Time Series Analysis: Univariate Stationary Processes

Master

Financial Economics

ANNÉE SCOLAIRE / ACADEMIC YEAR 2011-2012

Intervenant/Lecturer: Sergio FOCARDI

Dynamic Models

- Dynamic models represent processes through a Data Generation Process – DGP
- Which is a mechanism that produces future values given past values
- We will discuss linear dynamic models
- Auto Regressive Moving Average - ARMA – models for univariate series
- Vector Auto Regressive Moving Average – VARMA – models for multivariate series
- State-space models
- Equivalence between VARMA and state-space models

Bird's eye view

Type	Static model	Static model	Dynamic model	Dynamic model	Dynamic model
Model	Regressions	Factor models	ARCH and GARCH models	Dynamic factor models	ARMA VARMA VAR State Space models
Use	Prediction Exposure to risk factors	Dimensionality reduction Explanation Risk factors	Volatility modelling	Predictions	Predictions
Estimation	OLS and MLE	MLE and PCA	MLE	MLE	MLE

Economic intuition

- Financial markets are almost efficients but
- There are feedbacks. Why?:
- The economy is finite, prices cannot diverge indefinitely
- Agents act as servo-control mechanisms controlling risk and return
- People are subject to subtle patterns of reaction
- Somehow financial markets are managed

Empirical evidence of feedback

- Autocorrelation
- Momentum
- Reversals

Autocorrelation

- Consider a moment *t* and returns at *t*

$$R(t) = \frac{P(t) - P(t-1)}{P(t-1)}$$

- Form the correlation coefficient between returns at different moments *t,t+h*

$$\mu_R = E(R(t)) = \text{constant}$$
$$corr(R(t), R(t+h)) = E[(R(t - \mu_R)) \times (R(t+h) - \mu_R)]$$

- If returns are stationary means and correlations at different moments, called autocorrelations, are idependent from time

Empirical autocorrelation

- We cannot estimate autocorrelations as we have only one series, e.g., one return process
- Assuming returns are ergodic (means and time averages can be exchanged) we compute empirical autocorrelations for each series

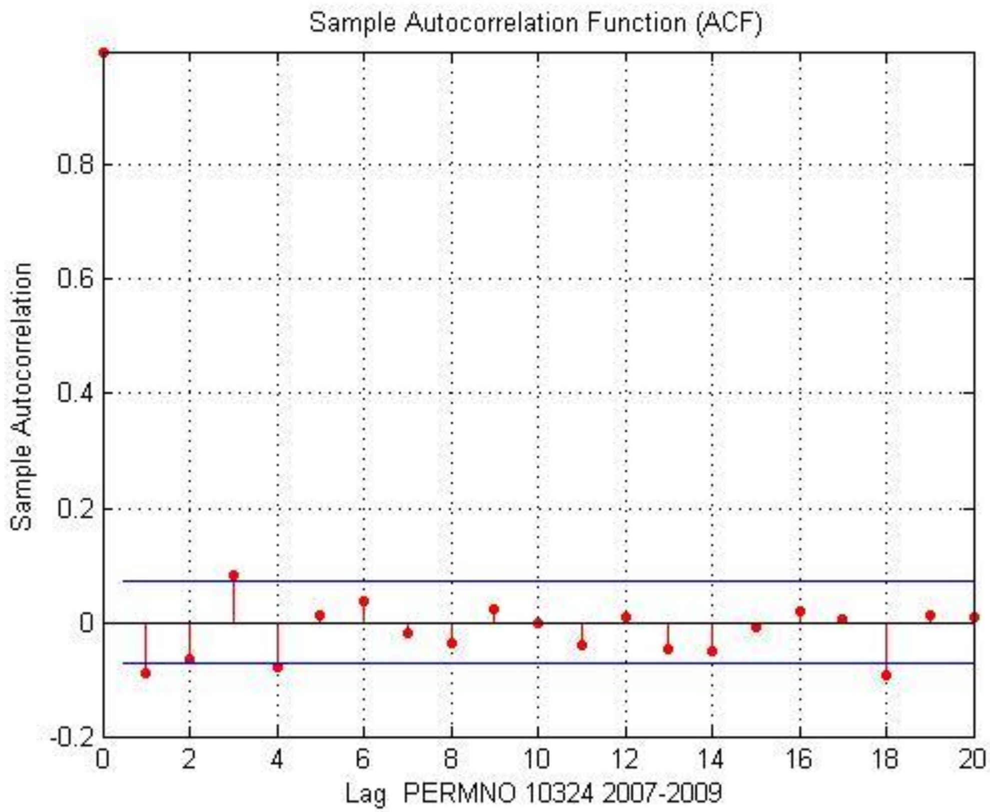
$$\hat{\mu}_R = \frac{\sum_{t=1}^{T} R(t)}{T}$$

$$\rho(h) = \frac{\sum_{t=1}^{T-h} \left(R(t) - \hat{\mu}_R\right) \times \left(R(t+h) - \hat{\mu}_R\right)}{T}$$

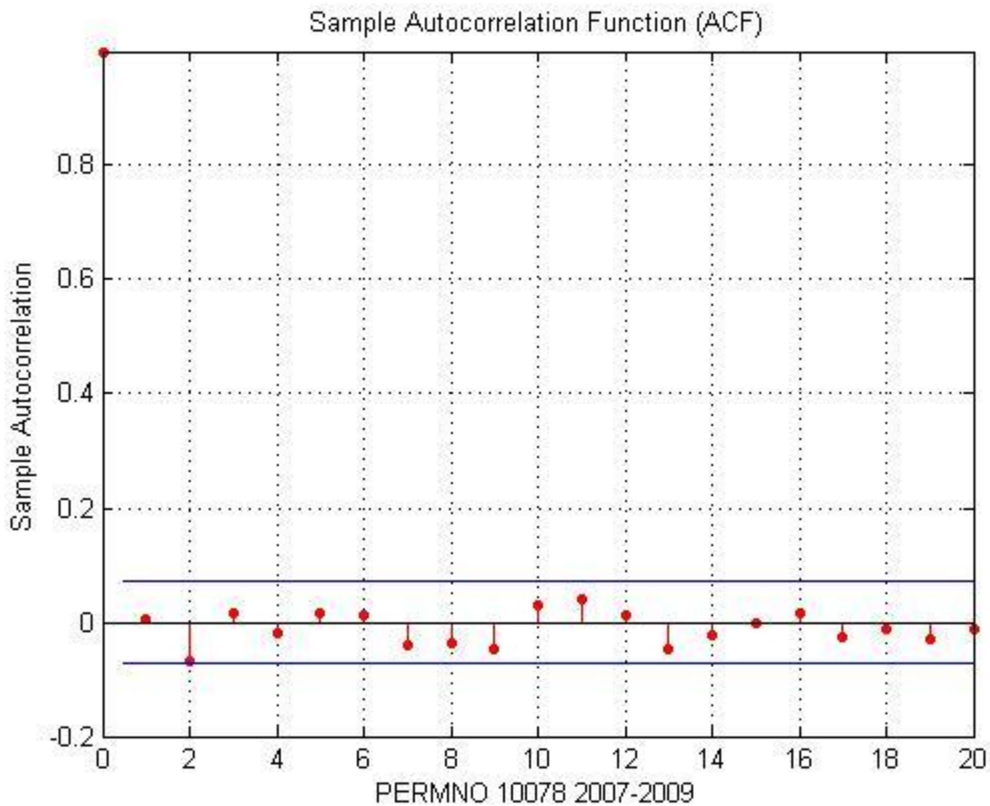
Autocorrelation in practice

- Very little autocorrelation of individual stocks
- Autocorrelations of portfolios and indexes
- Due to auto cross correlations

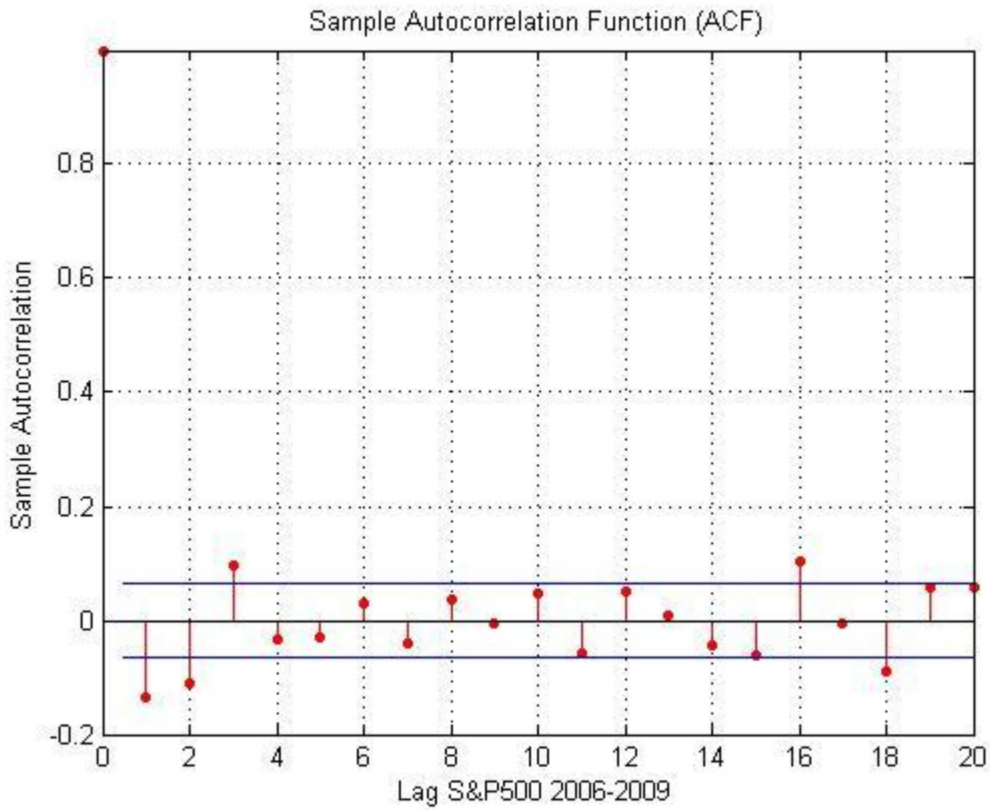
Autocorrelation function returns



Autocorrelation function returns



Autocorrelation SP500 VW 2006-2009

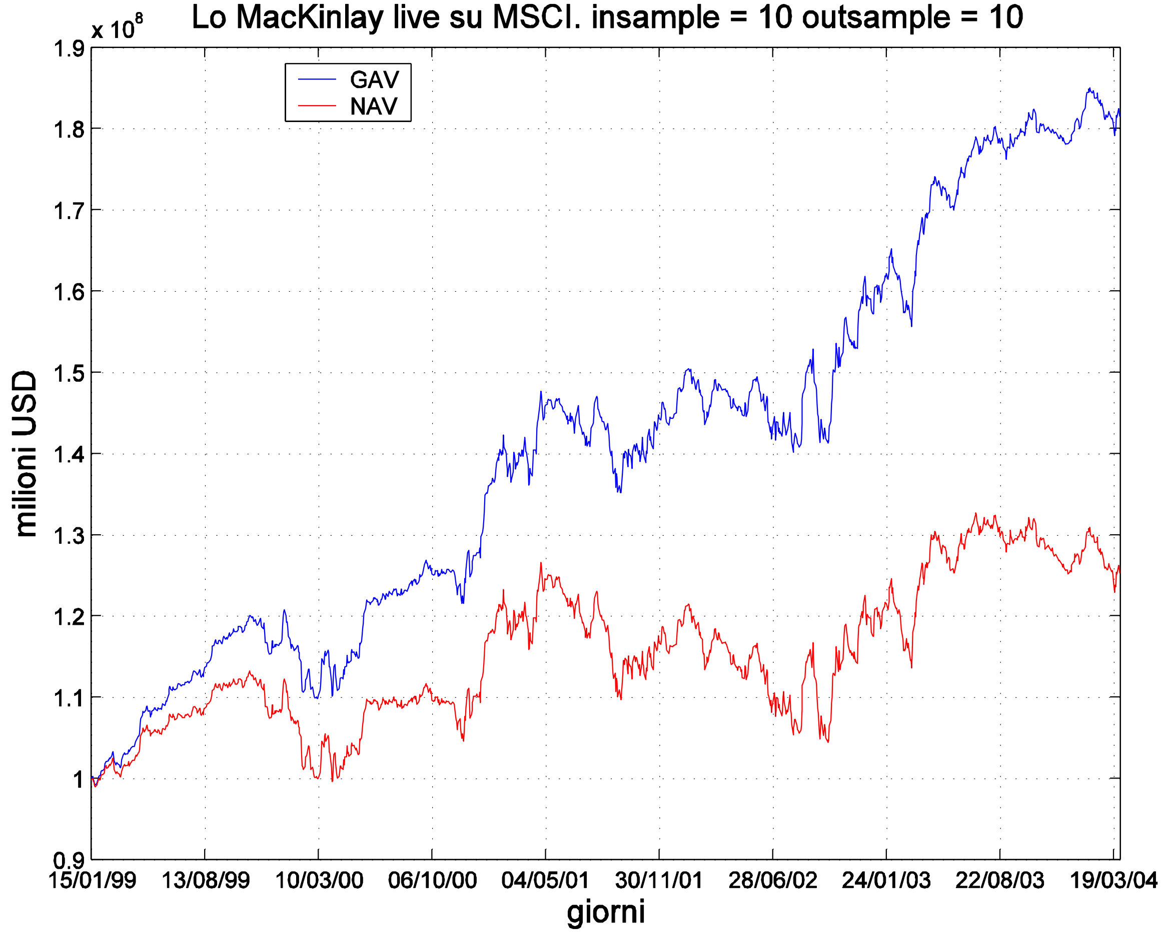


Reversal

- Well documented in the literature
- "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test", Andrew W. Lo and A. Craig MacKinlay
- Portfolio weights proportional to previous period excess returns:

$$w_{i,t} = \frac{1}{N}\left(r_{i,t} - r_{m,t}\right)$$

Profit of Lo-MacKinlay strategy



Empirical findings of momentum and reversals

Definition of momentum

- The phenomenon called momentum and reversals in the literature is…
- The existence of a well defined function that links the magnitude of realized returns to each pair consisting of an estimation window and a holding period

Two descriptive frameworks

- Jegadeesh-Titman (J-T) quantiles portfolios
- Lo-MacKinlay (L-M) global hedge portfolio
- L-M is analytically very useful, J-T closer to real-world implementations

Jegadeesh-Titman framework

- Jegadeesh & Titman first reported momentum in 1993
- Consider a universe of stocks
- Consider monthly returns
- At the beginning of each month consider returns on the past *J* months
- Order returns in ascending order and form ten equally weighted decile portfolios

Jegadeesh-Titman framework, ctd…

- The *J*-month period is called the formation period
- Select the highest and lowest decile portfolios
- Hold stocks for *K* months
- The *K*-months period is called the holding period
- Evaluate returns in the holding period

Jegadeesh-Titman framework, ctd…

- Consider a vector of prices sampled at given intervals: $P_{i,t}$
- For each stock, consider returns in the formation and holding periods

$$R_{J,i,t} = \frac{P_{i,t} - P_{i,t-J}}{P_{i,t-J}}$$

$$R_{K,i,t} = \frac{P_{i,t+K} - P_{i,t}}{P_{i,t}}$$

Jegadeesh-Titman framework, ctd...

- Order returns in ascending order, select the top and bottom deciles and form three portfolios:
  - An equally weighted portfolio L with stocks in the top decile
  - An equally weighted portfolio S with stocks in the bottom decile
  - A zero-cost, hedge portfolio long in L and short in S

Jegadeesh-Titman framework, ctd…

Call $R_{T,t}, R_{B,t}, R_{LS,t}$ the returns of these portfolios for each pair *J,K*:

$$R_{T,t}\left(J,K\right)$$

$$R_{B,t}\left(J,K\right)$$

$$R_{LS,t}\left(J,K\right) = R_{T,t}\left(J,K\right) - R_{B,t}\left(J,K\right)$$

Jegadeesh-Titman framework, ctd…

- Given that portfolios are equally weighted, are the average returns in the top and bottom deciles $R_{T,t}(J,K),\ R_{B,t}(J,K)$
- Call $R_{M,t}(K)$ the average return of all stocks in the holding period

Formulations of momentum and reversal phenomena

- Weak formulation: the sign of $E_t\left(R_{LS,t}\left(J,K\right)\right)$ is a well defined function of *J,K* and does not depend on time

- Stronger formulation: $E_t\left(R_{LS,t}\left(J,K\right)\right)$ is a well defined function of *J,K* and does not depend on *t*

Formulations of momentum and reversal phenomena, ctd…

Intermediate formulation:

- The conditional expectation $E_t\left(R_{LS,t}\left(J,K\right)\right)$ is a well defined function of (*J,K*) that changes slowly with time
- In other words: there is a term structure of momentum/reversal profits/losses that changes slowly with time

Extensions of the J-T framework

- Adding a short lag (typically one week) between the formation and holding periods
- Considering different percentiles
- Other, such as considering the entire term structure of momentum returns examined later

The Lo-MacKinlay framework

- Proposed by Lo and MacKinlay in 1990 to study the sources of momentum
- Consider a hedge portfolio formed by all stocks
- The weight of each stock is determined as follows

The Lo-MacKinlay framework, ctd...

- Consider the i-th stock. Its weight is:

$$w_{i,t+K} = \frac{1}{N}\left(R_{J,i,t} - R_{J,m,t}\right)$$

- where:

$$R_{J,m,t} = \frac{1}{N}\left(\sum_i R_{J,i,t}\right)$$

- is the average market return at time *t*

The Lo-MacKinlay framework, ctd…

- The Lo-MacKinlay framework allows to derive a number of results analytically, as we will see in the next module
- The J-T framework can be analyzed only empirically and via simulations

Bird's eye view of momentum & reversal

Momentum/reversals as described in the literature consist in the simultaneous presence of the following three phenomena:

- Short-time (*J*,*K*: 1 month) reversal
- Medium-time (*J*,*K*: 3-12 months) momentum
- Long-horizon (*J*,*K*: 2-5 years) reversals

Bird's eye view of momentum & reversal, ctd…

- Momentum alone is only a partial view of the phenomena
- It is the simultaneous presence of the three reversals-momentum-reversals phenomena that characterize price behavior…

Bird's eye view of momentum & reversal, ctd…

- And makes it difficult to find a tenable econometric explanation that also exhibits:
  - Low autocorrelations of returns
  - Substantial correlation of returns at the same time
  - Plus other stylized facts such as ARCH

Empirical results

- Let’s now describe in detail empirical results
- In this section we present raw empirical results of portfolio strategies
- As described in the literature and based on our own tests
- Without econometric explanation

Early findings

- The reversal phenomena were first reported in the financial literature by DeBondt and Thaler (1985)
- Long-short contrarian portfolios earn returns of about 8% per year when both the formation and the holding periods are 36 months
- Contrarian profits for each formation and holding period between 2 and 5 years
- Jegadeesh (1990) reports contrarian profits (reversals) of about 1.94% per month for short formation and holding periods of one month
- Lehmann (1990) reports contrarian profits for formation and holding periods of one week

Empirical results of J-T

- J-T 1993 examines momentum in USA stocks in the period 1965-1989
- They consider J:3-12, K: 3-12
- Findings are summarized in the following table

J-T table

	A				B			
J/K	3	6	9	12	3	6	9	12
3	0.0032	0.0058	0.0061	0.0069	0.0073	0.0078	0.0074	0.0077
6	0.0084	0.0095	0.0102	0.0086	0.0114	0.0110	0.0108	0.0090
9	0.0109	0.0121	0.0105	0.0082	0.0135	0.0130	0.0109	0.0085
12	0.0131	0.0114	0.0093	0.0068	0.0149	0.0121	0.0096	0.0069

Main points

- Maximum return of the L-S portfolio attained for *J*=12 and *K*=3
- Returns of the LS portfolio are higher when there is a one week lag: 0.0131 vs 0.0149
- When *J*=3, returns attain max for *K*=12
- When *J*=12, returns attain max for *K*=3
- For *J*=6 max for *K*=9, for *J*=9 max for *K*=6
- If a one week lag allowed, max always for *K*=3

Main points, ctd…

- There is a strong January effect:
- Every year in January momentum is negative
- Maximum returns are typically achieved for K>1

Long-term reversals

- J-T do not systematically explore long horizons
- However, they find reversals over long horizons, 60 months

Test update

- The test was repeated in the period 1990-1997
- Results are very similar to previous results

Lewellen 2000

- Lewellen repeated the momentum test using portfolios and the Lo-MacKinlay framework
- But considers portfolios and not individual stocks…
- Because earlier research by Grimblatt showed that momentum is almost as strong for industries as for stocks

Lewellen 2000, ctd…

- Lewellen created 15 portfolios:
  - A industry portfolios
  - B size portfolios
- Results shown in the following table

Lewellen table: industry portfolios

J/K	1	3	5	7	9	11	13	15	17
6	0.509	0.263	0.382	0.501	0.351	0.085	-0.078	-0.140	-0.223
12	0.673	0.432	0.309	0.257	0.124	-0.050	-0.128	-0.212	-0.174

Lewellen tables: size portfolios

J/K	1	3	5	7	9	11	13	15	17
6	0.268	0.173	0.283	0.523	0.402	0.261	0.039	0.044	0.055
12	0.525	0.417	0.429	0.416	0.299	0.218	0.232	0.287	0.243

Main points

- Momentum exists for portfolios…
- though they are smaller than in the J-T case
- Might be due to the framework
- Found reversals after 11 months

Karolyi-Kho test

- Karolyi-Kho performed bootstrap tests of several strategies
- They first tested momentum
- 9,807 stocks available from CRSP from NYSE, Amex, and Nasdaq exchanges during the period Jan 1963 to Dec 2000
- Results very similar to J-T
- Results in the following tables

K-K test tables

J/K	3 1965/ 1989	6	9	12	3 1965/ 2000	6	9	12
3	0.67	0.83	0.88	0.89	0.78	0.99	1.03	0.99
6	1.09	1.25	1.28	1.09	1.27	1.46	1.43	1.16
9	1.32	1.46	1.30	1.07	1.46	1.58	1.35	1.05
12	1.44	1.35	1.17	0.92	1.46	1.38	1.13	0.85

Momentum reversals non linear

- Momentum and reversals are non linear phenomena
- The cross sectional distribution of momentum and reversals are non linear
- We consider only linear phenomena
- Hence autocorrelation

Linear models

- We model autocorrelation through linear autoregressive models
- Assuming return distributions are not too different from normal distributions

Univariate autoregressive models

- Models
- Testing
- Estimation

The lag operator L

The lag operator shifts an infinite time series one place to the left

$$Lx(t) = x(t-1)$$

$$L^2 x(t) = L(Lx(t)) = Lx(t-1) = x(t-2)$$

Univariate stationary AR models

- An autoregressive process of order *p* – AR(*p*), is a process of the following form:
- $$x_t + a_1 x_{t-1} + \ldots + a_P x_{t-P} = \varepsilon_t$$
- Which can be written as:
- $$A(L)x_t = \left(1 + a_1 L + \ldots + a_P L^p\right)x_t =$$

$$= x_t + a_1 L x_t + \ldots + a_P L^p x_t = \varepsilon_t$$

Stationarity

- Let's consider the polynomial:

$$A(z) = 1 + a_1 z + \ldots + a_P z^p$$

- Where *z* is a complex variable. This equation is called the inverse characteristic equation:

$$A(z) = 1 + a_1 z + \ldots + a_P z^p = 0$$

- If the roots of this equation are strictly >1 in modulus, then the process is stationary and invertible with an inverse causal representation

Causal representation

If the autoregressive operator $A(L)$ is invertible, then it admits an inverse infinite moving average causal representation:

$$x_t = A^{-1}(L)\varepsilon_t = \sum_{i=0}^{+\infty} \lambda_i \varepsilon_{t-i}$$

$$\text{with } \sum_{i=0}^{+\infty} |\lambda_i| < +\infty$$

Univariate Stationary Moving Average Process

An Univariate Stationary Moving Average Process of order *q* MA(*q*) – admits the following representation:

$$x_t = \left(1 + b_1 L + \ldots + b_P L^q\right)\varepsilon_t =$$

$$= x_t + b_1 \varepsilon_t + \ldots + b_P \varepsilon_{t-q}$$

Causal Inverse Representation

If the roots of $B(z)$ are strictly greater than 1 in modulus, then the MA(q) process is invertible and admits the infinite causal autoregressive representation:

$$\varepsilon_t = B^{-1}(L)\varepsilon_t = \sum_{i=0}^{+\infty} \pi_i \varepsilon_{t-i}$$

$$\text{with } \sum_{i=0}^{+\infty} |\pi_i| < +\infty$$

Auto Regressive Moving Average processes of order *p,q*

- A stationary univariate process admits a minimal Auto Regressive Moving Average representation of order *p,q* ARMA(*p,q*) if it can be written as:
- $$x_t + a_1 x_{t-1} + a_p x_{t-p} = b_1 \varepsilon_t + \ldots + b_q \varepsilon_{t-q}$$
- or
- $$A(L) x_t = B(L) \varepsilon_t$$
- where ε_t is a serially uncorrelated, zero-mean white noise and the *A* and *B* polynomials have roots strictly >1 and no root in common

Representation

- If all the roots of the polynomial *A* are strictly greater than 1 in modulus, then the ARMA(*p,q*) process can be expressed as a moving average process:

$$x_t = \frac{B(L)}{A(L)} \varepsilon_t$$

- Conversely, if all the roots of the polynomial *B* are strictly greater than 1, then the ARMA(*p,q*) process can be expressed as an autoregressive process:

$$\varepsilon_t = \frac{A(L)}{B(L)} x_t$$

$t \geq 0$

Non-stationary univariate ARMA processes

- A process x_t with $t \geq 0$ is called an Integrated Auto Regressive Moving Average process – ARIMA(*p,d,q*) – if it satisfies a relationship of the type:

$$A(L)(I - \lambda L)^d x_t = B(L)\varepsilon_t$$

- where:
- The polynomials $A(L)$ and $B(L)$ have roots strictly>1 in modulus;
- ε_t is a white noise defined for *t*≥0;
- Initial conditions $\left(x_{-1},\ldots,x_{-p-d},\varepsilon_t,\ldots,\varepsilon_{-q}\right)$ independent from the white noise are given

Uniqueness

- The ARMA representation is not unique due to common polynomial factors
- The same process can be represented by models with different pairs
- One needs to determine a minimal ARMA(p,q) representation such that any other ARMA(p',q') have $p'>p, q'>q$
- Mathematically difficult except in the univariate case

Stationarity

- If the complex roots of the algebraic equation
  $\det(\mathbf{A}(z)) = 0$
- lay outside the unit circle, then the ARMA process:

$$\mathbf{A}(L)\mathbf{x}_t = \mathbf{B}(L)\boldsymbol{\varepsilon}_t$$

- is stationary
- Any stationary process admits an infinite moving average representation

Model selection criteria

- There are two basic approaches for assessing the appropriateness of ARMA models
- The first, the Box and Jenkins approach involves inspecting the computed sample autocorrelation functions (SACFs) and sample partial autocorrelation functions (SPACFs) of the time series
- The second approach is to select a set of possible (p, q) combinations and estimate the parameters accordingly.
- Then, the model is chosen with the Akaike information criterion (AIC) or the BIC criterion.
- G.E.P Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control,Revised Edition* (San Francisco: Holden-Day, 1976).

The Box–Jenkins method

- A popular recursive solution that consists of *identification*, *estimation* and *diagnostic checking*
- *Step 1*. The *identification step* in the Box–Jenkins approach tentatively determines the order of differencing, d, to induce stationarity, the autoregressive order, p and the moving average order, q
- The identification step is to "guess" the degree of integration. Given d, the ARMA – orders p and q for the differenced time series $\Delta^d y_t$ can be identified

The Box–Jenkins method

- *Step 2*. Given values for *d*, *p* and *q* from the identification step, the parameters of the ARIMA(*p, d, q*) model are estimated in the *estimation step*
- It delivers estimates of the ARMA coefficients *c*, $a_1, a_2, \dots, a_p, b_1, b_2, \dots, b_q$, and the residual variance σ^2.
- *Step 3*. In the *diagnostic–checking step* we examine the adequacy of the estimated model to determine whether or not all relevant information has been "extracted" from the data

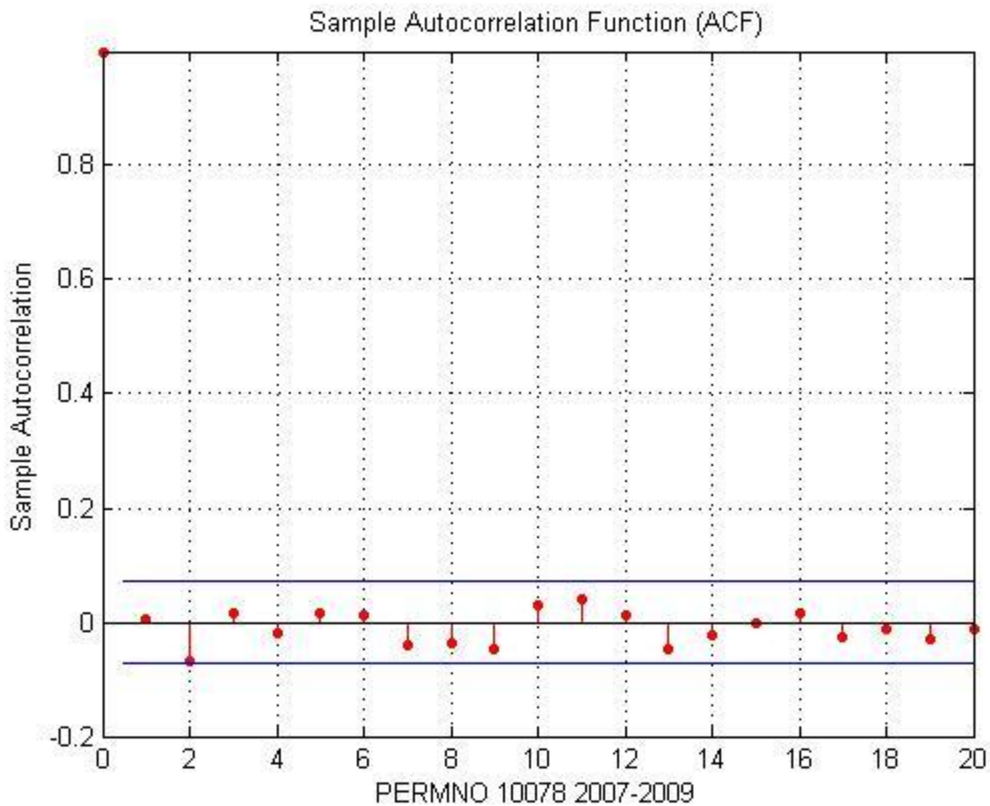
Tools and procedures

- Tools or procedures correspond to each of the described steps
- Inspection of ACF and unit root tests
- Estimation methods such as least squares estimation (LSE) and, maximum likelihood estimation (MLE).
- Analysis of residuals such as - Portmanteau statistics
- ABOVE ALL: Out of sample model validation

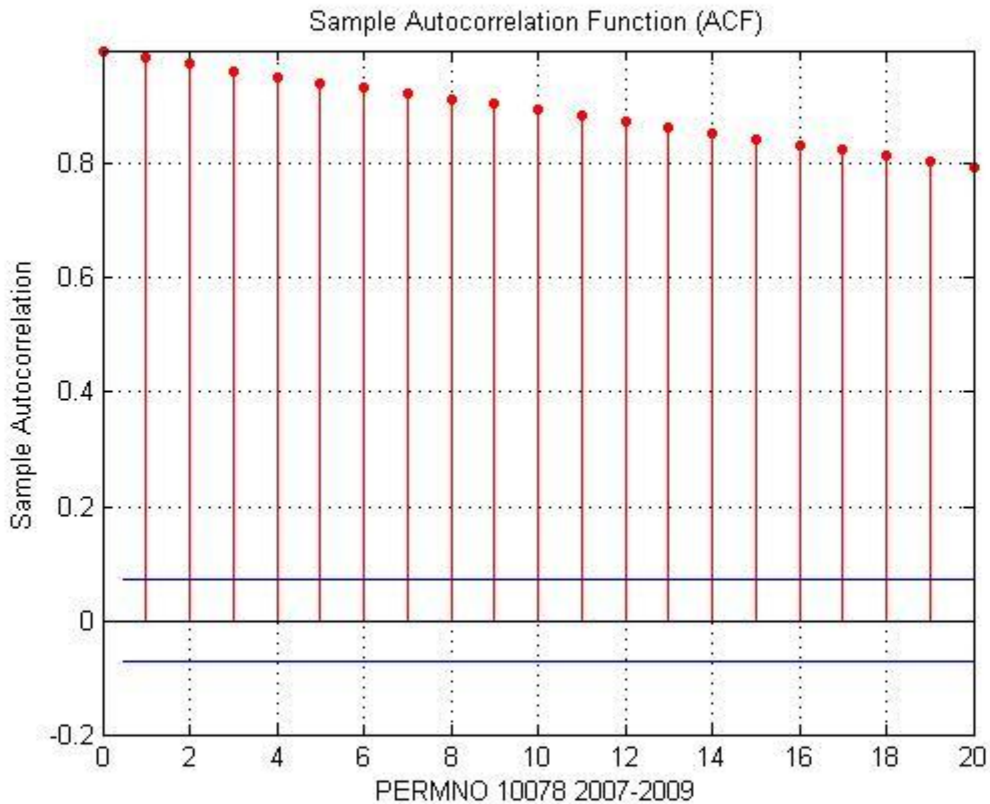
Inspection of ACF

- Inspection of the ACF gives hints if the process is stationary or not
- A stationary process implies a quick decay of the ACF
- An integrated process has a very slow decay of the ACF

Rapid decay of the ACF of returns



Slow decay of the ACF of prices



Formal test of stationarity

- Formal tests for deciding if a time series is integrated or stationary have been developed
- The best known is the (Augmented) Dickey-Fuller test ADF
- It tests $a=1$ against $a<1$ in the model

  $y(t)=ay(t-1)+b(t)$
- Non standard asymptotic distributions

Determine the number of lags

- The Box-Jenkins methodology requires inspecting the empirical ACF and the PACF (autocorrelation function at lags K keeping constant intermediate elements)
- And comparing the empirical ACF and PACF with theoretical models
- Requires experience and intuition

Determine the number of lags

- Formal model selection tests include AIC and BIC
- Given an ARMA(p,q) model call $\hat{\sigma}^2_{p,q}$ the variance of residuals $\hat{\sigma}^2_{p,q} = \frac{1}{T}\sum_{i=1}^{T}\hat{\varepsilon}_i^2(p,q)$
- Increasing the p and q $\hat{\sigma}^2_{p,q}$ becomes smaller
- But the model overfits that is fits unpredictable noise

Determine the number of lags

- The AIC and BIC criteria penalize models with many parameters
- The *Akaike information criterion* (AIC) minimizes $$AIC = \log\left(\hat{\sigma}^2_{p,q}\right) + \frac{2}{T}(p+q)$$

- The *Bayesian Information Criterion* (BIC) of Scwartz minimizes

$$BIC = \log\left(\hat{\sigma}^2_{p,q}\right) + \frac{\ln T}{T}(p+q)$$

Model estimation

The three major approaches to estimating ARMA models are

- The *Yule– Walker estimator* , which uses the Yule-Walker equations
- The *least squares estimator* (LSE) finds the parameter estimates that minimize the sum of the squared residuals
- The *maximum likelihood estimator* (MLE) maximizes the (exact or approximate) log-likelihood function associated with the specified model

The *Yule– Walker estimator*

- Yule-Walker (YW) equations can be used for parameter estimation of pure AR models
- YW equations for AR(p) process are

  $\rho_k = a_1 \rho_{k-1} + \cdots + a_p \rho_{k-p}$, $k = 1, 2, \ldots.$
- Or in matrix form

$$\begin{bmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & & \hat{\rho}_{p-2} \\ \hat{\rho}_2 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-3} \\ . & & & . \\ \hat{\rho}_{p-2} & & & \hat{\rho}_1 \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ a_{p-1} \\ a_p \end{bmatrix} = \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \hat{\rho}_3 \\ . \\ \hat{\rho}_{p-1} \\ \hat{\rho}_p \end{bmatrix}$$

- Hence $\hat{a} = \hat{T}^{-1} \hat{\rho}_p$

OLS estimators

- If only autoregressive terms are present OLS produces the usual regression linear estimator
- If moving average terms are present OLS leads to non linear minimization to be solved numerically

ML Estimators

- Requires specific distributional assumptions
- For example normal distributions
- Leads to non linear minimization problems to be solved numerically
- MATLAB vgxvarx instruction uses MLE meth
- Requires a VGX specification structure

Testing residuals for whiteness

- After modelling wee need to test that residuals are iid (white noise)
- Two tests:
- Durbin-Watson test
- Box-Pierce-Ljung test

Autocorrelation
Durbin-Watson test

- Given the AR(1) process, the null hypothesis of zero autocorrelation will be: H0 : r = 0
- The e are unobservable, so need to rely on the estimated residuals *e = Y -Xb.*
- These might exhibit some autocorrelation even if the null is true.
- No exact finite sample test valid for any X matrix
- The procedure:
- The DW statistic is given by:

$$DW = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2}$$

- The Durbin-Watson statistic is closely related to the first-order autocorrelation
- coefficient. Expanding the statistic, for large T: $DW \cong 2(1-\hat{\rho})$
- Heuristically, the range of d is from 0 to 4.
- DW < 2 for positive autocorrelation of the e
- DW > 2 for negative autocorrelation of the e
- DW=2 for zero autocorrelation of the e

Autocorrelation
Box-Pierce-Ljung-Box Statistic

- The Box-Pierce Q statistic is based on the squares of the first p autocorrelation coefficients of the OLS residuals. The statistic is defined as:

$$Q = T\sum_{j=1}^{p} r_j^2,\ r_j = \frac{\sum_{t=j+1}^{T} e_t e_{t-j}}{\sum_{t=1}^{T} e_t^2}$$

- The test
- Under the null of zero autocorrelation for the residuals, Q will have an asymptotic χ^2 distribution.
- Remark:
- An improved small-sample performance is expected from the revised Ljung-Box statistic:

$$Q' = T(T+2)T\sum_{j=1}^{p} \frac{r_j^2}{n-j}$$

Forecasting

- If a process follows an ARMA(p,q) model
- Predictions h steps ahead can be computed with the recursive formula

$$\hat{y}_t(h) = \sum_{i=1}^{p} a_i \hat{y}_t(h-i) + \sum_{j=0}^{q} b_j \hat{\varepsilon}_t(h-j)$$

Multivariate stationary ARMA processes

- A multivariate stationary process is called an Auto Regressive Moving Average Process of order *p,q* - ARMA(*p,q*) - if it satisfies an equation of the type:
- $$\mathbf{A}(L)\mathbf{x}_t = \mathbf{B}(L)\boldsymbol{\varepsilon}_t$$
- Where A and B are polynomial matrices in the lag operator of order *p* and *q* respectively:
- $$\mathbf{A}(L) = \sum_{i=1}^{p} \mathbf{A}_i L^i, \ \ \mathbf{A}_0 = \mathbf{I}, \ \ \mathbf{A}_p \neq 0$$

$$\mathbf{B}(L) = \sum_{j=1}^{q} \mathbf{B}_j L^j, \ \ \mathbf{B}_0 = \mathbf{I}, \ \ \mathbf{B}_q \neq 0$$

VAR

- Physical sciences are expressed through linear differential equations
- Differential equations link quantities and their rate of change
- Discretization of linear differential equations leads to finite difference models
- VAR models are difference equations

Vector Auto Regressive (VAR) Models

- In a VAR model, the current value of each variable is a linear function of the past values of all variables plus random disturbances
- In full generality, a VAR model can be written as follows:

$$\mathbf{x}_t = \mathbf{A}_1\mathbf{x}_{t-1} + \mathbf{A}_2\mathbf{x}_{t-2} + \cdots + \mathbf{A}_p\mathbf{x}_{t-p} + \mathbf{D}\mathbf{s}_t + \boldsymbol{\varepsilon}_t$$

where $\mathbf{x}_t = (x_{1,t},\cdots,x_{n,t})$ is a multivariate stochastic time series in vector notation, $\mathbf{A}_i,\ i = 1,2,\ldots,p,$ and $\mathbf{D}$ are deterministic *n*x*n* matrices, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t},\cdots,\varepsilon_{n,t})$ is a multivariate white noise with variance-covariance matrix $\boldsymbol{\Omega} = (\omega_1,\cdots,\omega_n)$ and $\mathbf{s}_t = (s_{1,t},\cdots,s_{n,t})$ is a vector of deterministic terms

VAR in lag notation

Using the lag-operator *L* notation, a VAR model can be written in the following form:

$$\mathbf{x}_t = \left(\mathbf{A}_1 L + \mathbf{A}_2 L^2 + \cdots + \mathbf{A}_n L^N\right)\mathbf{x}_t + \mathbf{D}\mathbf{s}_t + \boldsymbol{\varepsilon}_t$$

VAR in error-correction form

- VAR models can be written in equivalent forms in terms of the differences in the following error-correction form:

$$\Delta \mathbf{x}_t = \left(\mathbf{\Phi}_1 L + \mathbf{\Phi}_2 L^2 + \cdots + \mathbf{\Phi}_{n-1} L^{n-1} \right) \Delta \mathbf{x}_t + \mathbf{\Pi} L^n \mathbf{x}_t + \mathbf{D} \mathbf{s}_t + \boldsymbol{\varepsilon}_t$$

- where the first *n*-1 terms are in first differences and the last term is in levels
- The term in levels can be placed at any lag

Examples

$$\mathbf{x}_t = \mathbf{A}_1\mathbf{x}_{t-1} + \mathbf{A}_2\mathbf{x}_{t-2} + \mathbf{D}\mathbf{s}_t + \boldsymbol{\varepsilon}_t =$$

$$\mathbf{x}_t - \mathbf{x}_{t-1} = \left(\mathbf{A}_1 - \mathbf{I}\right)\mathbf{x}_{t-1} - \left(\mathbf{A}_1 - \mathbf{I}\right)\mathbf{x}_{t-2} + \left(\mathbf{A}_1 - \mathbf{I}\right)\mathbf{x}_{t-2} + \mathbf{A}_2\mathbf{x}_{t-2} + \mathbf{D}\mathbf{s}_t + \boldsymbol{\varepsilon}_t =$$

$$\Delta\mathbf{x}_t = \left(\mathbf{A}_1 - \mathbf{I}\right)\Delta\mathbf{x}_{t-1} + \left(\mathbf{A}_1 + \mathbf{A}_2 - \mathbf{I}\right)\mathbf{x}_{t-2} + \mathbf{D}\mathbf{s}_t + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\Phi}_1 = \left(\mathbf{A}_1 - \mathbf{I}\right),\ \boldsymbol{\Pi} = \left(\mathbf{A}_1 + \mathbf{A}_2 - \mathbf{I}\right)$$

$$\mathbf{x}_t = \mathbf{A}_1\mathbf{x}_{t-1} + \mathbf{A}_2\mathbf{x}_{t-2} + \mathbf{D}\mathbf{s}_t + \boldsymbol{\varepsilon}_t =$$

$$\mathbf{x}_t - \mathbf{x}_{t-1} = \left(\mathbf{A}_1 + \mathbf{A}_2 - \mathbf{I}\right)\mathbf{x}_{t-1} - \mathbf{A}_2\mathbf{x}_{t-1} + \mathbf{A}_2\mathbf{x}_{t-2} + \mathbf{D}\mathbf{s}_t + \boldsymbol{\varepsilon}_t =$$

$$\Delta\mathbf{x}_t = \left(-\mathbf{A}_2\right)\Delta\mathbf{x}_{t-1} + \left(\mathbf{A}_1 + \mathbf{A}_2 - \mathbf{I}\right)\mathbf{x}_{t-1} + \mathbf{D}\mathbf{s}_t + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\Phi}_1 = \left(-\mathbf{A}_2\right),\ \boldsymbol{\Pi} = \left(\mathbf{A}_1 + \mathbf{A}_2 - \mathbf{I}\right)$$

Equivalence VAR(*p*)-VAR(1)

- Consider a VAR*(p)* model:

$$\mathbf{x}_t = \left(\mathbf{A}_1 L + \mathbf{A}_2 L^2 + \cdots + \mathbf{A}_p L^p\right)\mathbf{x}_t + \mathbf{D}\mathbf{s}_t + \boldsymbol{\varepsilon}_t$$

- Adding variables, it can be transformed into a VAR(1) model:

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_t + \mathbf{S}_t + \mathbf{W}_t$$

Equivalence VAR(p)-VAR(1)

- Where:

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{bmatrix}, \ \mathbf{A} = \begin{bmatrix} \mathbf{A_1} & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_n & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{I}_n & \cdots & 0 & 0 \\ 0 & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{I}_n & 0 \end{bmatrix}, \ \mathbf{S}_t = \begin{bmatrix} \mathbf{Ds}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \mathbf{W}_t = \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- The characteristic equations of the two models have the same solutions

State space models

- A reasonable generalization of factor market models are state-space models :
- $$\mathbf{r}_t = \mathbf{a} + \mathbf{A}\mathbf{z}_t + \mathbf{B}\boldsymbol{\varepsilon}_t$$
  $$\mathbf{z}_{t+1} = \mathbf{C}\mathbf{z}_t + \mathbf{D}\boldsymbol{\varepsilon}_t$$
- The first equation is the usual regression of a factor market model
- The second equation is a VAR(1) model that describes the autoregressive dynamics of the factors

Equivalence of ARMA and State-Space Models

- ARMA models

$$\mathbf{A}(L)\mathbf{x}_t = \mathbf{B}(L)\boldsymbol{\varepsilon}_t$$

- and state-space models

$$\mathbf{r}_t = \mathbf{a} + \mathbf{A}\mathbf{z}_t + \mathbf{B}\boldsymbol{\varepsilon}_t$$

$$\mathbf{z}_{t+1} = \mathbf{C}\mathbf{z}_t + \mathbf{D}\boldsymbol{\varepsilon}_t$$

- are equivalent

Equivalence ARMA State-Space

- To see the equivalence between ARMA and state-space models, consider a univariate ARMA(*p*,*q*) model

$$x_t = \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{j=0}^{q} \psi_j \varepsilon_{t-j}, \ \ \psi_0 = 1$$

- This ARMA model is equivalent to the following state-space model:

$$x_t = \mathbf{C} z_t$$

$$z_t = \mathbf{A} z_{t-1} + \varepsilon_t$$

Equivalence ARMA State-Space

where $\mathbf{C} = \left[\varphi_1 \ldots \varphi_p 1 \psi_1 \ldots \psi_q \right]$

and

$$z_t = \begin{bmatrix} x_{t-1} \\ \vdots \\ x_{t-p} \\ \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-q} \end{bmatrix} \text{ and } \mathbf{A} = \begin{bmatrix} -\varphi_1 & \ldots & -\varphi_p & 1 & \psi_1 & \ldots & \psi_{q-1} & \psi_q \\ 1 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

Multivariate random walk

The multivariate random walk model of logprices is the simplest VAR model:

$$\mathbf{p}_t = \mathbf{p}_t + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t$$

$$\mathbf{r}_t = \Delta \mathbf{p}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t$$

Stationarity

- The stationarity and stability properties of a VAR*(p)* model depend on the roots of the det of the polynomial matrix

$$\mathbf{I} - \mathbf{A}_1 z - \mathbf{A}_2 z^2 - \cdots - \mathbf{A}_p z^p$$

- or $\det(\mathbf{I}z - \mathbf{A}) = 0$ in the equivalent VAR(1)representation
- Roots are the eigenvalues of **A**

Stationarity

- If all the roots of the above polynomial are strictly outside the unit circle, the VAR process is stationary
- In this case, the VAR process can be inverted and re-written as a infinite moving average of white noise
- If all the roots are outside the unit circle with the exception of a few roots on the unit circle, then the VAR process is integrated
- In this case it cannot be inverted as an infinite moving average. If some of the roots are inside the unit circle, then the process is explosive
- If the VAR process starts at some initial point characterized by initial values or distributions, then the process cannot be stationary.
- However, if all the roots are outside the unit circle, the process is asymptotically stationary.
- If some root is equal to 1, then the process can be differentiated to obtain an asymptotically stationary process

Integrated variables

- A series is integrated of order *n* if it can be transformed into a stationary series differencing *n* times
- In particular, a uni-variate time series *X* is integrated of order 1 if it can be represented as:

$$X_{t+1} = \rho X_t + b + \varepsilon_t$$

$$\rho = 1$$

ε_t stationary possibly autocorrelated.

Integrated series

- The key feature of an integrated time series is that random innovations never decay
- Most economic variables are integrated variables
- In particular, testing for integration in log price processes, one finds that the null of integration cannot be rejected in most cases
- For instance, testing the logprice processes in the S&P 500 using a standard test such as the ADF test, the null of integration cannot be rejected in more than 90% of the time series

Trend Stationary and Difference Stationary Processes

- Trend stationary:

$$y_t = s(t) + \eta_t$$

- Difference stationary:

$$\Delta y_t = \eta_t$$

Stochastic Trends

- Deterministic trend:

$$s(t) = E\left[x_t\right]$$

- In integrated processes, past shocks never decay
- An integrated process can be decomposed in the sum of three components:
- a deterministic trend,
- a stochastic trend,
- and a cyclic stationary process

Stochastic Trends

- Consider a process integrated of order 1:

$$\Delta\mathbf{x}_t = \mathbf{\Psi}(L) = \left(\sum_{i=0}^{\infty}\mathbf{\Psi}_i L^i\right)\boldsymbol{\varepsilon}_t$$

- Write:

$$\mathbf{\Psi}(L) = \mathbf{\Psi} + (1-L)\left(\sum_{i=0}^{\infty}\mathbf{\Psi}_i^* L^i\right)$$

Stochastic Trends

We can write

$$\Delta \mathbf{x}_t = (1-L)\mathbf{x}_t = \left[\mathbf{\Psi} + (1-L)\left(\sum_{i=0}^{\infty} \mathbf{\Psi}_i^* L^i \right) \right] \boldsymbol{\varepsilon}_t$$

$$\mathbf{x}_t = \frac{\mathbf{\Psi}}{(1-L)} \boldsymbol{\varepsilon}_t + \left(\sum_{i=0}^{\infty} \mathbf{\Psi}_i^* L^i \right) \boldsymbol{\varepsilon}_t + \mathbf{x}_{-1}$$

$$\mathbf{x}_t = \mathbf{\Psi} \sum_{i=1}^{t} \boldsymbol{\varepsilon}_i + \left(\sum_{i=0}^{\infty} \mathbf{\Psi}_i^* L^i \right) \boldsymbol{\varepsilon}_t + \mathbf{x}_{-1}$$

Stochastic Trends

- The process is decomposed in:

$$\left(\boldsymbol{\Psi}\sum_{i=1}^{t}\boldsymbol{\varepsilon}_i\right)_t$$

- Stationary (asymptotically stationary) component:

$$\left(\sum_{i=0}^{\infty}\boldsymbol{\Psi}_i^* L^i\right)\boldsymbol{\varepsilon}_t$$

Intercepts

- Model has a constant intercept:

$$\Delta\mathbf{x}_t = (1-L)\mathbf{x}_t = \mathbf{v} + \left[\mathbf{\Psi} + (1-L)\left(\sum_{i=0}^{\infty}\mathbf{\Psi}_i^* L^i\right)\right]\boldsymbol{\varepsilon}_t$$

- The process can have a linear trend:

$$\mathbf{x}_t = \frac{\mathbf{\Psi}}{(1-L)}\boldsymbol{\varepsilon}_t + \left(\sum_{i=0}^{\infty}\mathbf{\Psi}_i^* L^i\right)\boldsymbol{\varepsilon}_t + \mathbf{x}_{-1} + \frac{\mathbf{\Psi}}{(1-L)}\mathbf{v}$$

$$\mathbf{x}_t = \mathbf{\Psi}\sum_{i=1}^{t}\boldsymbol{\varepsilon}_i + \left(\sum_{i=0}^{\infty}\mathbf{\Psi}_i^* L^i\right)\boldsymbol{\varepsilon}_t + \mathbf{x}_{-1} + t\mathbf{u}$$

- where: $\mathbf{u} = \mathbf{\Psi}\mathbf{v}$

Cointegration

- Suppose that a set of time series integrated of order 1 is given
- Though each series is integrated of order 1, for instance an arithmetic random walk, there might be linear combinations of the series that are stationary
- If this happens, the series are said to be cointegrated
- The financial meaning of cointegration is the following: individual log price processes can be arithmetic random walks but there are portfolios, in general long-short portfolios, which are stationary, and thus mean-reverting around a constant mean

Key Defining Properties of Cointegration

- Linear combination of integrated processes are stationary
- There are common integrated trends
- One integrated variable can be *meaningfully* regresssed over other integrated variables

Cointegration

- The concept of cointegration was introduced by Granger in 1981. It can be expressed in the following way
- Suppose that a set of *n* time series, integrated of order 1, is given.
- If there is a linear combination of the series

$$\delta_t = \sum_{i=1}^{n} \beta_i x_{i,t}$$

- which is stationary, then the series are said to be cointegrated. Any linear combination as the one above is called a cointegrating relationship. Given *n* time series, there can be from none to at most *n*-1 cointegrating relationships

Uniqueness

- Cointegration vectors are not unique
- Given two cointegrating vectors:

$$\sum_{i=1}^{N} \alpha_i X_i, \quad \sum_{i=1}^{N} \beta_i X_i,$$

- A linear combination of the cointegrating vectors is another cointegrating vector:

$$A\sum_{i=1}^{N} \alpha_i X_i + \mathrm{B}\sum_{i=1}^{N} \beta_i X_i,$$

Common trends

- Stock and Watson first observed (1988) that a cointegrated model with *r* cointegrating relationships admits *n-r* common trends
- This means that all time series can be written in the form
- $$\mathbf{r}_t = \mathbf{a} + \mathbf{A}\mathbf{z}_t + \mathbf{B}\boldsymbol{\eta}_t$$
- where the *z* are common stochastic trends

Common trends

- Suppose there are *n* time series $x_{i,t}, i = 1,\ldots,n$ and *k*<*n* cointegrating relationships
- It can be demonstrated that there are *n-k* integrated time series $u_{j,t}$, $\mathrm{j} = 1,\ldots,n-k$, called common trends such that every time series can expressed as a linear combination of the common trends plus a stationary disturbance:

$$x_{it} = \sum_{j=1}^{n-k} \gamma_j u_{j,t} + \eta_{i,t}$$

Common Stochastic Trends

- Arrange the cointegrating relationships in
- a $n \times r$ matrix:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{1,1} & \cdots & \beta_{1,n-k} \\ \vdots & \ddots & \vdots \\ \beta_{n,1} & \cdots & \beta_{n,n-k} \end{pmatrix}$$

- The *r*-variate process: $\boldsymbol{\beta}'\mathbf{x}_t$ is stationary

Common stochastic trends

- Consider:

$$\mathbf{x}_t = \mathbf{\Psi}\sum_{i=1}^{t}\boldsymbol{\varepsilon}_i + \left(\sum_{i=0}^{\infty}\mathbf{\Psi}_i^* L^i\right)\boldsymbol{\varepsilon}_t + \mathbf{x}_{-1}$$

- It can be demonstrated that: $\boldsymbol{\beta}'\mathbf{\Psi} = 0$
- Stochastic trends have been removed:

$$\boldsymbol{\beta}'\mathbf{x}_t = \mathbf{z}_t = \boldsymbol{\beta}'\left(\sum_{i=0}^{\infty}\mathbf{\Psi}_i^* L^i\right)\boldsymbol{\varepsilon}_t + \boldsymbol{\beta}'\mathbf{x}_{-1}$$

Common Trends

- There is a $n \times r$ matrix $\mathbf{H}_1$ such that:

$$\mathbf{x}_t = (\mathbf{\Psi H})\left(\mathbf{H}^{-1}\sum_{i=1}^{t}\boldsymbol{\varepsilon}_i\right) + \left(\sum_{i=0}^{\infty}\mathbf{\Psi}_i^* L^i\right)\boldsymbol{\varepsilon}_t + \mathbf{x}_{-1} = \mathbf{A}\boldsymbol{\tau}_t + \left(\sum_{i=0}^{\infty}\mathbf{\Psi}_i^* L^i\right)\boldsymbol{\varepsilon}_t + \mathbf{x}_{-1}$$

$$\boldsymbol{\tau}_t = \mathbf{H}^{-1}\sum_{i=1}^{t}\boldsymbol{\varepsilon}_i$$

- are the common trends
- Deterministic trends must be removed separately

ECM

- Is there a general representation of cointegrated processes?
- The answer is positive. Granger was able to demonstrate that a multivariate integrated process is cointegrated if and only if it can be represented in the **Error Correction Model (ECM**) form.

ECM representation

- The ECM representation is a representation of a multivariate process in first differences with corrections in levels as follows:

$$\Delta \mathbf{x}_{t+1} = \left(\sum_{i=1}^{n-1} \mathbf{A} L^i \right) \Delta \mathbf{x}_t + \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{x}_t + \boldsymbol{\eta}_t$$

- where α is a *pxr* matrix, $\boldsymbol{\beta}$ is a a *pxr* matrix with and $\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\Pi}$ is a vector of stationary disturbances $\boldsymbol{\eta}_t$

ARDL

- The **Auto Regressive Distributed Lag (ARDL)** model takes into account exogenous variables that are not cointegrated among themselves (Pesaran and Shin)
- An ARDL model has the following form:

$$x_t = \alpha_0 + \alpha_1 t + \left(\Phi_{\mathbf{1}} L + \Phi_2 L^2 + \cdots + \Phi_p L^p\right) x_t +$$

$$+ \beta' z_t + \left(\beta'_{\mathbf{1}} L + \beta'_2 L^2 + \cdots + \beta'_q L^q\right) \Delta z + u_t$$

$$\Delta z_t = \left(P_{\mathbf{1}} L + P_2 L^2 + \cdots + P_s L^s\right) \Delta z_t + \boldsymbol{\varepsilon}_t$$

where the *z* are I(1) variables non cointegrated while *x* exhibits a long run relationships with the *zs.*

Stochastic and Deterministic Cointegration

- A process is said to be *stochastically cointegrated* if there are linear combinations of the process components that are stationary plus an eventual deterministic trend.
- In other words, stochastic cointegration removes the stochastic trends but not necessarily the deterministic trends.
- The process is said to be *deterministically cointegrated* if there are linear combinations of the process components, each including its own deterministic trend, that are stationary without any deterministic trend.
- In other words, deterministic cointegration removes both the stochastic trends and the deterministic trends.

Dynamic cointegration

- Cointegrating relationships are static relationships between variables taken at the same time
- The idea of dynamic cointegration is to introduce a small number of lags in the cointegrating relationship
- In other words, cointegration reduces the order of integration by applying linear regressions between variables. Dynamic cointegration reduces the order of integration by applying autoregressive modeling
- A VAR model with *n* lags

$$\mathbf{x}_t = \mathbf{A_1}\mathbf{x}_{t-1} + \mathbf{A}_2\mathbf{x}_{t-2} + \cdots + \mathbf{A}_n\mathbf{x}_{t-n} + \boldsymbol{\varepsilon}_t$$

- exhibits dynamic cointegration if there exists a **stationary** autoregressive combination of the variables of the type:
- $$\alpha'\mathbf{x}_t + \beta'\Delta\mathbf{x}_t$$
- Cointegration and dynamic cointegration can coexist

Equivalence
Cointegration State-Space

- In particular, cointegrated models

$$\Delta X_t = \mathbf{A}_1 \Delta X_{t-1} + \mathbf{A_2} \Delta X_{t-2} + \mathbf{\Pi} X_{t-1} + \varepsilon_t$$

- are equivalent to state space models

$$\begin{cases} \mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{u}_t \\ \mathbf{x}_t = \mathbf{C}\mathbf{z}_t + \mathbf{D}\mathbf{u}_t \end{cases}$$

Common Trends and State-Space

- Cointegrated processes can be represented as regressions on common trends

$$\mathbf{r}_t = \mathbf{a} + \mathbf{A}\mathbf{z}_t + \mathbf{B}\boldsymbol{\eta}_t$$

- Models with common trends are equivalent, obviously, to state-space models - but the number of state-variables and common trends might not coincide due to autocorrelation of residuals

Dimensionality Reduction

- The previous equivalences do not represent any dimensionality reduction
- They establish statistically equivalent representation
- The simplest example is the diagonalization of variance covariance matrix: per se, it does not implement any dimensionality reduction

Dynamic Factor Models

- Factor analytic models are represented by equations of the following type:

$$\mathbf{p}_t = \mathbf{s}_t + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{p}_{t-i} + \sum_{j=1}^{q} \mathbf{B}_j \mathbf{f}_{t-j} + \mathbf{u}_t$$

$$\mathbf{f}_{t+1} = \sum_{k=1}^{s} \mathbf{C}_k \mathbf{f}_{t-k} + \boldsymbol{\varepsilon}_t$$

- where the price processes $\mathbf{p}_t$ have an autoregressive distributed-lag dynamics, the factors $\mathbf{f}_t$ follow a VAR model and the noise terms $\mathbf{u}_t, \boldsymbol{\varepsilon}_t$ are mutually uncorrelated white noises

ARDL and Dynamic Factor Models

ARDL models are formally analogous to DFM but exogenous variables are typically observable and not hidden

Model Estimation Methods

Cointegration at work

Estimation and empirical evidence

- Empirical evidence of cointegration and reversals in equity price processes
- Estimation of ARDL models
- Estimation of cointegrated systems
- Estimation of state-space models

Dimensionality reduction

- Above techniques are not, per se, dimensionality reduction techniques
- Need to reduce the dimensionality of dynamic models in order to estimate them effectively and to extract information from noise
- Hence need for a conceptual model of price and return processes

Empirical facts

- Heteroschedasticity of volatility (well known ARCH/GARCH behaviour)
- Correlation return/volatility (ARCH-M) mixed evidence
- Reversals (Lo and MacKinlay)
- Momentum (Jagadeesh and Titman)
- Less than linear growth of variance (Lo)
- Autocorrelations

What model for stock price processes?

- Any model we choose is an approximation
- If we work with linear models, it is likely that there are structural breaks
- Different phenomena at different time horizons: reversals, momentum, etc
- A variety of phenomena at different time horizons affect modelling

Dynamic factors

- A possible modelling choice: dynamic factors analysis
- Conceptually, we assume that the market over time horizons in the range of 1-4 years can be described as an ARDL system with a small number of integrated factors

Estimation of cointegrated systems

- We can hope to arrive at only approximate solutions
- Large numbers of processes, hundreds or even thousands
- The key problem is to find the common dynamic factors in hundreds or thousands of processes

Estimation of ARDL models

- Pesaran and Shin extended the theory of ARDL models to cointegrated processes
- Estimation performed with standard regression techniques: OLS and MLE
- Critical point: residuals need to be serially uncorrelated
- Need to add lags; number of lags estimated with Akaike Information Criterion (AIC) or Schwartz's BIC
- Estimation sensitive to trends

Estimation of cointegrated systems

- Regression methods - Engle & Granger
- MLE methods - Johansen et alii
- Eigenvalues companion matrix - Ahlgren&Nyblom
- CCA – Bossaert, Pena & Poncela
- PCA - Stock & Watson
- Subspace methods – Bauer & Wagner

Engle-Granger

- It first estimates the long-run regression,
- then estimates the short-run dynamics
- Can be used only for pairs of processes

Johansen

- Consider the model:

$$\Delta \mathbf{x}_{t+1} = \left(\sum_{i=1}^{n-1} \mathbf{A} L^i \right) \Delta \mathbf{x}_t + \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{x}_t + \boldsymbol{\varepsilon}_t$$

- Johansen method estimates

$$\boldsymbol{\alpha}, \boldsymbol{\beta}$$

- with MLE estimates

Johansen

- State-of-the art estimate
- Critical values tabulated only up to a limited number of processes
- Does not scale up to hundreds or thousands

Eigenvalues Companion Matrix-Ahlgren&Nyblom

- Consider a process in a VAR(1 form):

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$$

- **A** is called the companion matrix of the original VAR*(p)* system
- **A** can be decomposed:

$$\mathbf{SAS}^{-1} = \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{\Omega} \end{pmatrix}$$

Eigenvalues Companion Matrix-Ahlgren&Nyblom

- Partition **S** conformably
- The processes:

$$\mathbf{S}_1\mathbf{X}$$

$$\mathbf{S}_2\mathbf{X}$$

- are respectively integrated and stationary

CCA-based methods - Bossaerts

- Canonical correlations
- Partition a random vector:

$$\mathbf{X} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}$$

- Partition conformably:

$$\mathbf{X} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}, \ \mathbf{x} = \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{y}} \\ \boldsymbol{\mu}_{\mathbf{z}} \end{pmatrix}, \ \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{\mathbf{yy}} & \boldsymbol{\Omega}_{\mathbf{yz}} \\ \boldsymbol{\Omega}_{\mathbf{yz}} & \boldsymbol{\Omega}_{\mathbf{zz}} \end{pmatrix}$$

CCA

- Consider two vectors of constants:

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_S \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_R \end{pmatrix}$$

- We want to maximize correlations between the variables:

$$\boldsymbol{\alpha}\mathbf{Y}, \ \boldsymbol{\beta}\mathbf{Z}$$

CCA

If vectors unitary then:

$$\boldsymbol{\Omega}_{\mathbf{yy}}^{-1}\boldsymbol{\Omega}_{\mathbf{yz}}\boldsymbol{\Omega}_{\mathbf{zz}}^{-1}\boldsymbol{\Omega}_{\mathbf{zy}}\boldsymbol{\alpha} = \rho^2\boldsymbol{\alpha}$$

$$\boldsymbol{\Omega}_{\mathbf{zz}}^{-1}\boldsymbol{\Omega}_{\mathbf{zy}}\boldsymbol{\Omega}_{\mathbf{yy}}^{-1}\boldsymbol{\Omega}_{\mathbf{yz}}\boldsymbol{\beta} = \rho^2\boldsymbol{\beta}$$

CCA-based methods

CCA-based methods search for canonical correlations between process variables at different time lags