1. Linear time series models

Javier Mencía
javier.mencia@bde.es

Banco de España, Financial Stability Department

Outline of the presentation

Outline

Motivation

- Time series econometrics focuses on the analysis of a collection of random variables over time.

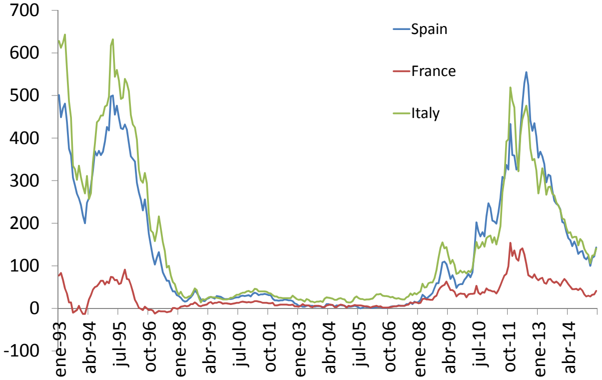
$$y_1, y_2, \cdots, y_T$$

- Time series models provide the framework to study the dynamic structure of the series.
- Time series are sampled at equally spaced intervals, which typically depend on the application: annual, quarterly, monthly, weekly, daily and intra-daily.

Outline

Correlation

Long term sovereign spreads with respect to German sovereign yields (bp)



Source: ECB Statistical Data Warehouse. Harmonised long-term interest rates for convergence assessment purposes.

- The correlation coefficient between two random variables X and Y measures the strength of the linear dependence between them:

$$\rho = cor(x_t, y_t) = \frac{cov(x_t, y_t)}{\sqrt{Var(x_t)Var(y_t)}} \in [-1, 1]$$

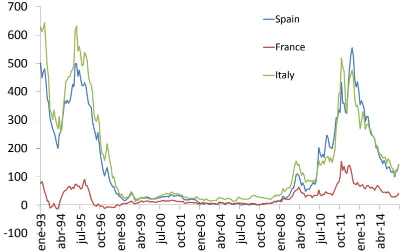
- $\rho = 0$ implies absence of linear dependence, while $\rho = \pm 1$ implies that $x_t = a + by_t$.
- For a finite sample with T observations, the correlation coefficient can be estimated by the sample counterpart:

$$\hat{\rho} = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2 \sum_{t=1}^{T}(y_t - \bar{y})^2}}$$

where $\bar{x}$ and $\bar{y}$ are the sample means.

Correlation

Correlation between sovereign spreads

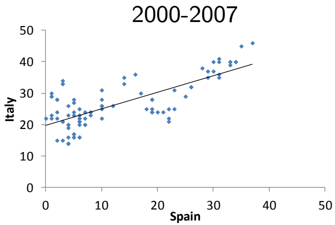


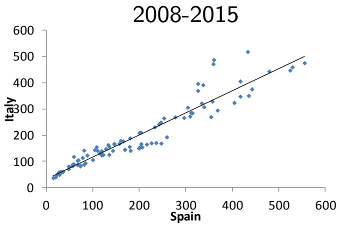
Correlation between long term sovereign spreads (%)

	1993-1999	2000-2007	2008-2015
ES-FR	77	88	83
ES-IT	99	75	95
IT-FR	75	73	93

Correlation

Correlation between sovereign spreads. Scatter plots





Correlation between long term sovereign spreads (%)

	2000-2007	2008-2015
ES-IT	75	95

- We will revise correlation in the context of multivariate dependence models (Part 3).

Autocorrelation

- The correlation between x_t and its own $i-th$ lag is called the autocorrelation of order i:

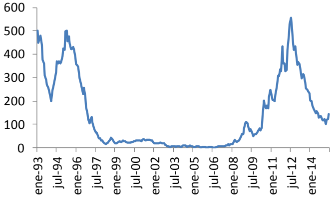
$$\rho_i = cor(x_t, x_{t-i}) = \frac{cov(x_t, x_{t-i})}{\sqrt{Var(x_t)Var(x_{t-i})}}$$

- By construction, $\rho_0 = 1$.
- The autocorrelation is a very useful concept to characterise the dynamic structure of a time series.

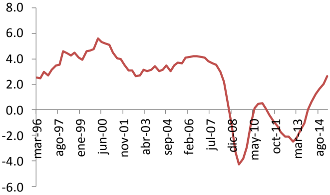
Autocorrelation

Examples based on Spanish data

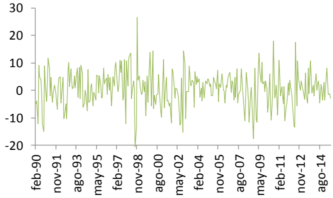
Sovereign spreads (bp)



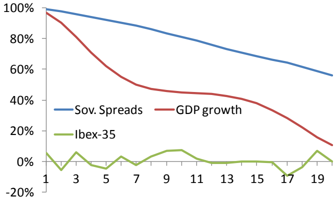
Real GDP annual growth rates (%)



Ibex-35 monthly returns (%)



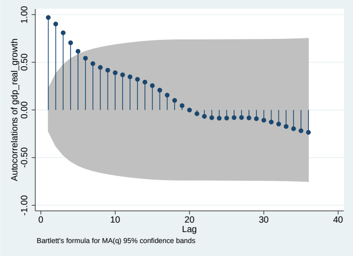
Autocorrelations (%)



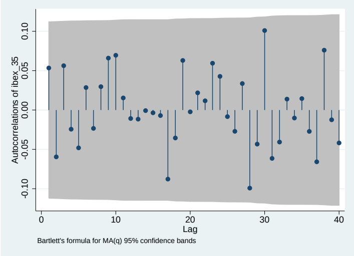
Autocorrelation

Examples based on Spanish data. Confidence intervals

Real GDP annual growth rates (%)



Ibex-35 monthly returns (%)



Stationarity

- Why computing the correlogram for the Ibex-35 monthly returns instead of levels?
- Stationarity is a crucial concept in time series analysis.
- Weak stationarity of x_t: expectation, variance and autocovariances are the same at every t.
  - $E(x_t) = E(x_{t+s}) = \mu$ is a finite constant.
  - $cov(x_t, x_{t-l}) = \gamma_l$ is finite and only depends on the lag l (not on t).
- The standard formulas to compute different statistics are only valid if all the observations come from the same distribution:

$$
\begin{aligned}
\bar{x}_a &= \frac{1}{T}(x_1 + x_2 + \cdots + x_T) \to \mu \\
\bar{x}_b &= \frac{1}{T}(x_{T+1} + x_{T+2} + \cdots + x_{2T}) \to \mu
\end{aligned}
$$

Stationarity

- Main sources of non-stationarity:
  - Deterministic component:

$$x_t = \alpha + \beta t + \varepsilon_t,$$

  - Stochastic component: unit root, random walk or integrated process of order 1 (I(1) for short):

$$\begin{aligned} y_t &= \gamma + y_{t-1} + \varepsilon_t \\ &= \gamma t + y_0 + \varepsilon_1 + \cdots + \varepsilon_t. \end{aligned}$$

- If $\varepsilon_t \sim iidN(0,1)$, then

$$\begin{aligned} x_t &\sim N(\alpha + \beta t, 1) \\ y_t &\sim N(y_0 + \gamma t, t). \end{aligned}$$

Stationarity

- Computing returns or taking first differences is consistent with assuming that there is a unit root:

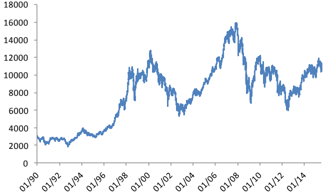
$$\Delta \log y_t = log(y_t) - log(y_{t-1}) \approx \frac{y_t - y_{t-1}}{y_{t-1}}$$

- In general, first differences are more common than detrending a deterministic component.
- In finance, unit roots are consistent with the efficient market hypothesis: in liquid markets, the current price should accurately reflect all the available information. Future prices should only change due to unexpected news, which implies that the best forecast of the price at $t+1$ is the price at t.

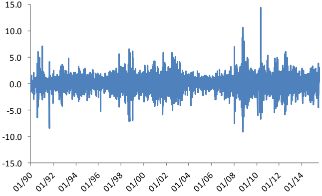
Stationarity

Dickey-Fuller Unit root test versus a deterministic trend for the Ibex-35 daily index

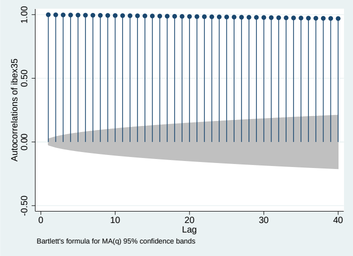
Ibex-35 index (%)



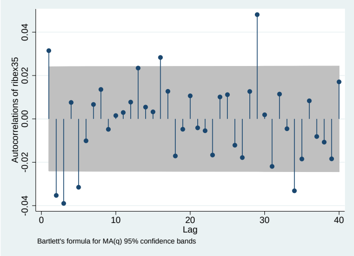
Ibex-35 daily returns (%)



Correlogram for the levels



Correlogram for the returns



Dickey-Fuller Unit root test versus a deterministic trend for the Ibex-35 daily index

- The null hypothesis is $H_0 : \rho = 1$ or $\delta_0 = 0$ versus $H_1 : \delta_0 \neq 0$.

$$y_t = \alpha + \rho y_{t-1} + \delta_1 t + \epsilon_t \quad \Rightarrow \quad \Delta y_t = \alpha + \delta_0 y_{t-1} + \delta_1 t + \epsilon_t$$

- The augmented version of this test controls for serial correlation in ϵ_t by including lags of Δy_t.

Dickey-Fuller Unit root test versus a deterministic trend for the Ibex-35 daily index

- For the Ibex-35 levels, the test does not reject the unit root, regardless of the presence of a trend.

```
. dfgls ibex35, maxlag(3) notrend;

DF-GLS for ibex35                                     Number of obs =   6659

               DF-GLS mu       1% Critical       5% Critical      10% Critical
  [lags]     Test Statistic       Value             Value             Value
------------------------------------------------------------------------------
    3           -0.140           -2.580            -1.950            -1.627
    2           -0.194           -2.580            -1.950            -1.627
    1           -0.247           -2.580            -1.951            -1.627

. dfgls ibex35, maxlag(3);

DF-GLS for ibex35                                     Number of obs =   6659

               DF-GLS tau      1% Critical       5% Critical      10% Critical
  [lags]     Test Statistic       Value             Value             Value
------------------------------------------------------------------------------
    3           -1.917           -3.480            -2.840            -2.553
    2           -1.995           -3.480            -2.841            -2.553
    1           -2.075           -3.480            -2.841            -2.553
```

Dickey-Fuller Unit root test versus a deterministic trend for the Ibex-35 daily index

- Once we compute returns, the test rejects the unit root at all conventional levels.

```
. dfgls ribex35, maxlag(3) notrend;

DF-GLS for ribex35                                   Number of obs =   6658

               DF-GLS mu      1% Critical      5% Critical     10% Critical
  [lags]     Test Statistic       Value            Value            Value
______________________________________________________________________________

    3          -32.869          -2.580           -1.950           -1.627
    2          -40.473          -2.580           -1.950           -1.627
    1          -50.424          -2.580           -1.951           -1.627


. dfgls ribex35, maxlag(3);

DF-GLS for ribex35                                   Number of obs =   6658

               DF-GLS tau     1% Critical      5% Critical     10% Critical
  [lags]     Test Statistic       Value            Value            Value
______________________________________________________________________________

    3          -38.685          -3.480           -2.840           -2.553
    2          -46.403          -3.480           -2.841           -2.553
    1          -56.096          -3.480           -2.841           -2.553
```

Seasonality

- Why computing the correlogram for the GDP annual growth rates when the frequency is quarterly?
- Seasonality refers to the cyclical or periodic behaviour that some economic time series exhibit.
- Typical examples are: GDP, unemployment rates, inflation indices, corporate earnings, ...
- It is a different source of non-stationarity that must be treated when modelling the data.
- Seasonal adjustment is the procedure to remove this effect in the applications where it is of secondary importance.
- Statistical institutes generally provide seasonally adjusted series for some variables (e.g. the GDP series used in the example).

Seasonality

- Similarly to stationarity, seasonality can be deterministic, stochastic or a combination of both.
- For most practical purposes, seasonal detrending removes most of the seasonal effects that we will encounter.
- Examples:
  - Quarterly GDP: annual first differences.

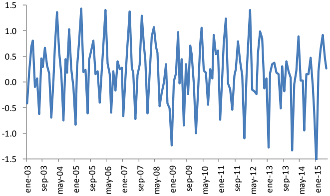
$$\Delta_4 GDP_t = GDP_t - GDP_{t-12}$$

  - Inflation index: 12-month first differences.

$$\Delta_{12} Inflation_t = Inflation_t - Inflation_{t-12}$$

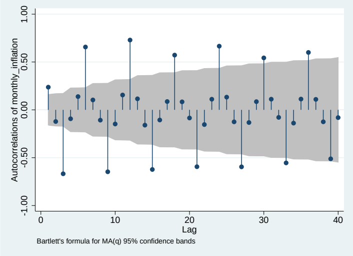
- Even if you are using seasonally adjusted data, it is a good policy to use the seasonally detrended series for modelling purposes.

Monthly returns (%)

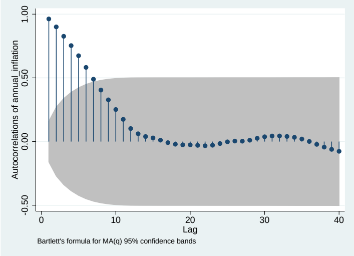


Annual returns (%)

Correlogram for the monthly returns



Correlogram for the annual returns



Outline

White noise and linear models

- A time series ε_t is white noise if it is a sequence of independent and identically distributed random variables with finite mean and variance.
- For instance, ε_t is Gaussian white noise if $\varepsilon_t \sim iid\ N(0, \sigma^2)$ for all t and $cov(\varepsilon_t, \varepsilon_{t-l}) = 0$ for $l = 1, 2, \cdots$.
- A time series x_t is linear if it can be expressed as a linear combination of white noise random variables:

$$x_t = \mu + \alpha_0 \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots$$

- This is known as the Wold decomposition. It is crucial to answer impulse response questions: how a shock at t affects the future.
- Linear time series models are econometric models to describe the pattern followed by the α weights with only a few parameters.

Moving Average

- A Moving Average of order p is a linear combination of $p+1$ white noise variables

$$x_t = \mu + \varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_2\varepsilon_{t-2} + \cdots + \alpha_p\varepsilon_{t-p},$$

- By construction, the correlogram of an $MA(p)$ process is zero for autocorrelations of order higher than p.
- The most simple example is the $MA(1)$ process,

$$x_t = \mu + \varepsilon_t + \alpha_1\varepsilon_{t-1},$$

  where $E(x_t) = \mu$, $Var(x_t) = (1+\alpha_1^2)\sigma^2$ and $cor(x_t, x_{t-1}) = \frac{\alpha_1}{1+\alpha_1^2}$.
- Hence, an $MA(1)$ can only generate first order autocorrelations between -0.5 and 0.5.

MA(1)

- The residual ε_t is unobservable, but we can write it as a function of the past of x_t:

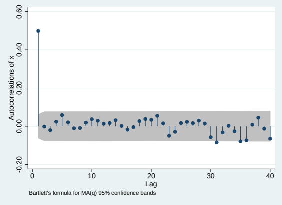
$$\varepsilon_t = (x_t - \mu) - \alpha_1(x_{t-1} - \mu) + \alpha_1^2(x_{t-2} - \mu) - \alpha_1^3(x_{t-3} - \mu) + \cdots$$

- This decomposition only converges to a finite value if $|\alpha_1| < 1$.
- If $|\alpha_1| < 1$, we can say that the process is invertible (i.e. we can back out the residuals).

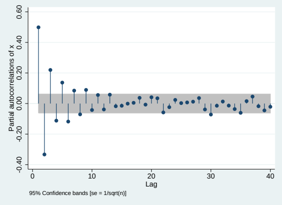
MA identification with correlograms and partial correlograms

$MA(1)$ simulation example with $\alpha_1 = 0.8$. $T = 1,000$

Correlogram



Partial correlogram



- The partial autocorrelation of order l of x_t is the $l-$th order correlation, conditioning on the previous lags.
- It can be obtained as δ_l from the following regression:

$$x_t = \delta_0 + \delta_1 x_{t-1} + \cdots + \delta_{l-1} x_{t-l+1} + \delta_l x_{t-l} + u_t$$

- The correlogram is particularly useful to identify the order of the $MA(1)$ process.

Autorregressive processes

- In an autorregressive process of order p, or $AR(p)$, a variable x_t is modelled as a function of its own first p lags.

$$x_t = \mu + \rho_1 x_{t-1} + \rho_2 x_{t-2} + \cdots + \rho_p x_{t-p} + \varepsilon_t,$$

where $\varepsilon_t \sim iid\ N(0, \sigma^2)$.

- The most simple example is once again an $AR(1)$,

$$x_t = \mu + \rho_1 x_{t-1} + \varepsilon_t,$$

where $E(x_t) = \mu/(1-\rho_1)$, $Var(x_t) = \sigma^2/(1-\alpha_1^2)$ and $cor(x_t, x_{t-l}) = \rho_1^l$ if the process is stationary.

Stationarity of an $AR(p)$ processes

- We can write an AR(1) as a function of its own past:

$$x_t = \mu(1 + \rho_1 + \rho_1^2 + \cdots) + \varepsilon_t + \rho_1\varepsilon_{t-1} + \rho_1^2\varepsilon_{t-2} + \cdots ,$$

- Hence, x_t will be stationary if

$$\begin{aligned} E(x_t) &= \mu(1 + \rho_1 + \rho_1^2 + \cdots), \\ Var(x_t) &= \sigma^2(1 + \rho_1^2 + \rho_1^4 + \cdots), \end{aligned}$$

  converge to a finite value, which requires that $|\rho_1| < 1$.

- In the general $AR(p)$ case, stationarity requires that the solutions of the equation

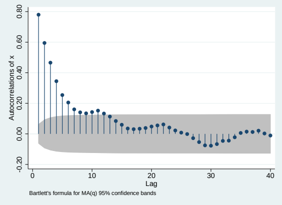
$$1 - \rho_1 z - \rho_2 z^2 - \cdots - \rho_p z^p = 0$$

  are all greater than 1 in absolute value.

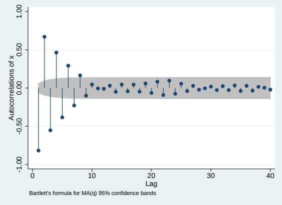
AR identification

Simulation example. $AR(1)$, $T = 1,000$

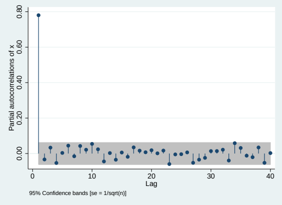
Correlogram, $\alpha_1 = 0.8$.



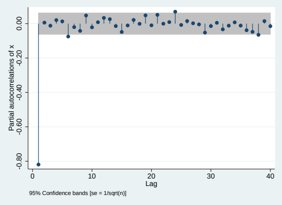
Correlogram, $\alpha_1 = -0.8$.



Partial correlogram, $\alpha_1 = 0.8$.



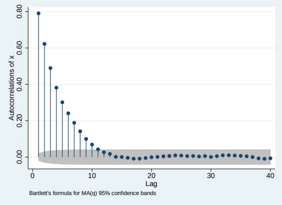
Partial correlogram, $\alpha_1 = -0.8$



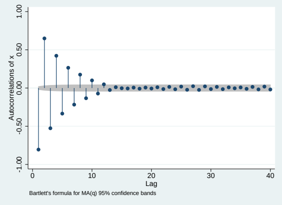
AR identification

Simulation example. $AR(1)$, $T = 10,000$

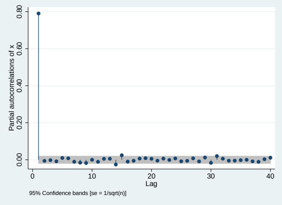
Correlogram, $\alpha_1 = 0.8$.



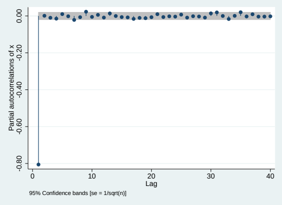
Correlogram, $\alpha_1 = -0.8$.



Partial correlogram, $\alpha_1 = 0.8$.



Partial correlogram, $\alpha_1 = -0.8$



ARMA processes

- Simple AR and MA models may be unsatisfactory, because a high order may be needed to get a good fit.
- The autoregressive moving average model (ARMA) overcomes this problem by combining the two possibilities so that the number of parameters is kept small.
- For instance, an ARMA(1,1) can be written as

$$x_t = \mu + \rho_1 x_{t-1} + \varepsilon_t + \alpha_1 \varepsilon_{t-1}.$$

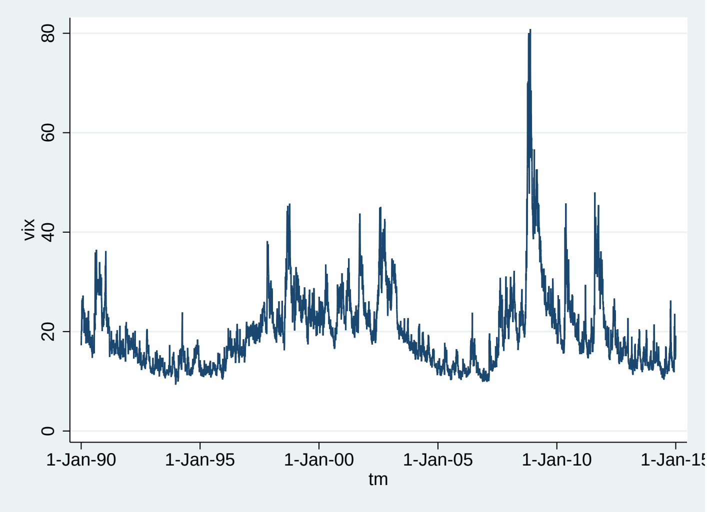
- ARMA models are more useful in volatility models in finance, due to the low correlation of stock returns.

Application. Modelling the VIX index

- VIX is the ticker symbol for the Chicago Board Options Exchange (CBOE) volatility index.
- VIX is computed in real time using a formula that uses prices on short term S&P500 index options.
- Formally, it is the market expectation of the volatility over the next 30 calendar days. Many comentators refer to it as a market fear gauge.
- It is possible to invest in volatility through futures and options on the VIX.
- The purpose of this exercise is to estimate the most appropriate ARMA model for the VIX.
- Modelling the VIX is crucial to price its derivatives, as well as for prediction purposes.

Application. Modelling the VIX index

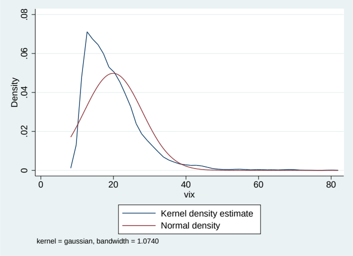
Historical evolution



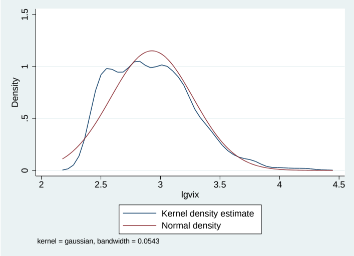
Summary statistics

	VIX	log-VIX
Mean	19.95	2.93
Std. dev.	8.01	0.35
Skewness	2.06	0.63
Kurtosis	10.40	3.33
Min	9.31	2.23
Max	80.86	4.39

VIX kernel density



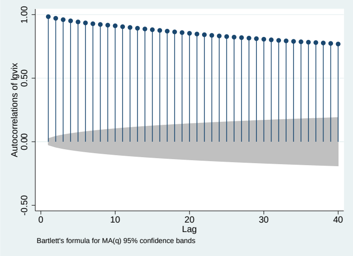
Log-VIX kernel density



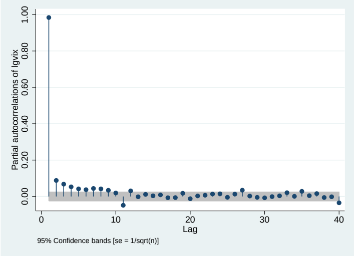
Application. Modelling the VIX index

Correlogram and partial correlogram of the log VIX

Correlogram



Partial correlogram



```
. dfgls lgvix, maxlag(10) notrend;

DF-GLS for lgvix                                    Number of obs =   6287

               DF-GLS mu        1% Critical        5% Critical       10% Critical
  [lags]     Test Statistic        Value              Value              Value
______________________________________________________________________________

    10          -4.631            -2.580             -1.950             -1.626
    9           -4.421            -2.580             -1.950             -1.626
    8           -4.514            -2.580             -1.950             -1.626
    7           -4.678            -2.580             -1.950             -1.626
    6           -4.890            -2.580             -1.950             -1.626
    5           -5.122            -2.580             -1.950             -1.627
    4           -5.329            -2.580             -1.950             -1.627
    3           -5.572            -2.580             -1.950             -1.627
    2           -5.899            -2.580             -1.951             -1.627
    1           -6.325            -2.580             -1.951             -1.627

Opt Lag (Ng-Perron seq t) = 10 with RMSE  .0607205
Min SC   = -5.587941 at lag  7 with RMSE  .0608383
Min MAIC = -5.592611 at lag 10 with RMSE  .0607205
```

- Therefore, the null hypothesis of a unit root is clearly rejected.

- We are going to estimate different ARMA models by maximum likelihood, and compare their fit.
- The likelihood is a reasonable way to compare the different models when the number of parameters is the same.
- We can control by the number of parameters by using the Akaike and Bayesian information criteria:

$$
\begin{aligned}
AIC &= -2\log(\mathit{likelihood}) + 2N_{parameters} \\
BIC &= -2\log(\mathit{likelihood}) + 2N_{parameters}\log(T)
\end{aligned}
$$

- Another way to check the fit is by checking that the residuals are indeed white noise.

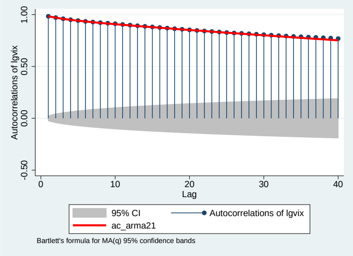
ARMA	log-lik	AIC	BIC	Parameters
(1,0)	8618	-17230	-17210	3
(2,0)	8643	-17278	-17251	4
(3,0)	8657	-17305	-17271	5
(4,0)	8666	-17320	-17280	6
(5,0)	8672	-17329	-17282	7
(6,0)	8676	-17336	-17282	8
(0,1)	1291	-2576	-2556	3
(0,2)	3544	-7080	-7053	4
(0,3)	4889	-9767	-9733	5
(1,1)	8647	-17286	-17259	4
(2,1)	8688	-17365	-17332	5

Application. Modelling the VIX index

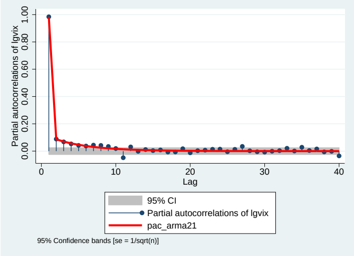
ARMA(2,1) estimates

```
. arima lgvix, ar(1/2) ma(1) technique(nr) nolog;

ARIMA regression

Sample:  1 - 6298                               Number of obs      =      6298
                                                Wald chi2(3)       =  5.90e+06
Log likelihood =  8687.722                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OIM
       lgvix |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lgvix        |
       _cons |    2.92721   .0741996    39.45   0.000     2.781781    3.072638
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   1.699206   .0364924    46.56   0.000     1.627682     1.77073
         L2. |  -.7010359   .0361288   -19.40   0.000    -.7718472   -.6302247
             |
          ma |
         L1. |  -.8186722    .029983   -27.30   0.000    -.8774378   -.7599066
-------------+----------------------------------------------------------------
      /sigma |   .0608916   .0005426   112.23   0.000     .0598283     .061955
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

Correlogram

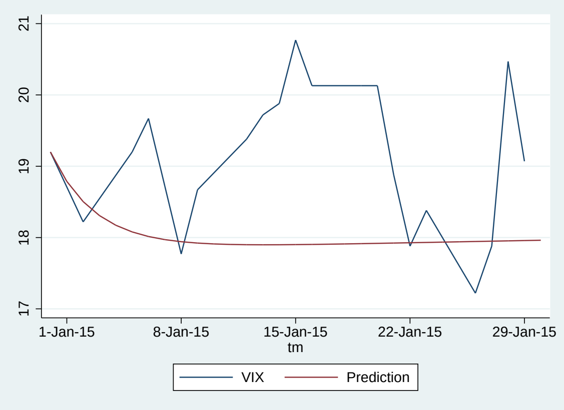


Partial correlogram



Application. Modelling the VIX index

Out-of-sample forecasts based on the ARMA(2,1) model



Outline

1. Introduction
2. Correlation, autocorrelation, stationarity and seasonality
3. Univariate linear models
4. Multivariate linear models
5. Cointegration

Multivariate models

- Since many economic and financial variables are interdependent, it may be necessary to model them jointly for certain applications.
- A multivariate time series $\mathbf{y}_t$ is a vector $(y_{1t}, y_{2t}, \cdots, y_{nt})'$ of n different time series processes that are measured concurrently.
- All the components of $\mathbf{y}_t$ should have the same properties in terms of stationarity, seasonality, ...
- The goal of this section is to develop multivariate extensions of the univariate linear models to model both the time series and cross-sectional dimensions of $\mathbf{y}_t$ jointly.

Cross-covariances

- We can define the covariance matrix of $\mathbf{y}_t$ as

$$\boldsymbol{\Gamma}_0 = E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_t - \boldsymbol{\mu})'] \text{ where } \boldsymbol{\mu} = E[\mathbf{y}_t].$$

- Analogously, we can define the cross-covariance matrix of order l as

$$\boldsymbol{\Gamma}_l = E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t-l} - \boldsymbol{\mu})'] \approx \frac{1}{T} \sum_{t=l+1}^{T} (\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_{t-l} - \bar{\mathbf{y}})',$$

  where $\bar{\mathbf{y}}$ is the sample mean vector.
- Cross-covariances capture the two most relevant features that multivariate linear models generate:
  - Time series univariate dimension (diagonal of $\boldsymbol{\Gamma}_l$): autocorrelation in the components of $\mathbf{y}_t$.
  - Cross-sectional dimension (off-diagonal terms): cross-correlation between different elements of $\mathbf{y}_t$.

- The $n \times 1$ vector ε_t is multivariate Gaussian white noise if $\varepsilon_t \sim \text{iid} N(0, \mathbf{\Sigma})$.
- Hence, $\mathbf{\Gamma}_0 = \mathbf{\Sigma}$ and $\mathbf{\Gamma}_1 = \mathbf{\Gamma}_2 = \cdots = \mathbf{0}$ in this case.
- Contemporaneous correlation can be obtained by linear combinations of independent standard normal variables: $\varepsilon_t = \mathbf{H}\mathbf{v}_t$, where $\mathbf{v}_t \sim \text{iid} N(0, \mathbf{I}_n)$.
- Hence, $\mathbf{H}$ should be such that $\mathbf{\Sigma} = Var(\varepsilon_t) = \mathbf{H} Var(\mathbf{v}_t)\mathbf{H}' = \mathbf{H}\mathbf{H}'$.
- Usually, $\mathbf{H}$ is chosen to be a lower triangular matrix. This is known as the Cholesky factorisation.

- Consider the following bivariate case:

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}$$

- Then, the Cholesky factorisation would be

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0.6 & 0.8 \end{pmatrix}$$

- And the white noise vector ε_t would be obtained as

$$\begin{aligned} \varepsilon_{1t} &= \mathbf{v}_{1t} \\ \varepsilon_{2t} &= 0.6\mathbf{v}_{1t} + 0.8\mathbf{v}_{2t} \end{aligned}$$

Vector autorregressive models: $VAR(p)$

- A Gaussian vector autorregressive process of order p, or $VAR(p)$ can be expressed as

$$\mathbf{y}_t = \mathbf{a} + \mathbf{B}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{B}_p \mathbf{y}_{t-p} + \varepsilon_t,$$

  where $\mathbf{a}$ is an $n \times 1$ vector, $\mathbf{B}_1, \cdots, \mathbf{B}_p$ are $n \times n$ matrices and ε_t is multivariate white noise: $\varepsilon_t \sim \text{iid} N(0, \mathbf{\Sigma})$.
- For a bivariate $VAR(1)$ case, we would have:

$$\begin{aligned} y_{1t} &= a_1 + b_{11} y_{1t-1} + b_{12} y_{2t-1} + \varepsilon_{1t}, \\ y_{2t} &= a_2 + b_{21} y_{1t-1} + b_{22} y_{2t-1} + \varepsilon_{2t}. \end{aligned}$$

- This is known as the reduced form representation, because it does not explicitly show the dependence through the contemporaneous correlation between ε_{1t} and ε_{2t}.

Impulse-response function in $VAR(p)$ models

- Following with the bivariate example, we can introduce the Cholesky decomposition in the reduced form equation:

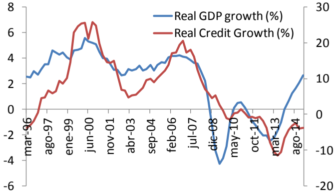
$$\begin{aligned} y_{1t} &= a_1 + b_{11} y_{1t-1} + b_{12} y_{2t-1} + h_{11} v_{1t}, \\ y_{2t} &= a_2 + b_{21} y_{1t-1} + b_{22} y_{2t-1} + h_{21} v_{1t} + h_{22} v_{2t}. \end{aligned}$$

- The impulse response function analyses the impact of adding an artificial shock δ at time τ, $\mathbf{v}_\tau = (v_{1\tau} + \delta, v_{2\tau})'$.
- The impulse response function shows the difference between the shocked and the original variables at times $t = \tau, \tau + 1, \tau + 2, \ldots$.

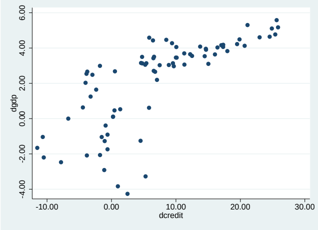
$$\begin{aligned} \Delta y_{1,\tau} = h_{11}\delta, \Delta y_{1,\tau+1} = b_{11} h_{11}\delta, \Delta y_{1,\tau+2} = b_{11}^2 h_{11} + b_{12} h_{11}\delta, \cdots \\ \Delta y_{2,\tau} = h_{21}\delta, \Delta y_{2,\tau+1} = b_{21} h_{21}\delta, \Delta y_{2,\tau+2} = b_{21}^2 h_{21} + b_{22} h_{21}\delta, \cdots \end{aligned}$$

- Note that the result depends on the ordering of the variables.

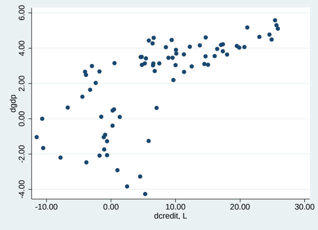
Annual growth levels (%)



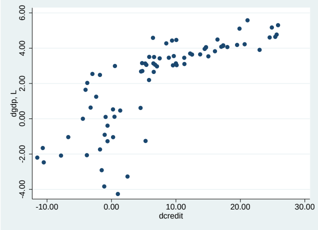
GDP and credit



GDP and lagged credit



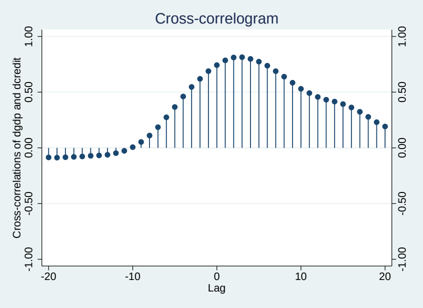
Lagged GDP and credit



```
. dfgls dgdp, maxlag(2) notrend;

DF-GLS for dgdp                                         Number of obs =    74

               DF-GLS mu      1% Critical        5% Critical      10% Critical
  [lags]     Test Statistic      Value              Value             Value
------------------------------------------------------------------------------

    2          -2.445          -2.609             -2.145            -1.842
    1          -2.349          -2.609             -2.158            -1.854

. dfgls dcredit, maxlag(2) notrend;

DF-GLS for dcredit                                      Number of obs =    74

               DF-GLS mu      1% Critical        5% Critical      10% Critical
  [lags]     Test Statistic      Value              Value             Value
------------------------------------------------------------------------------

    2          -1.409          -2.609             -2.145            -1.842
    1          -1.306          -2.609             -2.158            -1.854
```

- The unit root is rejected for GDP annual growth at the 5% and 10% levels, acceptable for just 74 observations.
- For consistency, we will use the same approach for the two variables.

$$Corr(\Delta GDP_t, \Delta Credit_{t-l})$$



	Log-likelihood	AIC	BIC
VAR(1)	-230.554	6.225	6.410
VAR(2)	-201.367	5.636	5.945
VAR(3)	-196.924	5.701	6.136

- Both the AIC and the BIC criteria support estimating a VAR(2), which represents a good compromise between flexibility and parsimony.

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dgdp						
dgdp						
L1.	1.597583	.0903477	17.68	0.000	1.420505	1.774662
L2.	-.6592625	.0956093	-6.90	0.000	-.8466533	-.4718717
dcredit						
L1.	.0290494	.0251178	1.16	0.247	-.0201806	.0782793
L2.	-.0265347	.0235995	-1.12	0.261	-.0727888	.0197195
_cons	.1191753	.0693111	1.72	0.086	-.016672	.2550225
dcredit						
dgdp						
L1.	.3302322	.3980577	0.83	0.407	-.4499466	1.110411
L2.	.137158	.4212394	0.33	0.745	-.6884561	.9627722
dcredit						
L1.	1.227604	.110665	11.09	0.000	1.010704	1.444503
L2.	-.3671478	.1039756	-3.53	0.000	-.5709361	-.1633594
_cons	.0239301	.3053736	0.08	0.938	-.5745913	.6224514

- VAR(1)

$$\mathbf{y}_t = \mathbf{a} + \mathbf{B}_1 \mathbf{y}_{t-1} + \varepsilon_t,$$

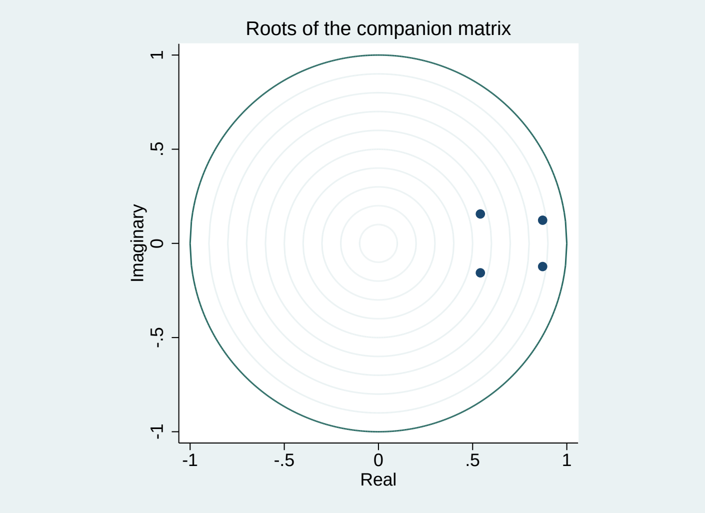
  Stationarity requires the eigenvalues of $\mathbf{B}_1$ to be smaller than 1 in modulus.

- VAR(2)

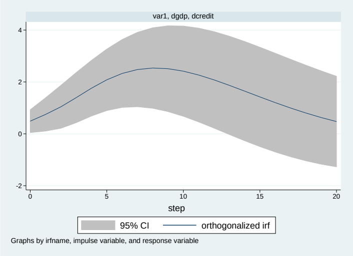
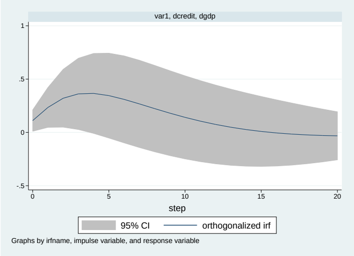
$$\mathbf{y}_t = \mathbf{a} + \mathbf{B}_1 \mathbf{y}_{t-1} + \mathbf{B}_2 \mathbf{y}_{t-2} + \varepsilon_t,$$

  The auxiliar matrix $\mathbf{B}_1^*$ should have eigenvalues smaller than 1 in modulus:

$$\mathbf{B}_1^* = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{I}_2 & \mathbf{0} \end{bmatrix}.$$



Roots of the companion matrix
Imaginary
1
.5
0
-.5
-1
-1
-.5
0
.5
1
Real

OIRF of Credit to a GDP positive shock	OIRF of GDP to a credit positive shock
	

VARMA models

- It is also possible to generalise the MA process to a vector moving average (VMA) multivariate process.
- Similarly, ARMA models can be generalised (VARMA).

$$\mathbf{y}_t = \mathbf{a} + \mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} + \varepsilon_t + \mathbf{C}_1\varepsilon_{t-1} + \cdots + \mathbf{C}_q\varepsilon_{t-q},$$

- The estimation of these processes is much more involved than the estimation of VAR processes.
- To the best of my knowledge, these models are not supported by Stata.

Outline

Cointegration

- As we have seen, many economic series are integrated processes of order 1, or $I(1)$.
- In addition, sometimes the spread between two (or more) I(1) series is stationary. Then, we say that these series are cointegrated.
- Cointegration is a measure of long term dependence between economic variables, by which these variables are "tied together" in the long run.
- Thus, cointegration is very different from correlation or even copulas, which tie returns (and returns are short term movements by definition).

- Cointegration between two variables x_t and y_t requires two conditions:
  1. Both x_t and y_t are $I(1)$ (i.e.: they have a unit root),...
  2. but there is a linear combination $z_t = x_t - \alpha y_t$ that is stationary (the cointegration relationship).
- The vector $(1, -\alpha)$ is called the cointegrating vector. It defines the long term equilibrium, since x_t cannot move very far from αy_t in the long run.
- Intuitively, x_t and y_t can be seen as two variables with a common stochastic (and non-stationary) trend:

$$
\begin{aligned}
x_t &= \alpha\delta_t + \varepsilon_{1t} \\
y_t &= \delta_t + \varepsilon_{2t}
\end{aligned}
$$

so that $z_t = \varepsilon_{1t} - \alpha\varepsilon_{2t}$ is stationary.

- Although x_t and y_t are tied together in the long run, unless the variance of $z_t = x_t - \alpha y_t$ is zero these two variables may experience transitory departures from this equilibrium in the short run.
- Cointegration theory establishes that the dynamics of cointegrated variables cannot be correctly described with a vector autoregressive model.
- The VAR does not exploit all the information, because it does not take into account the degree of disequilibrium between the variables given by the cointegration relationship.
- The correct specification is obtained by adding lag(s) of the disequilibrium terms in the VAR equations.
- The resulting formula is called the error correction model.

- Assuming that only the first lags matter, the error correction model can be written as

$$
\begin{aligned}
\Delta x_t &= \gamma_{10} + \gamma_{11}\Delta x_{t-1} + \gamma_{12}\Delta y_{t-1} + \gamma_{13} z_{t-1} + \varepsilon_{1t}, \\
\Delta y_t &= \gamma_{20} + \gamma_{21}\Delta x_{t-1} + \gamma_{22}\Delta y_{t-1} + \gamma_{23} z_{t-1} + \varepsilon_{2t},
\end{aligned}
$$

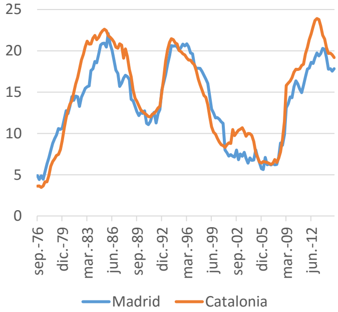
  where Δ denotes first differences and ε_{1t} and ε_{2t} are *iid* but potentially correlated variables.
- "Error correction" means that short term deviations from the long term equilibrium will eventually be corrected.
- If we assume that $\alpha > 0$, then there will only be and error correction mechanism if $\gamma_{13} < 0$ and $\gamma_{23} > 0$.

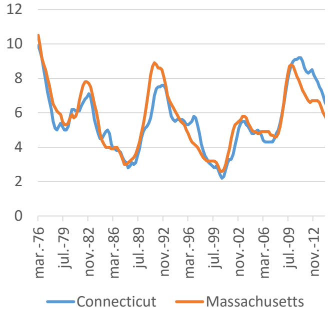
Identification of cointegration relationships

- Johansen's test investigate cointegration in general multivariate systems of n variables.
- For our bivariate example, this test seeks the linear combination of x_t and y_t that is more stationary.

Example

- Unemployment rates between regions (ES) or states (US) should not diverge indefinitely if restrictions to migration are small.
- This is specially the case between neighbouring regions.





Unint root tests

Spanish example

DF-GLS for Madrid — Number of obs = 15

[lags]	DF-GLS mu Test Statistic	1% Critical Value	5% Critical Value	10% Critica Value
3	-0.869	-2.593	-2.043	-1.732
2	-0.721	-2.593	-2.049	-1.738
1	-0.424	-2.593	-2.055	-1.743

DF-GLS for Catalonia — Number of obs = 15

[lags]	DF-GLS mu Test Statistic	1% Critical Value	5% Critical Value	10% Critica Value
3	-1.263	-2.593	-2.043	-1.732
2	-1.059	-2.593	-2.049	-1.738
1	-0.737	-2.593	-2.055	-1.743

- The two series have a unit root, but Johansen's test does not find cointegration.
- Perhaps this relationship is more of a highly correlated VAR model for the first differences than a long term cointegration relatioship.
- However, this test is very sensitive to changes in the assumptions (e.g. tend).

```
. vecrank Madrid Catalonia, lags(3) trend(rconstant);

                        Johansen tests for cointegration
Trend: rconstant                                        Number of obs =     153
Sample:  1977q2 - 2015q2                                         Lags =       3
-------------------------------------------------------------------------------
                                                             5%
maximum                                        trace    critical
  rank     parms       LL         eigenvalue  statistic    value
    0        8      -308.33058          .      15.5273*   19.96
    1       12      -303.83906     0.05702      6.5443     9.42
    2       14      -300.56692     0.04187
-------------------------------------------------------------------------------
```

- However, in the US case the two series have a unit root and Johansen's test does find cointegration.

```
. vecrank Connecticut Massachusetts, lags(3) trend(rconstant);

                       Johansen tests for cointegration
Trend: rconstant                                         Number of obs =     156
Sample:  1976q4 - 2015q3                                          Lags =       3
-------------------------------------------------------------------------------
                                                       5%
maximum                                      trace    critical
  rank    parms       LL       eigenvalue  statistic   value
    0       8       153.897         .       22.6775    19.96
    1      12     161.26448     0.09013      7.9426*    9.42
    2      14     165.23577     0.04964
-------------------------------------------------------------------------------
```

Cointegration relationship
US example

```
. vec Connecticut Massachusetts, lags(3) trend(rconstant);
```

Vector error-correction model

beta	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_ce1						
Connecticut	1	.	.	.	.	.
Massachusetts	-1.693616	.2348758	-7.21	0.000	-2.153964	-1.233268
_cons	3.974895	1.370262	2.90	0.004	1.289231	6.660559