Forecasting & Predictive Analytics

Guillaume Chevillon
chevillon@essec.edu

ESSEC Business School

October-December 2017
4th set of slides
Multivariate Time Series

Overview

- Vector Autoregressions
  - Basic examples
  - Properties (stationarity)
  - Granger Causality
  - Impulse Response functions
- Cointegration
  - Examining long-run relationships
  - Determining whether a VAR is cointegrated
  - Error correction models
  - Tests for cointegration (Engle-Granger and Johansen)
- Large dimensional datasets
  - Factor-Augmented VAR
  - Dynamic Panel Data

Why VAR analysis?

- Stationary VARs
  - Determine whether variable feedback into one another
  - Improve forecasts
  - Model the effect of a shock in one series to another
  - Differentiate between short run and long run dynamics
- Cointegration
  - Link random walks
  - Uncover long run relationships
  - Help medium to long term forecasting a lot
- Large Data
  - curse of dimensionality: too many parameters!

VAR Defined

- pth order vector autoregression VAR(p)

$$\mathbf{y}_t = \Phi_0 + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + ... + \Phi_p \mathbf{y}_{t-p} + \epsilon_t$$

- Bivariate VAR(1)

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \phi_{01} \\ \phi_{02} \end{bmatrix} + \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

- simply two nearly independent models

$$\begin{aligned} y_t &= \phi_{01} + \phi_{11,1} y_{t-1} + \phi_{12,1} x_{t-1} + \epsilon_{1t} \\ x_t &= \phi_{02} + \phi_{21,1} y_{t-1} + \phi_{22,1} x_{t-1} + \epsilon_{2t} \end{aligned}$$

Properties of a VAR

$$AR(1): y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$$
$$VAR(1): \mathbf{y}_t = \Phi_0 + \Phi_1 \mathbf{y}_{t-1} + \epsilon_t$$

- Moments
  - Mean: $(\mathbf{I}_K - \Phi_1)^{-1} \Phi_0$
  - Variance: $(\mathbf{I}_K - \Phi_1 \otimes \Phi_1)^{-1} vec(\Sigma_\epsilon)$
  - sth autocovariance: $\Phi_1^s \mathsf{V}[\mathbf{y}_t]$
  - -sth autocovariance: $\mathsf{V}[\mathbf{y}_t] \Phi_1^{s\prime}$
- Stationarity:
  - $|\lambda_i| < 1$ where λ_i are the eigenvalues of Φ_1
  - also roots of $|\mathbf{I}_K - \Phi_1 L| = \det(\mathbf{I}_K - \Phi_1 L)$ are outside the unit circle

Stock and Bond VAR

- VWM from CRSIP; 10 year bond returns from FRED
- May 1953 to December 2004 (620 months)

$$\begin{bmatrix} VWM_t \\ 10YR_t \end{bmatrix} = \begin{bmatrix} \phi_{01} \\ \phi_{02} \end{bmatrix} + \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} VWM_{t-1} \\ 10YR_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

- Market model

$$VWM_t = \phi_{01} + \phi_{11,1} VWM_{t-1} + \phi_{12,1} 10YR_{t-1} + \epsilon_{1t}$$

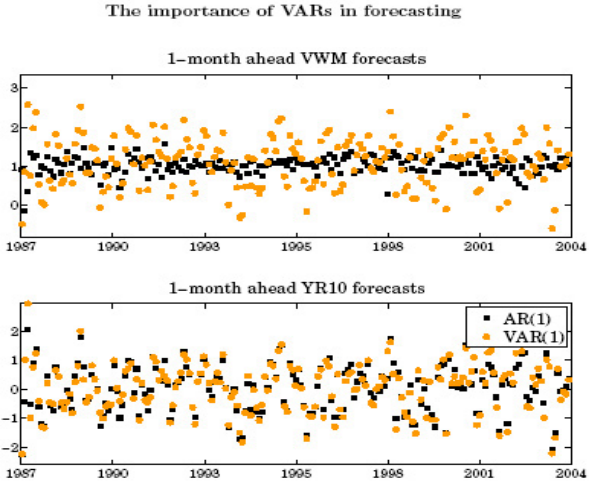
Long bond model

$$10YR_t = \phi_{02} + \phi_{21,1} VWM_{t-1} + \phi_{22,1} 10YR_{t-1} + \epsilon_{2t}$$

- Estimates

$$\begin{bmatrix} VWM_t \\ 10YR_t \end{bmatrix} = \begin{bmatrix} \underset{(0.00)}{0.996} \\ \underset{(0.68)}{0.046} \end{bmatrix} + \begin{bmatrix} \underset{(0.76)}{0.012} & \underset{(0.00)}{0.239} \\ \underset{(0.03)}{-0.058} & \underset{(0.00)}{0.334} \end{bmatrix} \begin{bmatrix} VWM_{t-1} \\ 10YR_{t-1} \end{bmatrix} + \Bigg[$$

Comparing VAR and AR forecasts



Campbell's VAR

- Campbell (1996) asset pricing paper: State vector evolves as a VAR
- State variables include
  - VWM: Value Weighted Market
  - LBR: Real labor income growth
  - RTB: Relative T-Bill (to recent values)
  - TERM: 10 year minus 3 month
  - DIV: dividend yield

	VWM_{t-1}	LBR_{t-1}	RTB_{t-1}	$TERM_{t-1}$	DIV_{t-1}
VWM_t	0.045	91.040	-0.265	0.444	-26.881
	(0.29)	(0.01)	(0.90)	(0.03)	(0.80)
LBR_t	0.000	-0.134	-0.002	0.000	-0.175
	(0.18)	(0.00)	(0.56)	(0.34)	(0.21)
RTB_t	-0.001	0.668	0.628	-0.020	1.936
	(0.09)	(0.17)	(0.00)	(0.00)	(0.22)
$TERM_t$	-0.010	-6.972	0.176	0.983	13.639
	(0.00)	(0.00)	(0.21)	(0.00)	(0.05)
DIV_t	0.000	-0.011	0.000	-0.000	-0.130
	(0.52)	(0.43)	(0.98)	(0.01)	(0.00)

Model Selection

- Step 1: Pick maximum lag length
- Information criteria

$$\mathsf{AIC} : \ln |\Sigma (p)| + \frac{2K^2 p}{T}$$
$$\mathsf{SC} : \ln |\Sigma (p)| + \frac{K^2 p \ln T}{T}$$

  $\Sigma (p)$.is the variance/covariance of the residuals using p lags; $|\cdot|$ is the determinant

- Hypothesis testing based
  - General to Specific
  - Specific to General
- Likelihood Ratio

$$\left(T - p_2 K^2\right) \left(\ln |\Sigma (p_1)| - \ln |\Sigma (p_2)|\right) \rightarrow \chi^2_{(p_2 - p_1) K^2}$$

Identification in Campbell's VAR

Maximum lag: 12 months

Lags	AIC	SIC	LR	P-val
0	6.78	5.91	14924	0.00
1	3.42	2.74	161.9	0.00
2	3.18	2.70	1428	0.00
3	0.28	0.00	35.75	0.07
4	0.29	0.20	24.92	0.46
5	0.32	0.43	120.9	0.00
6	0.11	0.41	23.92	0.52
7	0.14	0.64	30.21	0.21
8	0.15	0.84	22.16	0.62
9	0.17	1.06	26.47	0.38
10	0.17	1.25	23.39	0.55
11	0.18	1.45	68.83	0.00
12	0.00	1.47	N/A	N/A

Granger Causality

- First fundamentally new concept
- Examines whether lags of one variable are helpful in predicting another
- Defined in the negative: A scalar random variable $\{x_t\}$ is said not to Granger cause $\{y_t\}$ if

$$\mathsf{E}\left[y_t | x_{t-1}, y_{t-1}, x_{t-2}, y_{t-2}, ...\right] = \mathsf{E}\left[y_t | y_{t-1}, y_{t-2}, ...\right]$$

- Translates directly into a restriction in a VAR

$$\left[\begin{array}{c} x_t \\ y_t \end{array}\right] = \left[\begin{array}{cc} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{array}\right] \left[\begin{array}{c} x_{t-1} \\ y_{t-1} \end{array}\right] + \left[\begin{array}{cc} \phi_{11,2} & \phi_{12,2} \\ \phi_{21,2} & \phi_{22,2} \end{array}\right] \left[\begin{array}{c} x_{t-2} \\ y_{t-2} \end{array}\right] + \left[\right.$$

  In this model, $\{x_t\}$ does not GC $\{y_t\}$ if $\phi_{21,1} = \phi_{21,2} = 0$
- In the p lag model the null hypothesis is

$$\mathsf{H}_0 : \phi_{ij,1} = \phi_{ij,2} = ... = \phi_{ij,p} = 0$$

Granger Causality in Campbell's VAR

- Using model with lags 1,3 and 12

	VWM		LBR		RTB		TERM		DIV	
Exclusion	Test	P-val	Test	P-val	Test	P-val	Test	P-val	Test	P-val
VWM	-	-	3.08	0.38	2.07	0.56	15.2	0.00	103.6	0.00
LBR	12.3	0.01	-	-	4.3	0.23	14.4	0.00	0.678	0.88
RTB	2.81	0.42	10.1	0.02	-	-	15	0.00	7.22	0.07
TERM	12.4	0.01	3.26	0.35	288.3	0.00	-	-	0.54	0.91
DIV	2.63	0.45	3.43	0.33	16.3	0.00	8.9	0.03	-	-
All	31.5	0.00	27.1	0.01	351.9	0.00	51.9	0.00	135.4	0.00

Impulse Response functions

Second fundamentally new concept

- Complicated dynamics of a VAR make direct interpretation of coefficients difficult
- Solution is to examine impulse responses
- The impulse response function of y_i with respect to a shock in ϵ_j, for any j and i, is defined as the change in $y_{i,t+s}$, $s \geq 0$ for a unit shock in ffl$_{j,t}$
  - Hard to decipher
- As long as y_t is covariance stationarity it must have a VMA representation,

$$\mathbf{y}_t = \mu + \epsilon_t + \Theta_1 \epsilon_{t-1} + \Theta_2 \epsilon_{t-2} + ...$$

  Θ_j are the impulse responses!
- Why?
  - Directly measure the effect in period j of any shock

Considerations on shocks

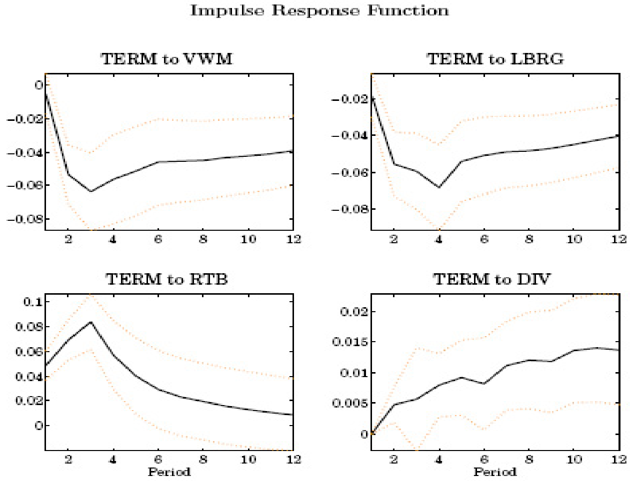
- Simple Bivariate VAR

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \phi_{01} \\ \phi_{02} \end{bmatrix} + \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

- How you “shock” matters
- Depends on the correlation between ϵ_{1t} and ϵ_{2t}
- 3 methods:
  - ignore autocorrelation and just shock ϵ_{jt} with a 1 standard deviation shock
  - use Choleski decomposition to factor Σ_ϵ and use $\Sigma^{1/2}\mathbf{e}$
  - use spectral decomposition to factor Σ_ϵ and use $\Sigma^{1/2}\mathbf{e}$

$$\Sigma = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}, \ \Sigma_C^{1/2} = \begin{bmatrix} 1 & 0 \\ .5 & .866 \end{bmatrix}, \ \Sigma_S^{1/2} = \begin{bmatrix} .96 & .26 \\ .26 & .96 \end{bmatrix}$$

Impulse Responses from Campbell's VAR



Cointegration

Cointegration is the VAR version of unit roots

- Establishes long run relationships between two unit root variables
  - Consumption: Unit root
  - Income: Unit root
  - Consumption - Income : ????
- $\mathsf{I}(1)$: Unit root, $\Delta\mathsf{I}(1) \sim \mathsf{I}(0)$
- $\mathsf{I}(0)$: Stationary

Definition

Assume x_t and y_t are $\mathsf{I}(1)$. They are cointegrated if there exists a vector $\beta = (\beta_1, \beta_2)'$ with both elements non-zero such that

$$\beta_1 x_t + \beta_2 y_t \sim \mathsf{I}(0)$$

-
  - Strong links between x_t and y_t
  - Both are random walks but difference is mean reverting
  - Mean reversion to the trend (stochastic trend)

Examples of Cointegration and Common Trends in Economics and Finance

- Cointegration naturally arises in economics and finance. In economics, cointegration is most often associated with economic theories that imply **equilibrium relationships** between time series variables
- In economics
  - The permanent income model implies cointegration between consumption and income, with consumption being the common trend.
  - Money demand models imply cointegration between money, income, prices and interest rates.
  - Growth theory models imply cointegration between income, consumption and investment, with productivity being the common trend.

Examples cont'd

- In macro-finance
  - Purchasing power parity implies cointegration between the nominal exchange rate and foreign and domestic prices.
  - Covered interest rate parity implies cointegration between forward and spot exchange rates.
  - The Fisher equation implies cointegration between nominal interest rates and inflation:
    $(1 + r_{t+1}) = (1 + i_t) \,/\, (1 + \pi_{t+1})$
  - The expectations hypothesis of the term structure implies cointegration between nominal interest rates at different maturities.
  - The present value model of stock prices states that a stock's price is an expected discounted present value of its expected future dividends or earnings.

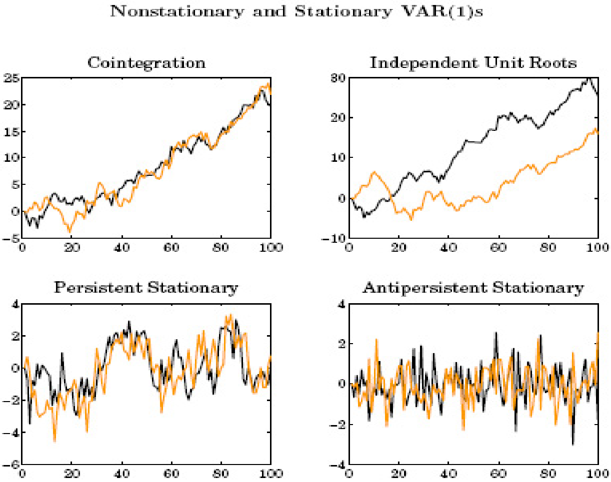
Remarks 1/2

- The equilibrium relationships implied by these economic theories are referred to as long-run equilibrium relationships, because the economic forces that act in response to deviations from equilibrium may take a long time to restore equilibrium.
- As a result, cointegration is modeled using long spans of low frequency time series data measured monthly, quarterly or annually.

Remarks 2/2

- In finance, cointegration may be a high frequency relationship or a low frequency relationship. Cointegration at a high frequency is motivated by arbitrage arguments.
  - The Law of One Price implies that identical assets must sell for the same price to avoid arbitrage opportunities. This implies cointegration between the prices of the same asset trading on different markets, for example.
  - Similar arbitrage arguments imply cointegration between spot and futures prices, and spot and forward prices, and bid and ask prices.
- Here the terminology long-run equilibrium relationship is somewhat misleading because the economic forces acting to eliminate arbitrage opportunities work very quickly. Cointegration is appropriately modeled using short spans of high frequency data in seconds, minutes, hours or days.

Stationary and Non-stationary VARs



How do we know whether a VAR is cointegrated?

- $$\begin{aligned} \mathbf{y}_t &= \Phi_{ij}\mathbf{y}_{t-1} + \epsilon_t \\ \Phi_{11} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} & \quad \Phi_{12} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \lambda_i = 1, 0.6 & \quad \lambda_i = 1, 1 \\ \Phi_{21} = \begin{bmatrix} .7 & .2 \\ .2 & .7 \end{bmatrix} & \quad \Phi_{12} = \begin{bmatrix} -.3 & .3 \\ .1 & -.2 \end{bmatrix} \\ \lambda_i = 0.9, 0.5 & \quad \lambda_i = -0.43, -0.06 \end{aligned}$$

- Eigenvalue condition determines whether a VAR(1) is cointegrated
  - Cointegrated if only 1 eigenvalue is unity.
  - If all less than zero: ?
  - If both 1: two independent unit roots

Error Correction Models

- Major point of cointegration: cointegration $\Leftrightarrow$ Error correction model
- What is an ECM?
  - Cointegrated VAR

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

  - ECM

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} -.2 & .2 \\ .2 & -.2 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

  - Normalized form

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} -.2 \\ .2 \end{bmatrix} [1 \quad -1] \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

  - $[1 \quad -1]$ is the cointegrating vector
  - $[-2 \quad 2]$ measures the speed of adjustment

Testing for Cointegration

Two tests for cointegration

- Engle-Granger
  - Simple and intuitive
  - Applicable with 1 cointegrating relationship
  - Exploits simple property of cointegrating relationship: difference is $\mathsf{I}(0)$
  - Most of the work is a simple OLS

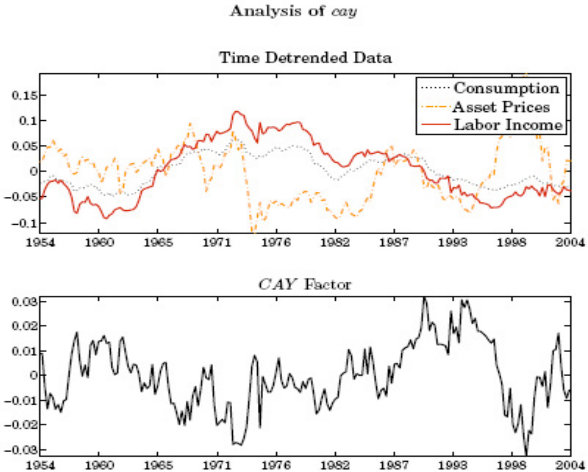
    $$y_t = \beta x_t + \epsilon_t$$

    Rest of work is testing $\widehat{\epsilon}_t$ for a unit root
- Johansen
  - More general

    $$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + \sum_{k=1}^{p} \Gamma_k \Delta \mathbf{y}_{t-k} + \epsilon_t$$

  - Tests the rank of $\Pi = \alpha\beta'$ directly

Consumption-Aggregate Wealth



Large Dimensional Datasets

- So far K variables in VAR(p) with $(pK+1)\,K$ parameters $<< T$ observations: what if $pK^2 >> T$?
- reduce dimensionality of the system:
    - Variable selection (e.g. autometrics, LASSO....)
    - (dynamic) Factor Extraction (PCA, PLS...)
- Factor models work better for forecasting.

Forecasting with Factors

- Common Factors: for each variable x_{it}

$$x_{it} = \lambda_i (L)' F_t + \varepsilon_{xit}$$

  so this implies for y_t

$$y_{t+1} = \beta (L) F_t + \gamma (L) y_t + u_{t+1}$$

- The factors F_t are often obtained nonparametrically, e.g. via Principal Component Estimation.
- Factor-Augmented VAR
- Reduce the number of variables K.
  - Simplest approach: use Principal Component Analysis.
  - (dynamic) Factor Extraction