	df	Prob.
GHP_LA	41.72100	1	0.0000
All	41.72100	1	0.0000

Example: House Price Index in Los Angeles and Riverside MSAs

- when we examine the first equation, so test the effect of the Riverside market on the Los Angeles market, we obtain a small value of the test statistic (1.178) and large p-value (0.2776)
- this means that we fail to reject the null and conclude that the Riverside market does not Granger-cause the Los Angeles market
- in other words, the Riverside market does not have predictive ability for the Los Angeles market
- when we examine the second equation, in which we are testing the effect of the Los Angeles market on the Riverside market, the value of the statistic is very large (41.721) and the p-value is 0
- this means that we strongly reject the null hypothesis that changes in house prices in Los Angeles does not Granger-cause changes in house prices in Riverside
- we conclude that the Los Angeles market has predictive ability for the Riverside market
- this is additional evidence for our original hypothesis that the real estate sector in Riverside depends highly on the Los Angeles sector

11.4 Impulse-Response Functions

- a VAR allows us to track how shocks to one variable are transmitted to the other variables in the system
- suppose that there is a shock to Los Angeles economy - an increase in the federal defense spending that positively affects businesses in the area that work directly or indirectly for the Department of Defense
- businesses in Los Angeles will increase the number of employees, this will increase demand for houses in Los Angeles, which in turn increases their prices and provides incentives for people to consider houses in Riverside instead, and that causes prices of houses to rise there too

11.4 Impulse-Response Functions

- consider again the VAR(1) for growth of house prices in Los Angeles Y_t and in Riverside X_t

$$Y_t = c_1 + \alpha_{11} Y_{t-1} + \beta_{11} X_{t-1} + \varepsilon_{1t}$$
$$X_t = c_2 + \alpha_{21} Y_{t-1} + \beta_{21} X_{t-1} + \varepsilon_{2t}$$

  a positive shock ε_{1t} will immediately increase Y_t, which in next period increases Y_{t+1}, and also X_{t+1}

- then, an increase in Y_{t+1} causes an increase in Y_{t+2} and also X_{t+2}, in addition an increase in X_{t+1} causes an increase in Y_{X+2} and also Y_{t+2}
- this continues on for $t+3, t+4, \ldots\ldots$
- a shock to the i-th variable thus not only directly affects the i-th variable but is also transmitted to all of the other endogenous variables through the dynamic (lag) structure of the VAR

11.4 Impulse-Response Functions

- the **impulse-response functions** (IRFs) measure the change in the variables of the VAR over time after a one time shock
- formally impulse-response function is defined as

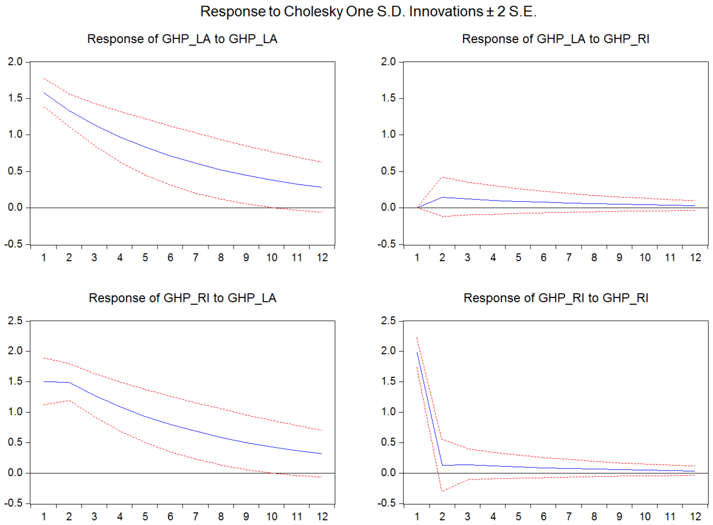
$$\psi_{1,s}^{Y} = \frac{\partial Y_{t+s}}{\partial \varepsilon_{1t}} \approx \frac{\Delta Y_{t+s}}{\Delta \varepsilon_{1t}} \qquad \psi_{1,s}^{X} = \frac{\partial X_{t+s}}{\partial \varepsilon_{1t}} \approx \frac{\Delta X_{t+s}}{\Delta \varepsilon_{1t}}$$

and

$$\psi_{2,s}^{Y} = \frac{\partial Y_{t+s}}{\partial \varepsilon_{2t}} \approx \frac{\Delta Y_{t+s}}{\Delta \varepsilon_{2t}} \qquad \psi_{2,s}^{X} = \frac{\partial X_{t+s}}{\partial \varepsilon_{2t}} \approx \frac{\Delta X_{t+s}}{\Delta \varepsilon_{2t}}$$

- IRFs thus provides insight into the persistence and magnitude of the shocks
- it is common to set the size of the shock used to construct IRFs to 1 standard deviation
- to obtain IRFs in EViews after estimating a VAR, either click on the **Impulse** button, or select **View** $\rightarrow$ **Impulse Response**

Example: House Price Index in Los Angeles and Riverside MSAs



11.4 Impulse-Response Functions

- a problem arises when error terms ε_{1t} and ε_{2t} are correlated - IRFs assume that only one shock occurs at a time
- solution to this problem is quite technical, errors terms ε_{1t} and ε_{2t} are transformed in a certain way so that the covariance of transformed shocks will be zero
- from a practical point of view, the implication is that when error terms are highly correlated ordering of variables in the VAR matters for the shape of IRFs

11.4 Impulse-Response Functions

- in general, the choice of the ordering of the variables should be based on a prior knowledge on the transmission of shocks - in a VAR a shock to variable i is assumed to not have any contemporaneous effect on any variable ordered before it, so $j < i$, it will only have a contemporaneous effect on any variable ordered after it, so $j > i$
- in our example ordering (LA, RI) implies that shock in the Riverside housing market does not have any contemporaneous effect in the Los Angeles market (but effect may be coming in future periods) while a shock in the Los Angeles market contemporaneously affects both Riverside and Los Angeles markets
- fortunately, ordering of variables in a VAR matters only for the IRS, for forecasting purposes ordering does not matter

11.5 Forecasting with VAR

- because every equation in the VAR has an autoregressive specification, forecast can be constructed in a similar recursive way as for a univariate AR
- for example, for a two variable VAR(1)

$$Y_t = c_1 + \alpha_{11} Y_{t-1} + \beta_{11} X_{t-1} + \varepsilon_{1t}$$
$$X_t = c_2 + \alpha_{21} Y_{t-1} + \beta_{21} X_{t-1} + \varepsilon_{2t}$$

under the quadratic loss function, $f_{t,h}^i = \mu_{t+h|t}^i$, so the 1 step ahead forecast is

$$f_{t,1}^Y = c_1 + \alpha_{11} Y_t + \beta_{11} X_t$$
$$f_{t,1}^X = c_2 + \alpha_{21} Y_t + \beta_{21} X_t$$

2 step ahead forecast is

$$f_{t,2}^Y = c_1 + \alpha_{11} f_{t,1}^Y + \beta_{11} f_{t,1}^X$$
$$f_{t,2}^X = c_2 + \alpha_{21} f_{t,1}^Y + \beta_{21} f_{t,1}^X$$

and in general, s step ahead forecast

$$f_{t,s}^Y = c_1 + \alpha_{11} f_{t,s-1}^Y + \beta_{11} f_{t,s-1}^X$$
$$f_{t,s}^X = c_2 + \alpha_{21} f_{t,s-1}^Y + \beta_{21} f_{t,s-1}^X$$

Example: House Price Index in Los Angeles and Riverside MSAs

- to construct a forecast for an estimated VAR in EViews click on **Forecast** button or choose **Proc** $\rightarrow$ **Forecast...**
- the window that open then asks us to provide similar options as with univariate ARs

Example: House Price Index in Los Angeles and Riverside MSAs