Lecture 5: Vector Autoregressions and Factor Models

UCSD, Winter 2017

Allan Timmermann[1]

[1]UC San Diego

1. Vector Autoregressions

2. Forecasting with VARs
   - Present value example
   - Impulse response analysis

3. Cointegration

4. Forecasting with Factor Models

5. Forecasting with Panel Data

From univariate to multivariate models

- Often information other than a variable's own past values are relevant for forecasting
- Think of forecasting Hong Kong house prices
  - exchange rate, GDP growth, population growth, interest rates might be relevant
  - past house prices in Hong Kong also matter (AR model)
- In general we can get better models by using richer information sets
- How do we incorporate additional information sources?
  - Vector Auto Regressions (VARs) (small set of predictors)
  - Factor models (many possible predictors)

Vector Auto Regressions (VARs)

- Vector autoregressions generalize univariate autoregressions to the multivariate case by letting y_t be an $n \times 1$ vector and so extend the information set to $\mathcal{I}_t = \{y_{it}, y_{it-1}, ..., y_{i1}\}$ for $i = 1, ..., n$
- Many of the properties of VARs are simple multivariate generalizations of the univariate AR model
- The **Wold representation theorem** also extends to the multivariate case and hence VARs and VARMA models can be used to approximate covariance stationary multivariate (vector) processes
- VARMA: Vector AutoRegressive Moving Average

VARs: definition

- A pth order VAR for an $n \times 1$ vector y_t takes the form:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + ... + A_p y_{t-p} + \varepsilon_t, \ \varepsilon_t \sim WN(0, \Sigma)$$

- $A_i : n \times n$ matrix of autoregressive coefficients for $i = 1, ..., p$:

$$A_i = \begin{pmatrix} A_{i11} & A_{i12} & \cdots & A_{i1n} \\ A_{i21} & A_{i22} & \cdots & A_{i2n} \\ \vdots & & & \\ A_{in1} & & & A_{inn} \end{pmatrix}$$

- $\varepsilon_t : n \times 1$ vector of innovations. These can be correlated across variables
- VARs have the same regressors appearing in each equation
- Number of parameters: $\underbrace{n^2 \times p}_{A_1,...,A_p} + \underbrace{n}_{c} + \underbrace{n(n+1)/2}_{\Sigma}$

Why do we need VARs for forecasting?

- Consider a VAR with two variables: $y_t = (y_{1t}, y_{2t})'$
- $\mathcal{I}_T = \{y_{1T}, y_{1T-1}, ...y_{11}, y_{2T}, y_{2T-1}, ..., y_{21}\}$
- Suppose y_1 depends on past values of y_2. Forecasting y_1 one step ahead (y_{1T+1}) given $\mathcal{I}_T$ is possible if we know today's values, y_{1T}, y_{2T}
- Suppose we want to predict y_1 two steps ahead (y_{1T+2})
  - Since y_{1T+2} depends on y_{2T+1}, we need a forecast of y_{2T+1}, given $\mathcal{I}_T$
- We need a joint model for predicting y_1 and y_2 given their past values
  - This is provided by the VAR

Example: Bivariate VAR(1)

- Joint model for the dynamics in y_{1t} and y_{2t} :

$$\begin{aligned} y_{1t} &= \phi_{11} y_{1t-1} + \phi_{12} y_{2t-1} + \varepsilon_{1t}, \quad \varepsilon_{1t} \sim WN(0, \sigma_1^2) \\ y_{2t} &= \phi_{21} y_{1t-1} + \phi_{22} y_{2t-1} + \varepsilon_{2t}, \quad \varepsilon_{2t} \sim WN(0, \sigma_2^2) \end{aligned}$$

- Each variable depends on one lag of the other variable and one lag of itself
- ϕ_{12} measures the impact of the past value of y_2, y_{2t-1}, on current y_{1t}. When $\phi_{12} \neq 0$, y_{2t-1} affects y_{1t}
- ϕ_{21} measures the impact of the past value of y_1, y_{1t-1}, on current y_{2t}. When $\phi_{21} \neq 0$, y_{1t-1} affects y_{2t}
- The two variables can also be **contemporaneously** correlated if the innovations ε_{1t} and ε_{2t} are correlated and are influenced by common shocks:

$$Cov(\varepsilon_{1t}, \varepsilon_{2t}) = \sigma_{12}$$

If $\sigma_{12} \neq 0$, shocks to y_{1t} and y_{2t} are contemporaneously correlated

Forecasting with Bivariate VAR I

- One-step-ahead forecast given $\mathcal{I}_T = \{y_{1T}, y_{2T}, ..., y_{11}, y_{21}\}$:

$$\begin{aligned} f_{1T+1|T} &= \phi_{11} y_{1T} + \phi_{12} y_{2T} \\ f_{2T+1|T} &= \phi_{21} y_{1T} + \phi_{22} y_{2T} \end{aligned}$$

- To compute two-step-ahead forecasts, use the chain rule:

$$\begin{aligned} f_{1T+2|T} &= \phi_{11} f_{1T+1|T} + \phi_{12} f_{2T+1|T} \\ f_{2T+2|T} &= \phi_{21} f_{1T+1|T} + \phi_{22} f_{2T+1|T} \end{aligned}$$

- Using the expresssions for $f_{1T+1|T}$ and $f_{2T+1|T}$, we have

$$\begin{aligned} f_{1T+2|T} &= \phi_{11}(\phi_{11} y_{1T} + \phi_{12} y_{2T}) + \phi_{12}(\phi_{21} y_{1T} + \phi_{22} y_{2T}) \\ f_{2T+2|T} &= \phi_{21}(\phi_{11} y_{1T} + \phi_{12} y_{2T}) + \phi_{22}(\phi_{21} y_{1T} + \phi_{22} y_{2T}) \end{aligned}$$

Forecasting with Bivariate VAR II

- Collecting terms, we have

$$\begin{aligned} f_{1T+2|T} &= (\phi_{11}^2 + \phi_{12}\phi_{21})y_{1T} + \phi_{12}(\phi_{11} + \phi_{22})y_{2T} \\ f_{2T+2|T} &= \phi_{21}(\phi_{11} + \phi_{22})y_{1T} + (\phi_{12}\phi_{21} + \phi_{22}^2)y_{2T} \end{aligned}$$

- To forecast y_1 two steps ahead we need to forecast both y_1 and y_2 one step ahead.
  - This can only be done if we have forecasting models for both y_1 and y_2
- Therefore, we need to use a VAR for multi-step forecasting of time series that depend on other variables

Predictive (Granger) causality

- Clive Granger (1969) used a variable's predictive content to develop a definition of causality that depends on the conditional distribution of the predicted variable, given other information
- Statistical concept of causality closely related to forecasting
- Basic principles:
  - cause should precede (come before) effect
  - a causal series should contain information useful for forecasting that is not available from the other series (including their past)
- Granger causality in the bivariate VAR:
  - If $\phi_{12} = 0$, then y_2 does **not** Granger cause y_1 : past values of y_2 do not improve our predictions of future values of y_1
  - If $\phi_{21} = 0$, then y_1 does **not** Granger cause y_2 : past values of y_1 do not improve our predictions of future values of y_2
  - For all other values of ϕ_{12} and ϕ_{21}, y_1 will Granger cause y_2 and/or y_2 will Granger cause y_1
  - Include more lags?

Granger causality tests

- Each variable predicts every other variable in the general VAR
- In VARs with many variables, it is quite likely that some variables are not useful for forecasting all the other variables
- Granger causality findings might be overturned by adding more variables to the model — y_{2t} may simply predict y_{1t+1} because other information (y_{3t} which causes both y_{1t+1} and y_{2t+1}) has been left out (omitted variable)

Estimation of VARs

- In sufficiently large samples and under conventional assumptions, the least squares estimates of $(A_1, ..., A_p)$ will be normally distributed around the true parameter value
- Standard errors for each regression are computed using the OLS estimates
- OLS estimation is asymptotically efficient
- OLS estimates are generally biased in small samples

VARs in matlab

- 5-variable sample code on Triton Ed: **varExample.m**
- model = *vgxset*('n',5,'nAR',nlags,'Constant',true); % set up the VAR model
- [modelEstimate,modelStdEr,LL] = *vgxvarx*(model,Y(1:estimationEnd,:)); %estimate the VAR
- numParams = *vgxcount*(model); %number of parameters
- [aicForecast,aicForecastCov] = *vgxpred*(modelEstimates{aicLags,1},forecastHorizon,[],); %forecast with VAR

Difficulties with VARs

- VARs initially became a popular forecasting tool because of their relative simplicity in terms of which choices need to be made by the forecaster
- When estimating a VAR by classical methods, only two choices need to be made to construct forecasts
  - which variables to include (choice of $y_1, ..., y_n$)
  - how many lags of the variables to include (choice of p)
- Risk of overparameterization of VARs is high
  - The general VAR has $n(np+1)$ mean parameters plus another $n(n+1)/2$ covariance parameters
    - For $n = 5, p = 4$ this is 105 mean parameters and 15 covariance parameters
  - Bayesian procedures reduce parameter estimation error by shrinking the parameter estimates towards some target value

Choice of lag length

- Typically we search over VARs with different numbers of lags, p
- With a vector of constants, n variables, p lags, and T observations, the BIC and AIC information criteria take the forms

$$\begin{aligned} BIC(p) &= \ln|\hat{\Sigma}_p| + n(np+1)\frac{\ln(T)}{T} \\ AIC(p) &= \ln|\hat{\Sigma}_p| + n(np+1)\frac{2}{T} \end{aligned}$$

- $\hat{\Sigma}_p = T^{-1}\sum_{t=t}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'$ is the estimate of the residual covariance matrix
- The objective is to identify the model (indexed by p) that **minimizes** the information criterion
- The sample code **varExample.m** chooses the VAR, using up to 12 lags (maxLags)

Multi-period forecasts

- VARs are ideally designed for generating multi-period forecasts. For the VAR(1) specification

$$y_{t+1} = Ay_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim WN(0, \Sigma)$$

the $h-$step-ahead value can be written

$$y_{t+h} = A^h y_t + \sum_{i=1}^{h} A^{h-i} \varepsilon_{t+i}$$

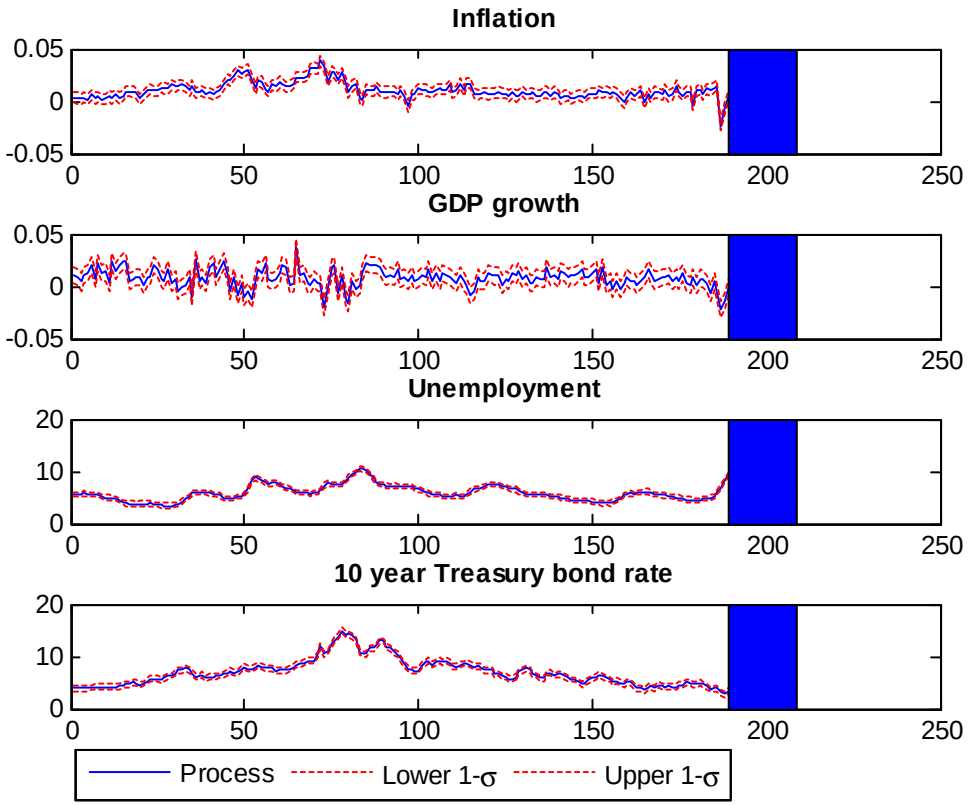
- The forecast under MSE loss is then

$$f_{t+h|t} = A^h y_t$$

- Just like in the case with an AR(1) model!

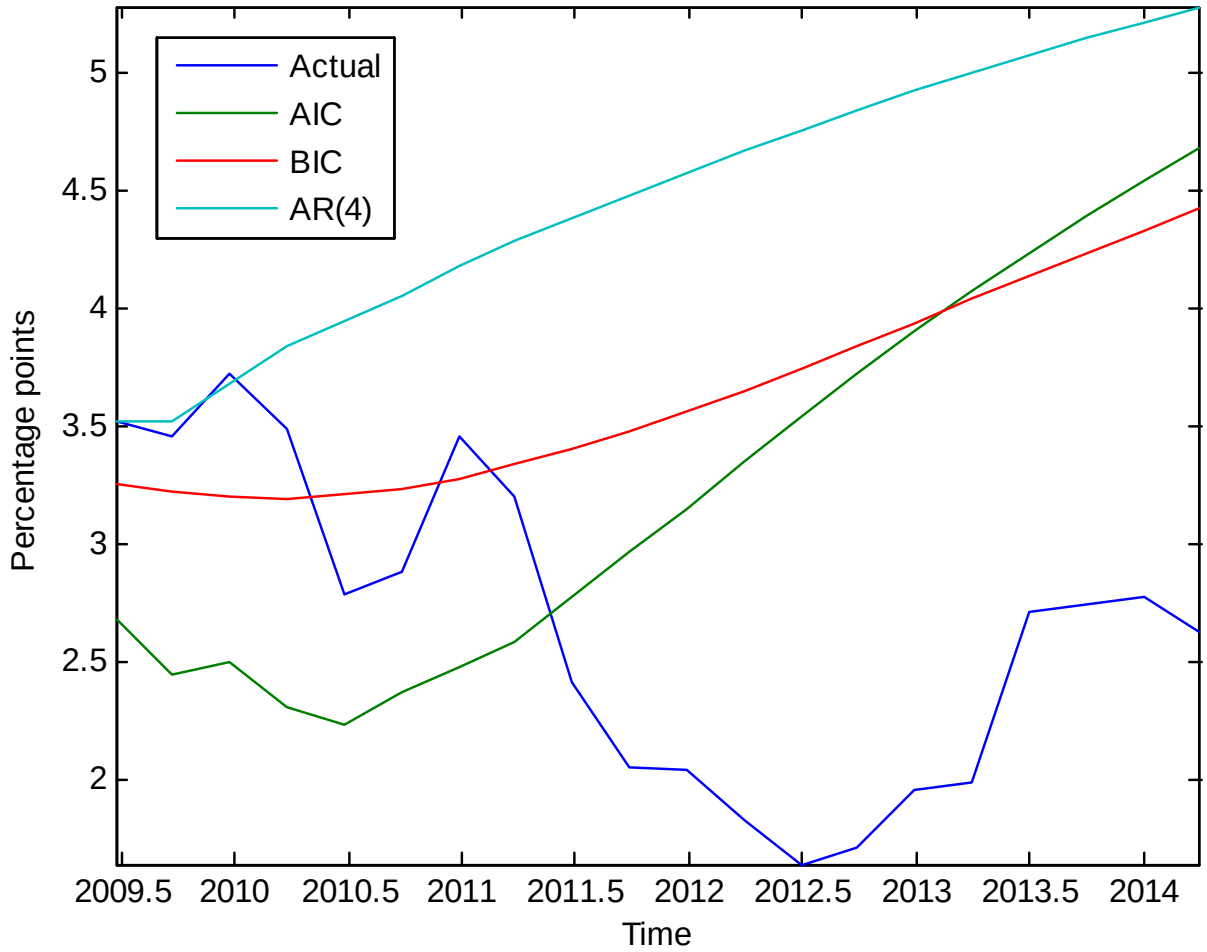
Multi-period forecasts: 4-variable example

- Forecasts using 4-variable VAR with quarterly inflation rate, unemployment rate, GDP growth and 10-year yield
- *vgxplot*(modelEstimates,Y,aicForecast,aicForecastCov); % plot forecast



Multi-period forecasts of 10-year yield (cont.)

Reserve last 5-years of data for forecast evaluation



Example: Campbell-Shiller present value model I

- Campbell and Shiller (1988) express the continuously compounded stock return in period $t+1$, r_{t+1}, as an approximate linear function of the logarithms of current and future stock prices, p_t, p_{t+1} and dividends, d_{t+1}:

$$r_{t+1} = k + \rho p_{t+1} + (1-\rho) d_{t+1} - p_t$$

  ρ is a scalar close to (but below) one, and k is a constant
- Rearranging, we get a recursive equation for log-prices:

$$p_t = k + \rho p_{t+1} + (1-\rho) d_{t+1} - r_{t+1}$$

- Iterating forward and taking expectations conditional on current information, we have

$$p_t = \frac{k}{1-\rho} + (1-\rho) E_t \left[\sum_{j=0}^{\infty} \rho^j d_{t+1+j} \right] - E_t \left[\sum_{j=0}^{\infty} \rho^j r_{t+1+j} \right]$$

Example: Campbell-Shiller present value model II

- Stock prices depend on an infinite sum of expected future dividends and expected returns
- Key to the present value model is therefore how such expectations are formed
  - VARs can address this question since they can be used to generate multi-period forecasts
- To illustrate this point, let z_t be a vector of state variables with $z_{1t} = p_t$, $z_{2t} = d_t$, $z_{3t} = x_t$; x_t are predictor variables
- Define selection vectors $e_1 = (1\ 0\ 0)'$, $e_2 = (0\ 1\ 0)'$, $e_3 = (0\ 0\ 1)'$ so $p_t = e_1' z_t, d_t = e_2' z_t, x_t = e_3' x_t$
- Suppose that z_t follows a VAR(1):

$$\begin{aligned} z_{t+1} &= Az_t + \varepsilon_{t+1} \quad \Rightarrow \\ E_t[z_{t+j}] &= A^j z_t \end{aligned}$$

Example: Campbell-Shiller present value model III

- If expected returns $E_t[r_{t+1+j}]$ are constant and stock prices only move due to variation in dividends, we have (ignoring the constant and assuming that we can invert $(I - \rho A)$)

$$\begin{aligned} p_t &= (1-\rho) E_t \left[\sum_{j=0}^{\infty} \rho^j d_{t+1+j} \right] \\ &= (1-\rho) e_2' \sum_{j=0}^{\infty} \rho^j A^{j+1} z_t = (1-\rho) e_2' A (I - \rho A)^{-1} z_t \end{aligned}$$

- Nice and simple expression for the present value stock price!
- The VAR gives us a simple way to compute expected future dividends $E_t[d_{t+1+j}]$ for all future points in time given the current information in z_t
- Can you suggest other ways of doing this?

Impulse response analysis

- Stationary vector autoregressions (VARs) can equivalently be expressed as vector moving average (VMA) processes:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ...$$

- Impulse response analysis shows how variable i in a VAR is affected by a shock to variable j at different horizons:

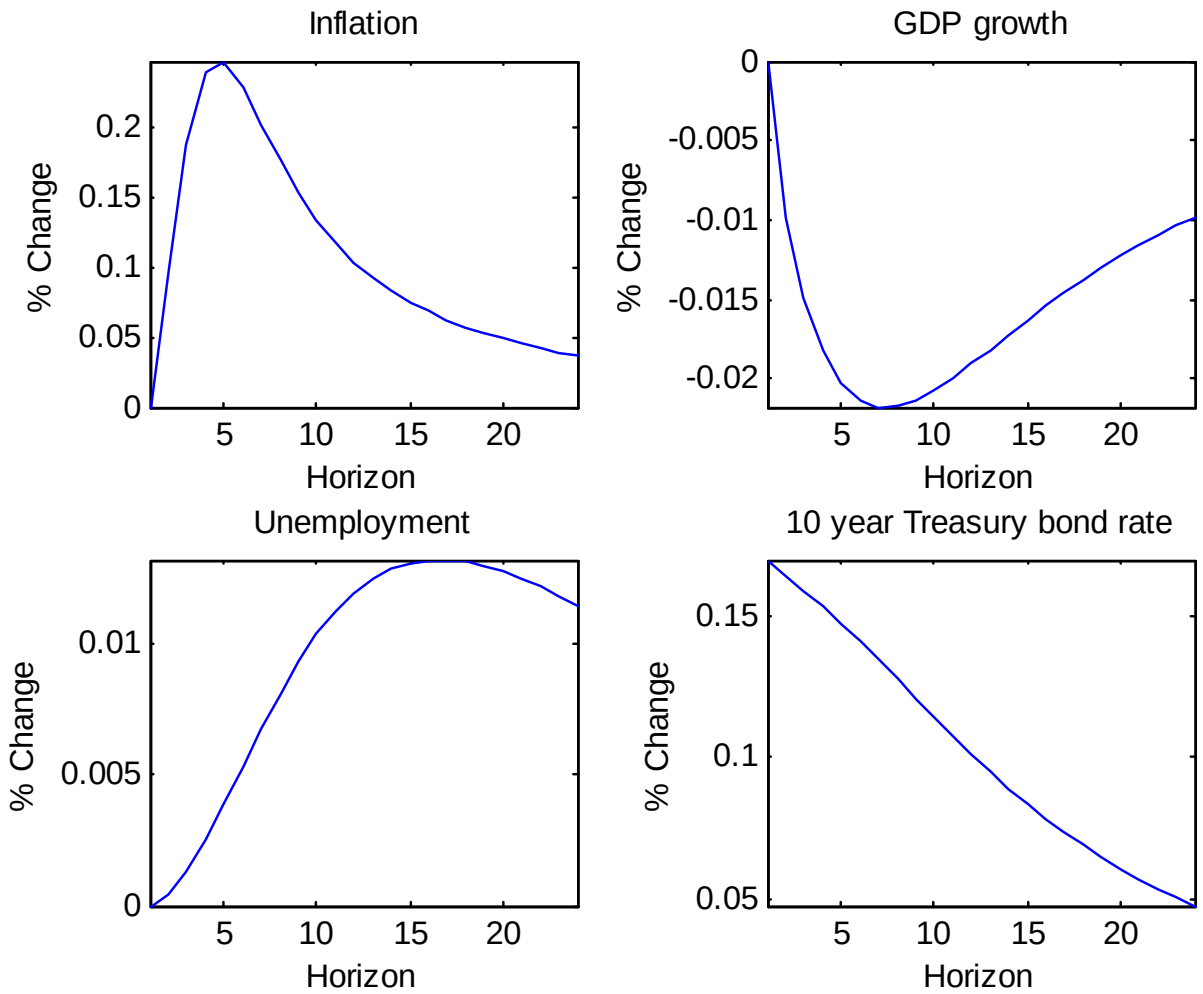
$$\frac{\partial y_{it+1}}{\partial \varepsilon_{jt}} \quad \text{1-period impulse}$$
$$\frac{\partial y_{it+2}}{\partial \varepsilon_{jt}} \quad \text{2-period impulse}$$
$$\frac{\partial y_{it+h}}{\partial \varepsilon_{jt}} \quad h\text{-period impulse}$$

- Suppose we find out that variable j is higher than we expected (by one unit). Impulse responses show how much we revise our forecasts of future values of y_{it+h} due to this information
- How does an interest rate shock affect future unemployment and inflation?

Impulse response analysis in matlab

- Four-variable model: inflation, GDP growth, unemployment, 10-year Treasury bond rate
- impulseHorizon = 24; % 24 months out
- W0 = zeros(impulseHorizon,4); %baseline scenario of zero shock
- W1 = W0;
- W1(1,4) = sqrt(modelEstimates{aicLags,1}.Q(4,4)); % one standard deviation shock to variable number four (interest rate)
- Yimpulse = *vgxproc*(modelEstimates{aicLags,1},W1,[],Y(1:estimationEnd,:)); %impulse response
- Ynoimpulse = *vgxproc*(modelEstimates{aicLags,1},W0,[],Y(1:estimationEnd,:));

Impulse response analysis: shock to 10-year yield



Nobel Prize Award, 2003 press release

"Most macroeconomic time series follow a stochastic trend, so that a temporary disturbance in, say, GDP has a long-lasting effect. These time series are called nonstationary; they differ from stationary series which do not grow over time, but fluctuate around a given value. Clive Granger demonstrated that the statistical methods used for stationary time series could yield wholly misleading results when applied to the analysis of nonstationary data. His significant discovery was that specific combinations of nonstationary time series may exhibit stationarity, thereby allowing for correct statistical inference. Granger called this phenomenon **cointegration**. He developed methods that have become invaluable in systems where short-run dynamics are affected by large random disturbances and long-run dynamics are restricted by economic equilibrium relationships. Examples include the relations between wealth and consumption, exchange rates and price levels, and short and long-term interest rates."

- This work was done at UCSD

Cointegration

- Consider the variables

$$\begin{aligned} x_t &= x_{t-1} + \varepsilon_t && x \text{ follows a random walk (nonstationary)} \\ y_{1t} &= x_t + u_{1t} && y_1 \text{ is a random walk plus noise} \\ y_{2t} &= x_t + u_{2t} && y_2 \text{ is a random walk plus noise} \end{aligned}$$

- $\varepsilon_t, u_{1t}, u_{2t}$ are all white noise (or at least stationary)
- x_t is a unit root process: $(1-L)x_t = \varepsilon_t$, so $L = 1$ is a "root"
- y_1 and y_2 behave like random walks. However, their difference

$$y_{1t} - y_{2t} = u_{1t} - u_{2t}$$

  is stationary (mean-reverting)
- Over the long run, $y_1 - y_2$ will revert to its equilibrium value of zero

Cointegration (cont.)

- Future levels of random walk variables are difficult to predict
- It is much easier to predict *differences* between two sets of cointegrated variables
  - Example: Forecasting the *level* of Brent or WTI (West Texas Intermediate) crude oil prices five years from now is difficult
  - Forecasting the *difference* between these prices (or the logs of their prices) is likely to be easier
  - In practice we often study the logarithm of prices (instead of their level), so percentage differences cannot become too large

Cointegration (cont.)

- If two variables are cointegrated, they must both individually have a stochastic trend (follow a unit root process) and their individual paths can wander arbitrarily far away from their current values
  - There exists a linear combination that ties the two variables closely together
  - Future values cannot deviate too far away from this equilibrium relation
- Granger representation theorem: Equilibrium errors (deviations from the cointegrating relationship) can be used to predict future changes
- Examples of **possible** cointegrated variables:
  - Oil prices in Shanghai and Hong Kong—if they differ by too much, there is an arbitrage opportunity
  - Long and short interest rates
  - Baidu and Alibaba stock prices (pairs trading)
  - House prices in two neighboring cities
  - Chinese A and H share prices for same company. Arbitrage opportunities?

Vector Error Correction Models (VECM)

- Vector error correction models (VECMs) can be used to analyze VARs with nonstationary variables that are cointegrated
  - Cointegration relation restricts the long-run behavior of the variables so they converge to their cointegrating relationship (long-run equilibrium)
- Cointegration term is called the **error-correction** term
  - This measures the deviation from the equilibrium and allows for short-run predictability
  - In the long-run equilibrium, the error correction term equals zero

Vector Error Correction Models (cont.)

- VECM for changes in two variables, y_{1t}, y_{2t} with cointegrating equation $y_{2t} = \beta y_{1t}$ and lagged error correction term $(y_{2t-1} - \beta y_{1t-1})$:

$$\begin{aligned} \Delta y_{1t} &= \alpha_1 \underbrace{(y_{2t-1} - \beta y_{1t-1})}_{\text{lagged error correction term}} + \lambda_1 \Delta y_{1t-1} + \varepsilon_{1t} \\ \Delta y_{2t} &= \alpha_2 (y_{2t-1} - \beta y_{1t-1}) + \lambda_2 \Delta y_{2t-1} + \varepsilon_{2t} \end{aligned}$$

- In the short run y_1 and y_2 can deviate from the equilibrium $y_{2t} = \beta y_{1t}$
  - Lagged error correction term $(y_{2t-1} - \beta y_{1t-1})$ pulls the variables back towards their equilibrium
  - α_1 and α_2 measure the speed of adjustment of y_1 and y_2 towards equilibrium
  - Larger values of α_1 and α_2 mean faster adjustment

Vector Error Correction Models (cont.)

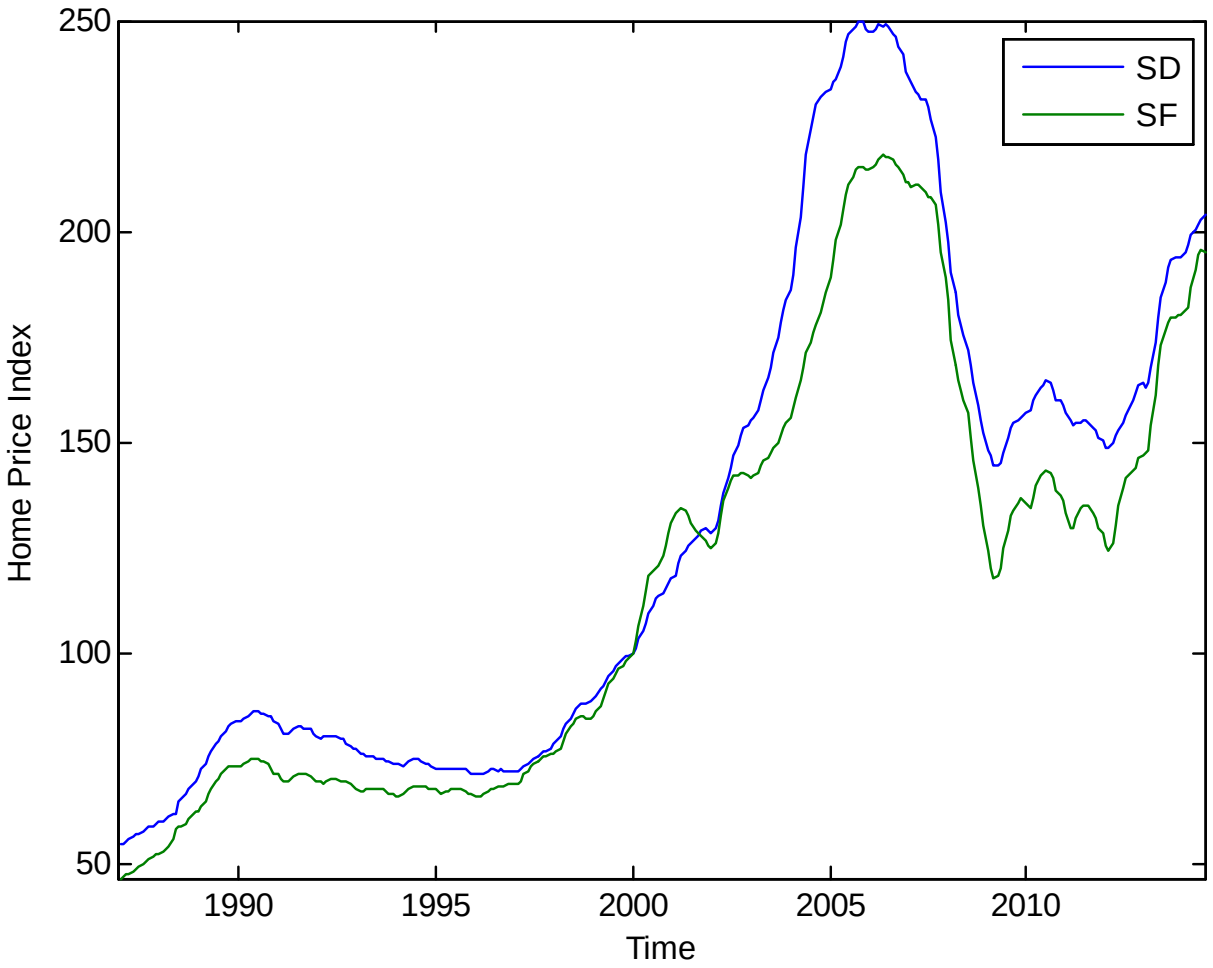
- In many applications (particularly with variables in logs), $\beta = 1$. Then a forecasting model for the changes Δy_{1t} and Δy_{2t} could take the form

$$\Delta y_{1t} = c_1 + \underbrace{\sum_{i=1}^{p} \lambda_{1i} \Delta y_{1t-i}}_{p \text{ AR lags}} + \alpha_1 \underbrace{(y_{2t-1} - y_{1t-1})}_{\text{error correction term}} + \varepsilon_{1t}$$

$$\Delta y_{2t} = c_2 + \sum_{i=1}^{p} \lambda_{2i} \Delta y_{2t-i} + \alpha_2 (y_{2t-1} - y_{1t-1}) + \varepsilon_{2t}$$

- This can be estimated by OLS since you know the cointegrating coefficient, $\beta = 1$
- Include more lags of the error correction term $(y_{2t-1} - y_{1t-1})$? Adjustments may be slow

House prices in San Diego and San Francisco



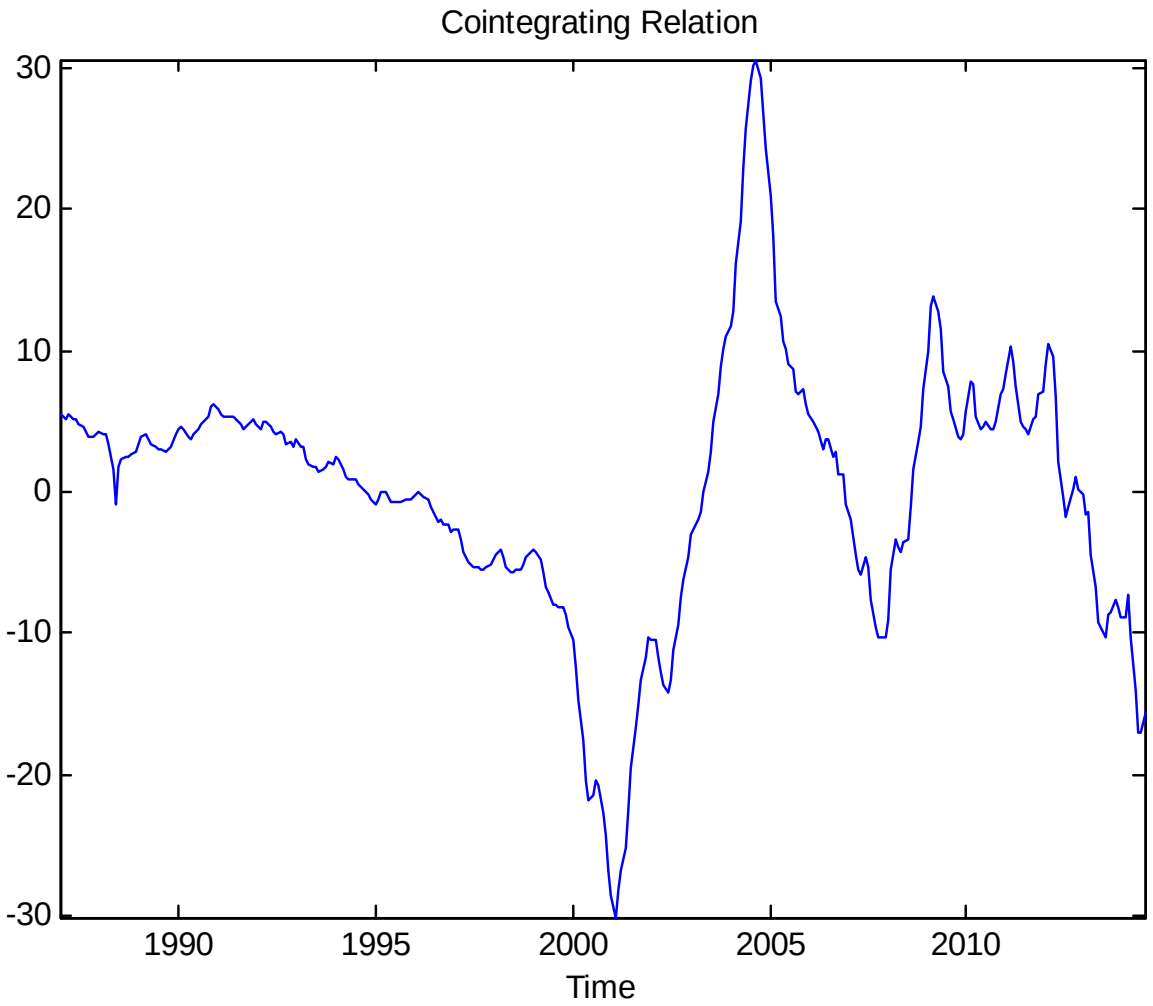
Simple test for cointegration

- Regress San Diego house prices on San Francisco house prices and test if the residuals are non-stationary
  - use logarithm of prices (?)
- Null hypothesis is that there is no cointegration (so there is no linear combination of the two prices that is stationary)
- If you reject the null hypothesis (get a low p-value), this means that the series are cointegrated
- If you don't reject the null hypothesis (high p-value), the series are not cointegrated
- Often test has low power (fails to reject even when the series are cointegrated)

Test for cointegration in matlab

- See **VecmExample.m** file on Triton Ed
- In matlab: *egcitest* : Engle-Granger cointegration test
- $[h, pValue, stat, cValue] = egcitest(Y)$
- "Engle-Granger tests assess the null hypothesis of no cointegration among the time series in Y. The test regresses Y(:,1) on Y(:,2:end), then tests the residuals for a unit root.
- Values of h equal to 1 (true) indicate rejection of the null in favor of the alternative of cointegration. Values of h equal to 0 (false) indicate a failure to reject the null."
- $p-$value of test for SD and SF house prices: 0.9351. We fail to reject the null that the house prices are not cointegrated. Why?

House prices in San Diego and San Francisco



Forecasting with Factor models I

- Suppose we have a very large set of predictor variables, $x_{it}, i = 1, ...n$, where n could be in the hundreds or thousands
- The simplest forecasting approach would be to consider a linear model with all predictors included:

$$y_{t+1} = \alpha + \sum_{i=1}^{n} \beta_i x_{it} + \phi_1 y_t + \varepsilon_{yt+1}$$

- This model can be estimated by OLS, assuming that the total number of parameters, $n+2$, is small relative to the length of the time series, T
- Often $n > T$ and so linear regression methods are not feasible
- Instead it is commonly assumed that the $x-$variables only affect y through a small set of r **common factors**, $F_t = (F_{1t}, ..., F_{rt})'$, where r is much smaller than N (typically less than ten)

Forecasting with Factor models II

- This suggests using a common factor forecasting model of the form

$$y_{t+1} = \alpha + \sum_{i=1}^{r} \beta_{iF} F_{it} + \phi_1 y_t + \varepsilon_{yt+1}$$

- Suppose that $n = 200$ and $r = 3$ common factors
  - The general forecasting model requires fitting 202 mean parameters: $\alpha, \beta_1, ..., \beta_{200}, \phi_1$
  - The simple factor model only requires estimating 5 mean parameters: $\alpha, \beta_{1F}, \beta_{2F}, \beta_{3F}, \phi_1$

Forecasting with Factor models

- The identity of the common factors is usually unknown and so must be extracted from the data
- Forecasting with common factor models can therefore be thought of as a two-step process
  1. Extract estimates of the common factors from the data
  2. Use the factors, along with past values of the predicted variable, to select and estimate a forecasting model
- Suppose a set of factor estimates, $\hat{F}_{it}$, has been extracted. These are then used along with past values of y to estimate a model and generate forecasts of the form:

$$\hat{y}_{t+1|t} = \hat{\alpha} + \sum_{i=1}^{r} \hat{\beta}_{iF} \hat{F}_{it} + \hat{\phi}_1 y_t$$

- Common factors can be extracted using the principal components method

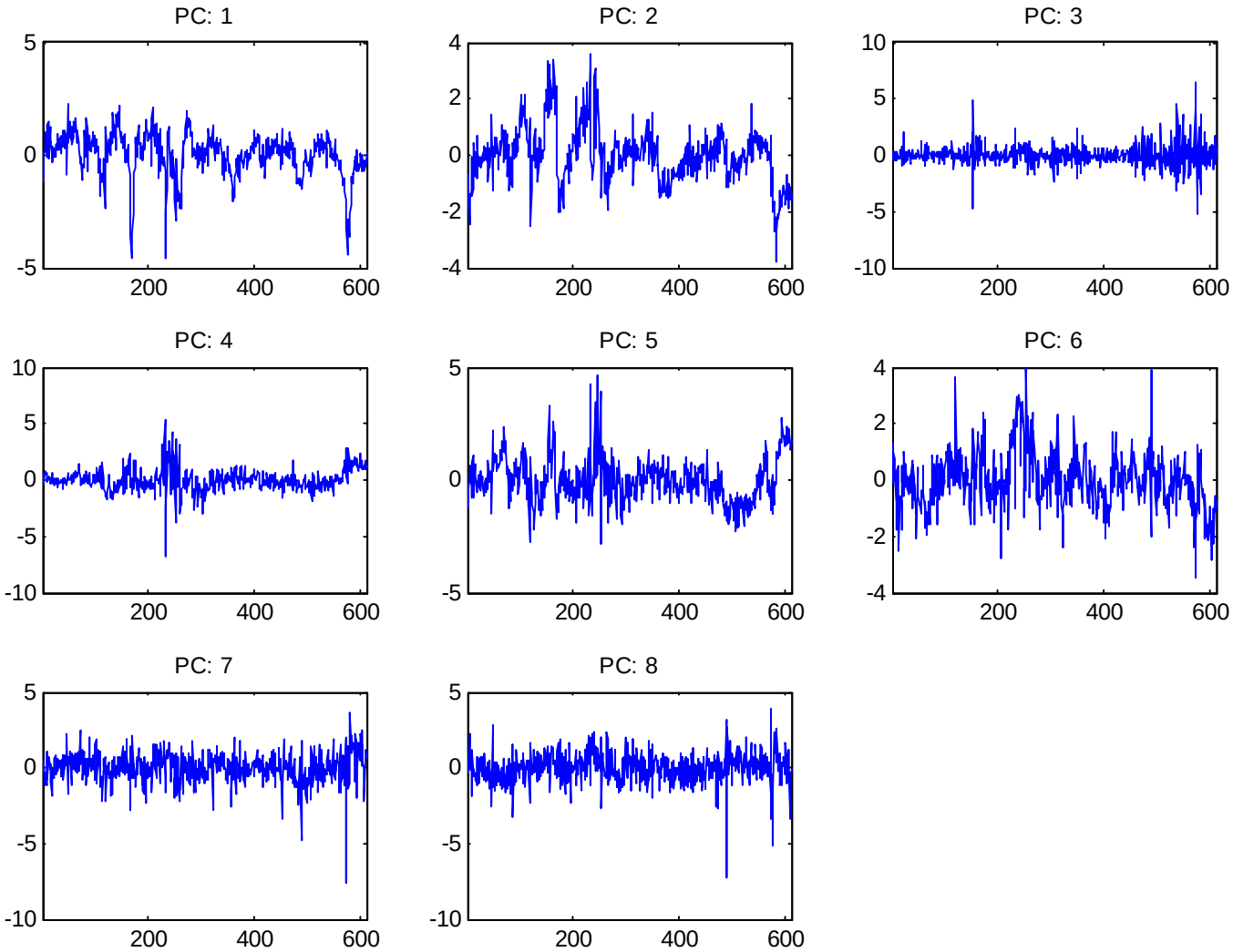
Principal components

- Wikipedia: "Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components. The principal components are orthogonal because they are the eigenvectors of the covariance matrix, which is symmetric. PCA is sensitive to the relative scaling of the original variables."

Empirical example

- Data set with $n = 132$ predictor variables
- Available in macro_raw_data.xlsx on Triton Ed. Uses data from Sydney Ludvigson's NYU website
- Data series have to be transformed (e.g., from levels to growth rates) before they are used to form principal components
- Extract $r = 8$ common factors using principal components methods

Empirical example (cont.): 8 principal components



Forecasting with panel data I

- Forecasting methods can also be applied to cross-sections or panel data
- Key requirement is that the predictors are predetermined in time. For example, we could build a forecasting model for a large cross-section of credit-card holders using data on household characteristics, past payment records etc.
  - The implicit time dimension is that we know whether a payment in the data turned out fraudulent
- Panel regressions take the form

$$y_{it} = \alpha_i + \lambda_t + X_{it}'\beta + u_{it}, \quad i = 1, ..., n, \quad t = 1, ..., T$$

- α_i : fixed effect (e.g., firm, stock, or country level)
- λ_t : time fixed effect
  - How do we predict λ_{t+1}?
- Panel models can be estimated using regression methods
- Do slope coefficients β vary across units (β_i)?
  - bias-variance trade-off