ECONOMETRICS Part II
PhD LBS

Luca Gambetti
UAB, Barcelona GSE

February-March 2014

Contacts

Prof.: Luca Gambetti
email: luca.gambetti@uab.es
webpage: http://pareto.uab.es/lgambetti/

Description

This is an introductory Time Series Analysis with applications in macroeconomics.

Goal of the course

The main objective of the course is to provide the students with the knowledge of a comprehensive set of tools necessary for empirical research with time series data.

Contents

References

1. P. J.Brockwell, and R. A. Davis, (2009), Time Series: Theory and Methods,Springer-Verlag: Berlin.
2. J. D. Hamilton (1994), Time Series Analysis, Princeton University Press: Princeton.
3. H. Lutkepohl (2005), New Introduction to Multiple Time Series, Springer-Verlag: Berlin.
4. T.J. Sargent (1987) Macroeconomic Theory, Academic Press.

Grades

Take-Home Exam.

Econometric Software

MATLAB.

1. INTRODUCTION

1 What does a macroeconometrican do?

”Macroeconometricians do four things: describe and summarize macroeconomic data, make macroeconomic forecasts, quantify what we do or do not know about the true structure of the macroeconomy, and advise (and sometimes become) macroeconomic policymakers.” Stock and Watson, JEP, 2001.

Except advising and becoming policymakers, this is what we are going to do in this course.

2 Preliminaries

2.1 Lag operators

- The lag operator L maps a sequence $\{x_t\}$ into a sequence $\{y_t\}$ such that $y_t = Lx_t = x_{t-1}$, for all t.

- It can be applied repeatedly on a process, for instance $L(L(Lx_t)) = L^3 x_t = x_{t-3}$.

- If we apply L to a constant c, $Lc = c$.

- Inversion: L^{-1} is the inverse of L, $L^{-1}x_t = x_{t+1}$. It is such that $L^{-1}(L)x_t = x_t$.

- The lag operator and multiplication operator are commutative $L(\beta x_t) = \beta L x_t$ (β a constant).

- The lag operator is distributive over the addition operator $L(x_t + w_t) = Lx_t + Lw_t$

- Polynomials in the lag operator:

$$\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + ... + \alpha_p L^p$$

is a polynomial in the lag operator of order p and is such that

$$\alpha(L)x_t = \alpha_0 x_t + \alpha_1 x_{t-1} + ... + \alpha_p x_{t-p}$$

with α_i ($i = 1, ..., p$) constant.

• Absolutely summable (one-sided) filters. Let $\{\alpha_j\}_{j=0}^{\infty}$ be a sequence of absolutely summable coefficients, i.e. $\sum_{j=0}^{\infty} |\alpha_j| < \infty$. We define the filter

$$\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + ...$$

which gives

$$\alpha(L)x_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + ...$$

• Square summable (one-sided) filters. Let $\{\alpha_j\}_{j=0}^{\infty}$ be a sequence of square summable coefficients, i.e. $\sum_{j=0}^{\infty} \alpha_j^2 < \infty$. We define the filter

$$\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + ...$$

which gives

$$\alpha(L)x_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + ...$$

• Absolute summability implies square summability: if $\sum_{j=0}^{\infty} |\alpha_j| < \infty$ then $\sum_{j=0}^{\infty} \alpha_j^2 < \infty$.

• $\alpha(0) = \alpha_0$.

• $\alpha(1) = \alpha_0 + \alpha_1 + \alpha_2 + ...$

• $\alpha(L)(bx_t + cw_t) = \alpha(L)bx_t + \alpha(L)cw_t$.

• $\alpha(L)x_t + \beta(L)x_t = (\alpha(L) + \beta(L))x_t$.

• $\alpha(L)[\beta(L)x_t] = \beta(L)[\alpha(L)x_t]$.

• Polynomials factorization: consider

$$\alpha(L) = 1 + \alpha_1 L + \alpha_2 L^2 + ... + \alpha_p L^p$$

Then

$$\alpha(L) = (1 - \lambda_1 L)(1 - \lambda_1 L)...(1 - \lambda_p L)$$

where λ_i $(i = 1, ..., p)$ are the reciprocals of the roots of the polynomial in L.

• Lag polynomials can also be inverted. For a polynomial $\phi(L)$, we are looking for the values of the coefficients α_i of $\phi(L)^{-1} = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + ...$ such that $\phi(L)^{-1}\phi(L) = 1$.

Example: $p = 1$. Let $\phi(L) = (1 - \phi L)$ with $|\phi| < 1$. To find the inverse write

$$(1 - \phi L)(\alpha_0 + \alpha_1 L + \alpha_2 L^2 + ...) = 1$$

note that all the coefficients of the non-zero powers of L must be equal to zero. So This gives

$$\begin{aligned} \alpha_0 &= 1 \\ -\phi + \alpha_1 &= 0 \quad \Rightarrow \alpha_1 = \phi \\ -\phi\alpha_1 + \alpha_2 &= 0 \quad \Rightarrow \alpha_2 = \phi^2 \\ -\phi\alpha_2 + \alpha_3 &= 0 \quad \Rightarrow \alpha_3 = \phi^3 \end{aligned}$$

and so on. In general $\alpha_k = \phi^k$, so $(1 - \phi L)^{-1} = \sum_{j=0}^{\infty} \phi^j L^j$ provided that $|\phi| < 1$.

It is easy to check this because

$$(1 - \phi L)(1 + \phi L + \phi^2 L^2 + ... + \phi^k L^k) = 1 - \phi^{k+1} L^{k+1}$$

so

$$(1 + \phi L + \phi^2 L^2 + ... + \phi^k L^k) = \frac{1 - \phi^{k+1} L^{k+1}}{(1 - \phi L)}$$

and $\sum_{j=0}^{k} \phi^j L^j \to \frac{1}{(1-\phi L)}$ as $k \to \infty$.

Let T be an index which has an ordering relation defined on it. Thus, if $t_1, t_2 \in T$, then $t_1 \leq t_2$ or $t_1 > t_2$. Usually, $T = \mathbb{R}$ or $T = \mathbb{Z}$, the sets of real or integer numbers, respectively.

Stochastic process. A stochastic process is a family of random variables $\{X_t, t \in T\}$ indexed by some set T .

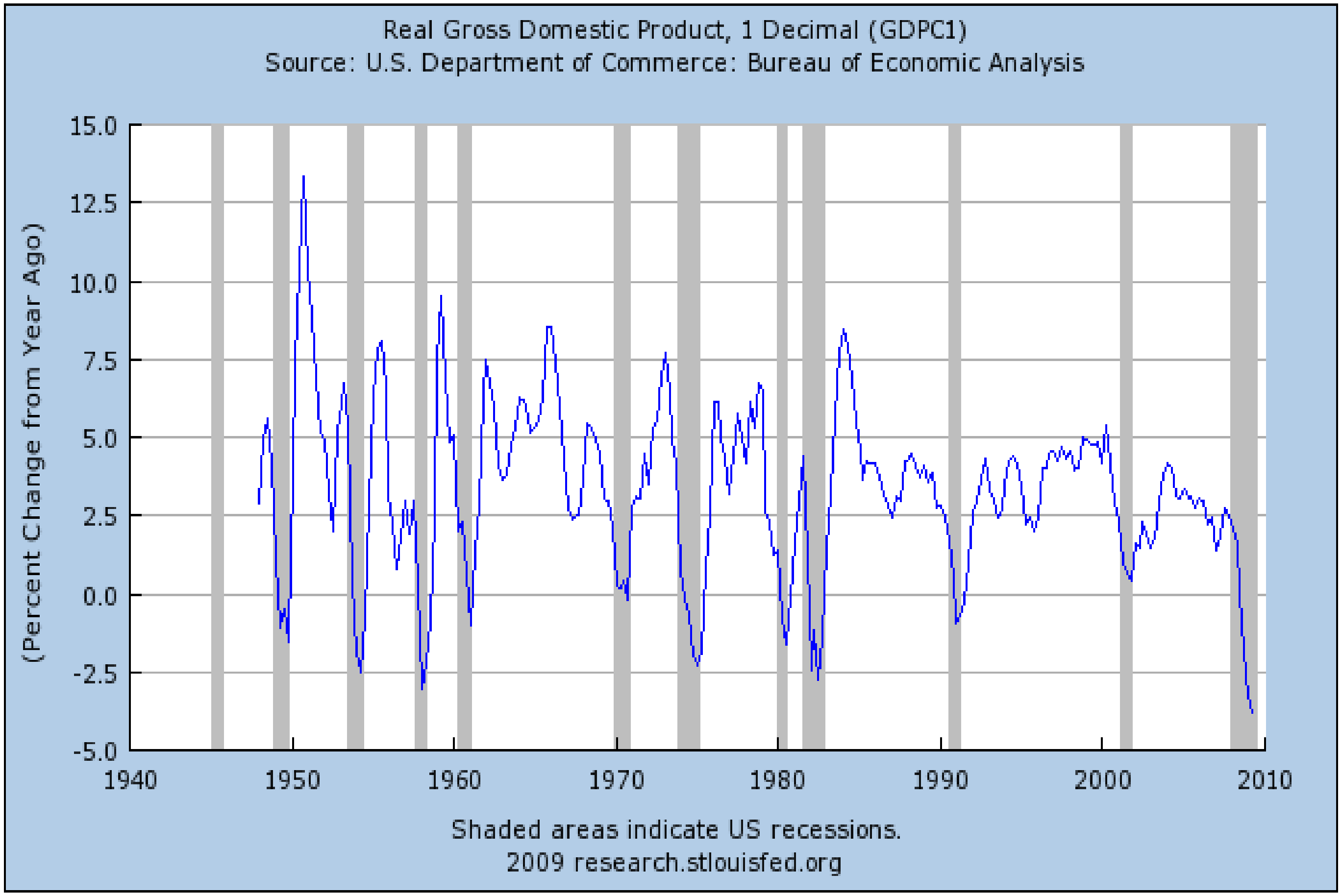
A stochastic process can be discrete or continuous according to whether T is continuous, e.g., $T = \mathbb{R}$, or discrete, e.g., $T = \mathbb{Z}$.

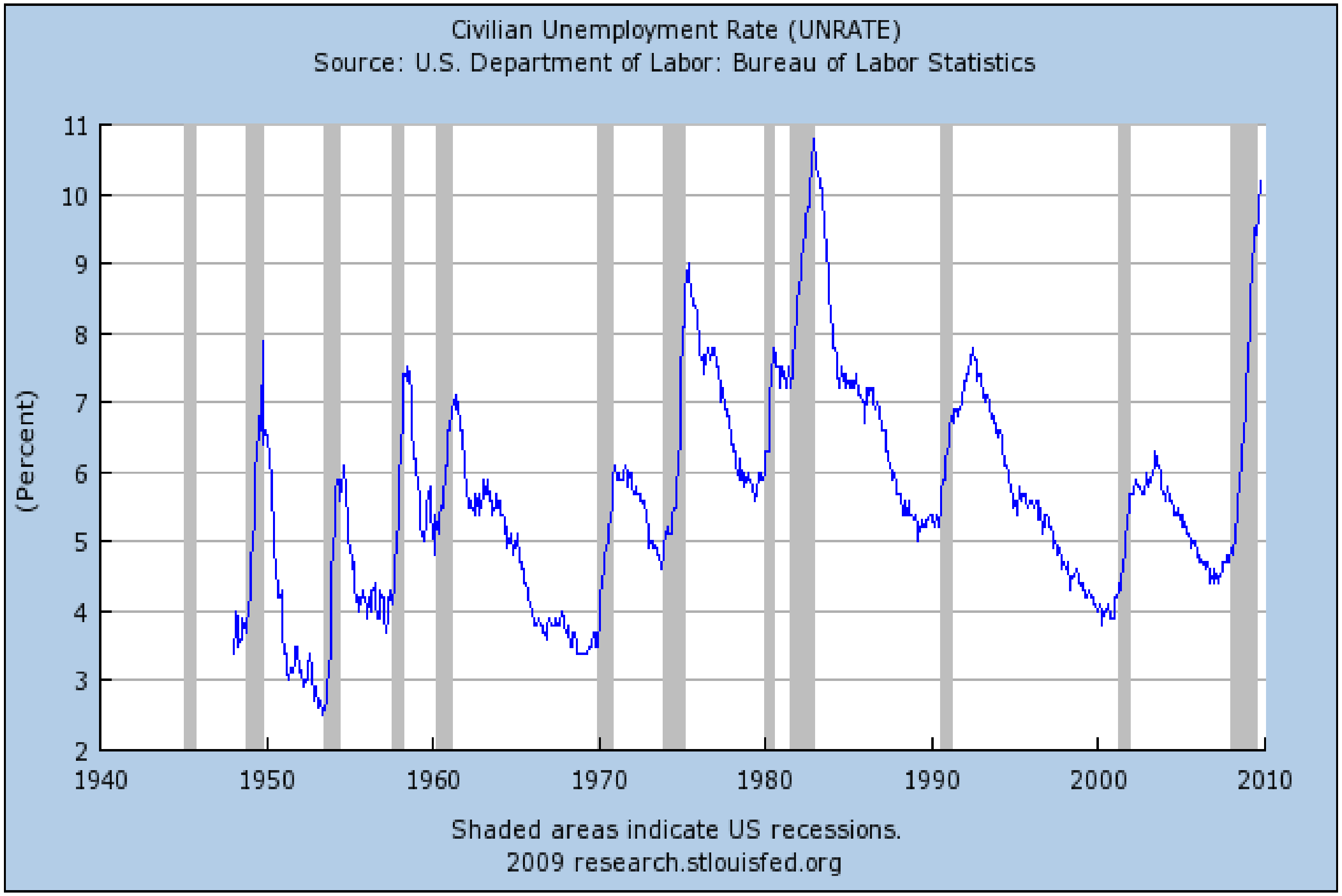
Time series. *A time series is part of a realization of a stochastic process*

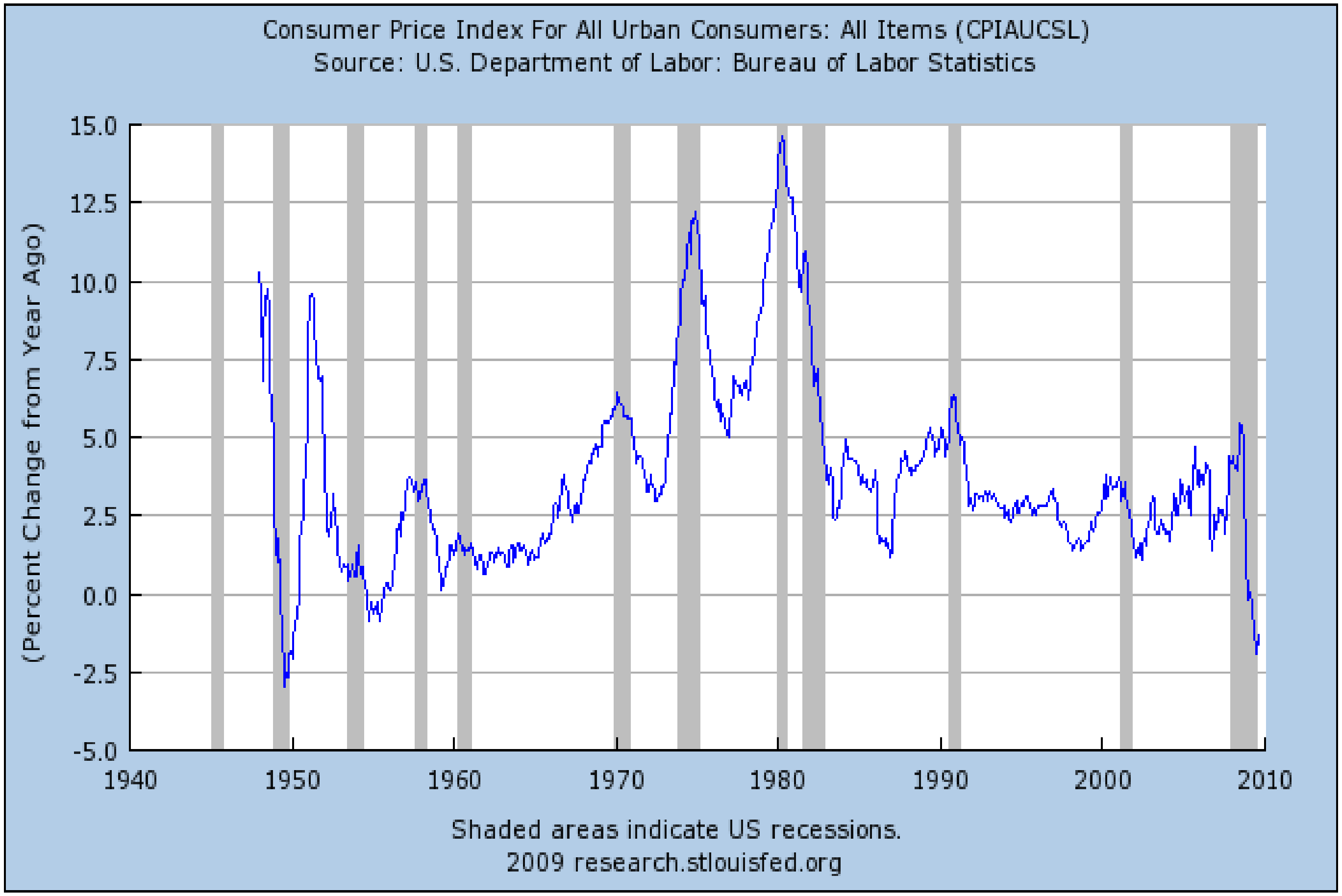
Example 1 Let the index set be $T = \{1, 2, 3\}$ and let the space of outcomes Ω be the possible outcomes associated with tossing one dice:

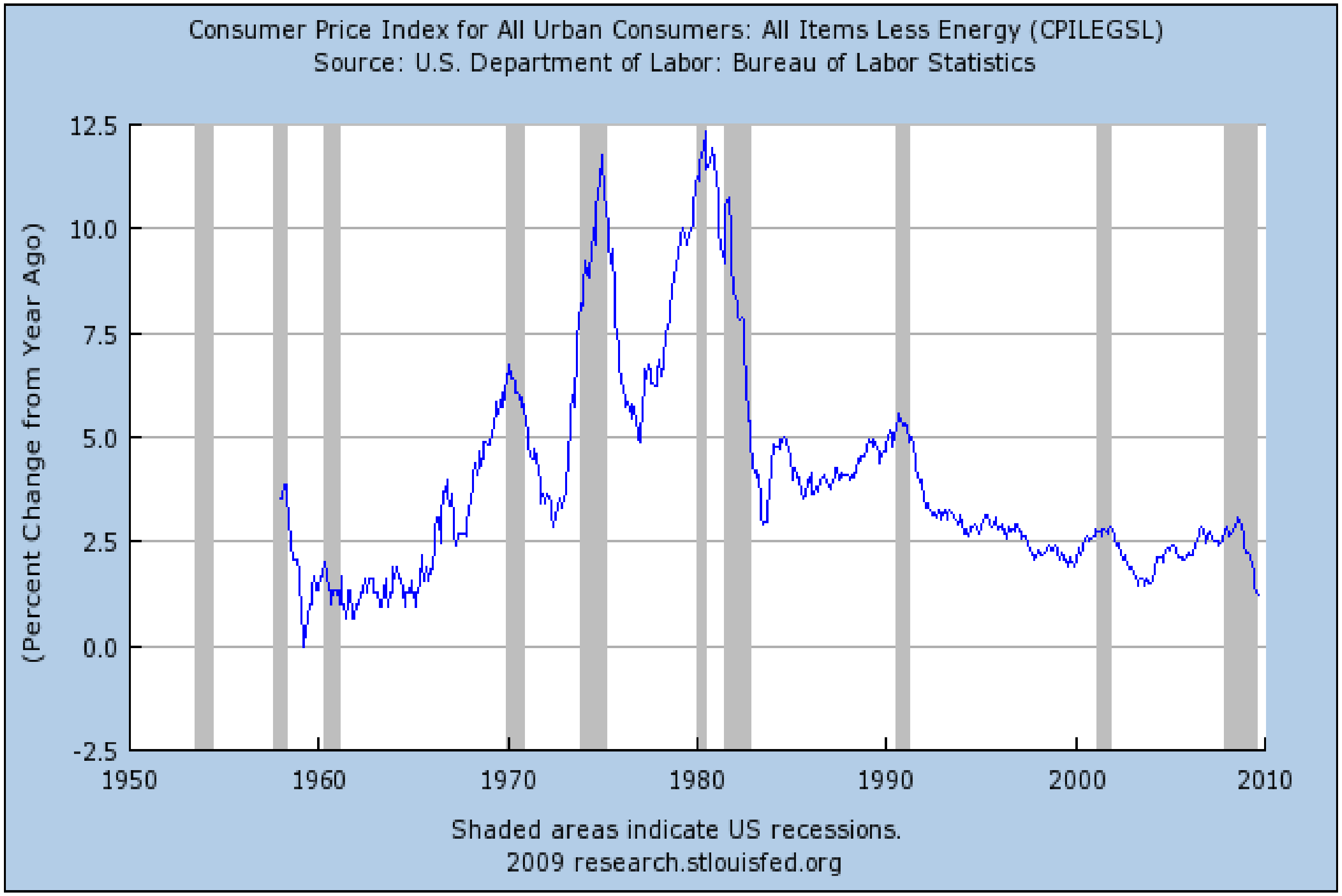
$$\Omega = 1, 2, 3, 4, 5, 6$$

Define $X_t = t + [\text{value on dice}]^2 t$. Therefore for a particular ω, say $\omega = 3$, the realization or path would be (10, 20, 30), and this stochastic process has 6 different possible realizations (associated to each of the values of the dice).


Real Gross Domestic Product, 1 Decimal (GDPC1)
Source: U.S. Department of Commerce: Bureau of Economic Analysis
(Percent Change from Year Ago)
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
1940
1950
1960
1970
1980
1990
2000
2010
Shaded areas indicate US recessions.
2009 research.stlouisfed.org


Civilian Unemployment Rate (UNRATE)
Source: U.S. Department of Labor: Bureau of Labor Statistics
(Percent)
11
10
9
8
7
6
5
4
3
2
1940
1950
1960
1970
1980
1990
2000
2010
Shaded areas indicate US recessions.
2009 research.stlouisfed.org


Consumer Price Index For All Urban Consumers: All Items (CPIAUCSL)
Source: U.S. Department of Labor: Bureau of Labor Statistics
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
(Percent Change from Year Ago)
1940
1950
1960
1970
1980
1990
2000
2010
Shaded areas indicate US recessions.
2009 research.stlouisfed.org


Consumer Price Index for All Urban Consumers: All Items Less Energy (CPILEGSL)
Source: U.S. Department of Labor: Bureau of Labor Statistics
(Percent Change from Year Ago)
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
1950
1960
1970
1980
1990
2000
2010
Shaded areas indicate US recessions.
2009 research.stlouisfed.org

2.2 Stationarity

This course will mainly focus on stationary processes. There are two definitions of stationarity: strict and weak (or second order).

Strict Stationarity *The time series* X_t *is said to be strictly stationary if the joint distributions of* $(X_{t_1}, ...X_{t_k})'$ *and* $(X_{t_1+h}, ...X_{t_k+h})'$ *are the same for all positive integers for all* $t_1, ..., t_k$*,* $h \in \mathbb{Z}$*.*

Interpretation: This means that the graphs over two equal-length time intervals of a realization of the time series should exhibit similar statistical characteristics.

In order to define the concept of weak stationarity we first need to introduce the concept of autocovariance function. This function is a measure of dependence between elements of the sequence X_t.

The autocovariance function *If* X_t *is a process such that* $Var(X_t) < \infty$ *for each* $t \in T$*, then the autocovariance function* $\gamma_t(r, s)$ *of* X_t *is defined by*

$$\gamma(r, s) = Cov(X_r, X_s) = E[(X_r - E(X_r))(X_s - E(X_s))], \tag{1}$$

Weak Stationarity *The time series X_t is said to be weakly stationary if*

(i) $E|X_t|^2 < \infty$ for all t.

(ii) $E(X_t) = \mu$ for all t.

(iii) $\gamma(r, s) = \gamma(r + t, s + t)$ for all t, r, s.

Notice that for a covariance stationary process $\gamma(r, s) = \gamma(r - s, 0) = \gamma(h)$.
In summary: weak stationarity means that the mean, the variance are finite and constant and that the autocovariance function only depends on h, the distance between observations.

If $\gamma(.)$ is the autocovariance function of a stationary process, then it satisfies

(i) $\gamma(0) \geq 0$.

(ii) $|\gamma(h)| \leq \gamma(0)$ for all $h \in \mathbb{Z}$.

(iii) $\gamma(-h) = \gamma(h)$ for all $h \in \mathbb{Z}$

Autocorrelation function, ACF *For a stationary process X_t, the autocorrelation function at lag h is defined as*

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = Corr(X_{t+h}, X_t) \text{ for all } t, h \in \mathbb{Z}.$$

Partial correlation function (PACF). *The partial autocorrelation $\alpha(.)$ of a stationary time series is defined by*

$$\alpha(1) = Corr(X_2, X_1) = \rho(1)$$

and

$$\alpha(k) = Corr(X_{k+1} - P(X_{k+1}|1, X_2, ..., X_k), X_1 - P(X_1|1, X_2, ..., X_k)) \;\; fork \geq 2$$

An equivalent definition is that the kth partial autocorrelation $\alpha(k)$ is defined as the last coefficient in the linear projection of Y_t on its k most recent values.

$$Y_t = \alpha_1^{(k)} Y_{t-1} + \alpha_2^{(k)} Y_{t-2} + ... + \alpha_k^{(k)} Y_{t-k} + \varepsilon_t$$

$\alpha(k) = \alpha_k^{(k)}$.

Strict stationarity implies weak stationarity, provided the first and second moments of the variables exist, but the converse of this statement is not true in general.

There is an one important case where both concepts of stationary are equivalent.

Gaussian Time series *The process X_t is a Gaussian time series if and only if the joint density of $(X_{t_1}, ..., X_{t_n})'$ is Gaussian for all $t_1, t_2, ..., t_n$*

If X_t is a stationary Gaussian time series, then it is also strictly stationary, since for all $n = \{1, 2, ...\}$ and for all $h, t_1, t_2, ... \in \mathbb{Z}$, the random vectors $(X_{t_1}, ..., X_{t_n})'$, and $(X_{t_1+h}, ..., X_{t_n+h})'$ have the same mean, and covariance matrix, and hence they have the same distribution.

2.3 Ergodicity

Consider a stationary process X_t, with $E(X_t) = \mu$ for all t. Assume that we are interested in estimating μ. The standard approach for estimating the mean of a single random variable consists of computing its sample mean

$$\bar{X} = (1/N) \sum_{i=1}^{N} X_t^{(i)}$$

(we call this ensemble average) where the X_i's are different realizations of the variable X_t.

When working in a laboratory, one could generate different observations for the variable X_t under identical conditions.

However, when analyzing economic variables over time, we can only observe a unique realization of each of the random variable X_t so that it is not possible to estimate μ by computing the above average.

However, we can compute a time average

$$\bar{X} = (1/T) \sum_{t=1}^{T} X_t$$

Whether time averages converge to the same limit as the ensemble average, $E(X_t)$, hast to do with the concept of Ergodicity.

Ergodicity for the mean *A covariance stationary process* X_t *is said to be ergodic for the mean if* $\bar{X} = (1/T)\sum_{t=1}^{T} X_t$ *converges in probability to* $E(X_t)$ *as* T *gets large.*

Ergodicity for the second moments *A covariance stationary process is said to be ergodic for the second moments if*

$$[1/(T-j)] \sum_{t=j+1}^{T} (X_t - \mu)(X_{t-j} - \mu) \xrightarrow{p} \gamma_j$$

Important result: A sufficient condition for ergodicity in mean of a stationary process is that $\sum_{h=0}^{\infty} |\gamma(h)| < \infty$. If the process is also Gaussian then the above condition also implies ergodicity for all the higher moments.

For many applications, ergodicity and stationarity turn out to amount for the same requirements. However, we present now an example of a stationary process that is not ergodic.

Example Consider the following process. $y_0 = u_0$ with $u_0 \sim N(0, \sigma^2)$ and $y_t = y_{t-1}$ for $t = 1, 2, 3, ...$ It easy to see that the process is stationary. In fact

$$E(y_t) = 0,\ E(y_t^2) = \sigma^2,\ E(y_{t-j}, y_{t-i}) = \sigma^2, i \neq j$$

However

$$(1/T)\sum_{t=1}^{T} y_t = (1/T)\sum_{t=1}^{T} u_1 = u_1 \neq 0$$

2.4 Some processes

iid sequences The sequence ε_t is i.i.d with zero mean and variance σ^2 , written $\varepsilon \sim iid(0, \sigma^2)$, (independent and identically distributed) if all the variables are independent and share the same univariate distribution.

White noise The process ε_t is called white noise, written $\varepsilon \sim WN(0, \sigma^2)$ if it is weakly stationary with $E(\varepsilon) = 0$ and autocovariance function $\gamma(0) = \sigma^2$ and $\gamma(h) = 0$ for $h \neq 0$.

Note that an i.i.d sequence with zero mean and variance σ^2 is also white noise. The converse is not true in general.

Martingale difference sequence, m.d.s. A process ε_t, with $E(\varepsilon_t) = 0$ is called a martingale difference sequence if

$$E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2} ...) = 0, \quad t = 2, 3, ...$$

Random Walk Consider the process

$$y_t = y_{t-1} + u_t, \quad t = 0, 1, ...$$

where u is a $WN(0, \sigma^2)$ and y_0 fixed. Substituting backward

$$y_t = y_0 + u_1 + u_2 + ... + u_{t-2} + u_{t-1} + u_t$$

It is easy to see that $E(y_t) = y_0$. Moreover the variance is

$$\gamma(0) = E\left(\sum_{j=1}^{t} u_j\right)^2 = \left(\sum_{j=1}^{t} E(u_j^2)\right) = t\sigma^2$$

and

$$\begin{aligned} \gamma(h) &= E\left(\sum_{j=1}^{t} u_j \sum_{k=1}^{t-h} u_j\right) \\ &= \sum_{k=1}^{t-h} E(u_k^2) \\ &= (t-h)\sigma^2 \end{aligned}$$

The autocorrelation is

$$\begin{aligned} \rho(h) &= \frac{(t-h)\sigma^2}{(t\sigma^2(t-h)\sigma^2)^{1/2}} \\ &= \left(1 - \frac{h}{t}\right)^{1/2} \end{aligned} \tag{2}$$

3 Linear projections

Let X_t be a $(k+1) \times 1$ vector of variables, with non-singular variance-covariance matrix and let Y_t be a variable. Consider the linear functions of X_t

$$P(Y_t|X_{1t}, ..., X_{kt}) = \beta' X_t \tag{3}$$

satisfying

$$E[(Y_t - \beta' X_t) X_t'] = 0$$

Then (3) is the linear projection of Y_t onto X_t and β is the projection coefficient satisfying

$$\beta' = E(Y_t X_t')[E(X_t X_t')]^{-1}$$

The linear projection has the following properties

1. If $E(Y|X) = X\beta$, then $E(Y|X) = P(Y|X)$.

2. $P(aZ + bY|X) = aP(Z|X) + bP(Y|X)$

4. $P(Y|X) = P(P(Y|X, Z)|X)$

4 Moments estimation

The sample mean is the natural estimator for the expected value of a process.

The sample mean of $Y_1, Y_2, ..., Y_T$ is defined as

$$\bar{Y}_T = (1/T) \sum_{t=1}^{T} Y_t$$

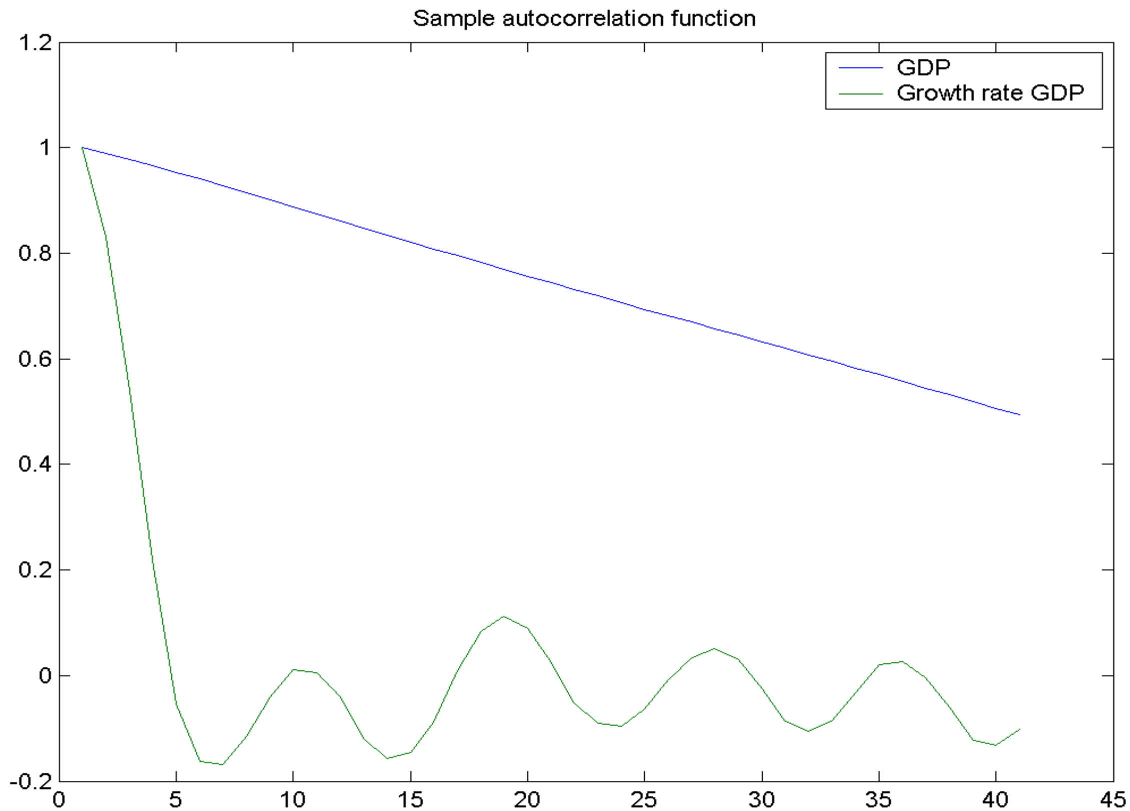
The sample autocovariance of $Y_1, Y_2, ..., Y_T$ is defined as

$$\hat{\gamma}(j) = (1/T) \sum_{t=j+1}^{T} (Y_t - \bar{Y}_T)(Y_{t-j} - \bar{Y}_T)$$

The sample autocorrelation of $Y_1, Y_2, ..., Y_T$ is defined as

$$\hat{\rho}(j) = \frac{\hat{\gamma}(j)}{\hat{\gamma}(0)}$$

The sample autocovariance and autocorrelation can be computed for any data set, and are not restricted to realizations of a stationary process. For stationary processes both functions will show a rapid decay towards zero as h increases. However, for non-stationary data, these functions will exhibit quite different behavior. For instance for variables with trend, the autocorrelation reduces very slowly.


Sample autocorrelation function
GDP
Growth rate GDP
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
0
5
10
15
20
25
30
35
40
45

To compute the k-th partial correlation, one simply has to run an OLS regression including the most recent m values of the variable. The last coefficient would be the k-th. autocorrelation, that is,

$$Y_t = \hat{\alpha}_1^{(k)} Y_{t-1} + \hat{\alpha}_2^{(k)} Y_{t-2} + ... + \hat{\alpha}_k^{(k)} Y_{t-k} + \hat{\varepsilon}_t$$

where $\hat{\varepsilon}_t$ denotes OLS residuals, so that $\hat{\alpha}(k) = \hat{\alpha}_k^(k)$.

2. ARMA[1]

[1]This part is based on Hamilton textbook.

1 MA

1.1 MA(1)

Let ε_t be WN with variance σ^2 and consider the zero mean[2] process

$$Y_t = \varepsilon_t + \theta\varepsilon_{t-1} \tag{1}$$

where θ is a constant. This time series is called *first-order moving average process* denote by MA(1).

1.1.1 Moments

The expectation of Y_t is given by

$$\begin{aligned} E(Y_t) &= E(\varepsilon_t + \theta\varepsilon_{t-1}) \\ &= E(\varepsilon_t) + \theta E(\varepsilon_{t-1}) \\ &= 0 \end{aligned} \tag{2}$$

Clearly with a constant term μ in (10) the expectation would be μ.

[2] All we see works even for non zero mean processes.

The variance is given by

$$\begin{aligned} E(Y_t)^2 &= E(\varepsilon_t + \theta\varepsilon_{t-1})^2 \\ &= E(\varepsilon_t)^2 + \theta^2 E(\varepsilon_{t-1})^2 + 2\theta E(\varepsilon_t\varepsilon_{t-1}) \\ &= \sigma^2 + \theta^2\sigma^2 \\ &= (1+\theta^2)\sigma^2 \end{aligned} \tag{3}$$

The first autocovariance is

$$\begin{aligned} E(Y_tY_{t-1}) &= E(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2}) \\ &= E(\varepsilon_t\varepsilon_{t-1}) + E(\theta\varepsilon_t\varepsilon_{t-2}) + E(\theta\varepsilon_{t-1}^2) + E(\theta^2\varepsilon_{t-1}\varepsilon_{t-2}) \\ &= \theta\sigma^2 \end{aligned}$$

Higher autocovariances are all zero, $E(Y_tY_{t-j}) = 0$ for $j > 1$.

Since the mean and covariances are not functions of time the process is *stationary* regardless on the value of θ. Moreover the process is also ergodic for the first moment since $\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$. If ε_t is also Gaussian then the process is ergodic for all moments.

The j-th autocorrelation is

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{(1+\theta^2)}$$

for $j = 1$ and zero for $j > 1$.

Figure 1 displays the autocorrelation functions for

$$Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}$$

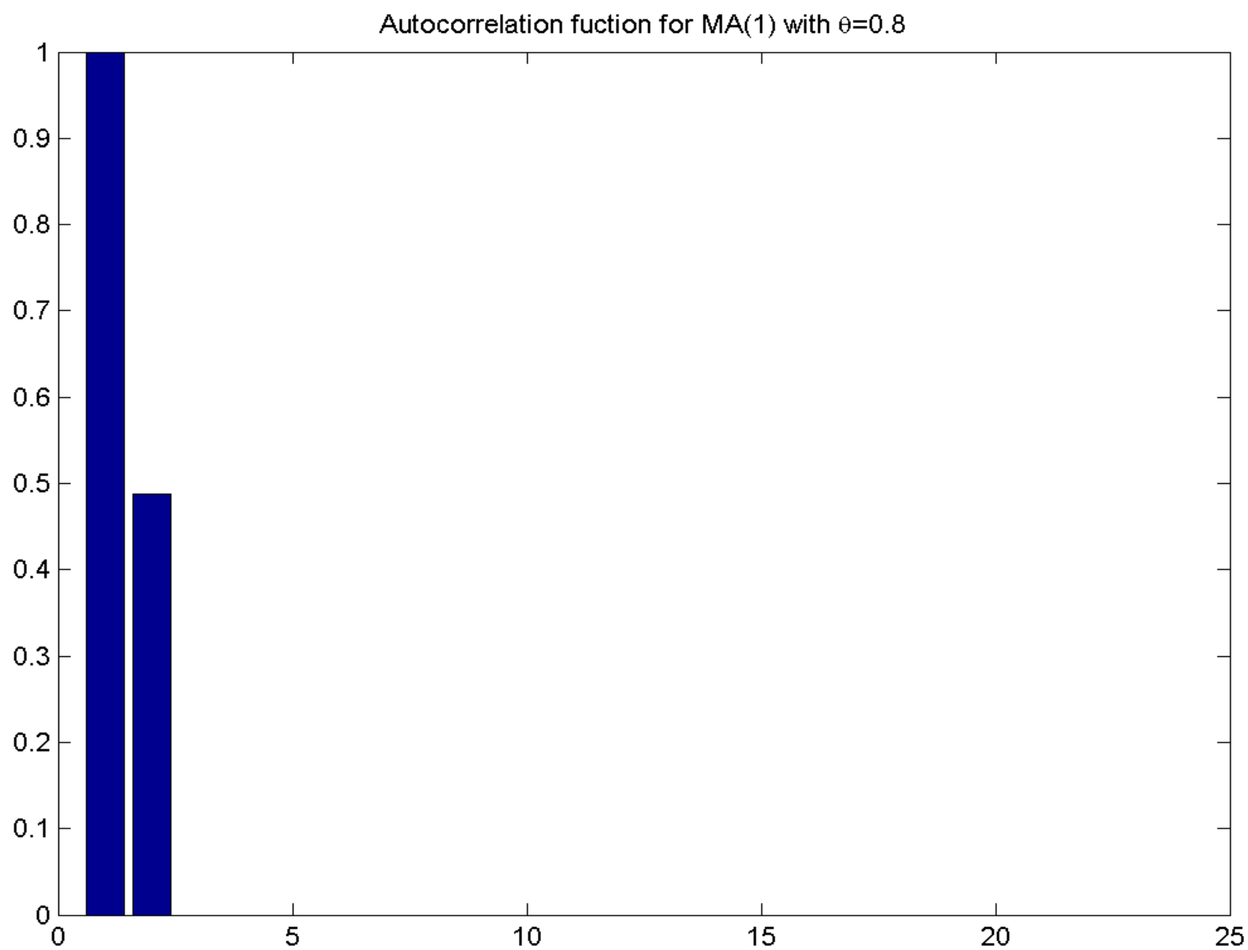


Figure 1

Note that the first order autocorrelation can be plotted as a function of θ as in Figure 2.

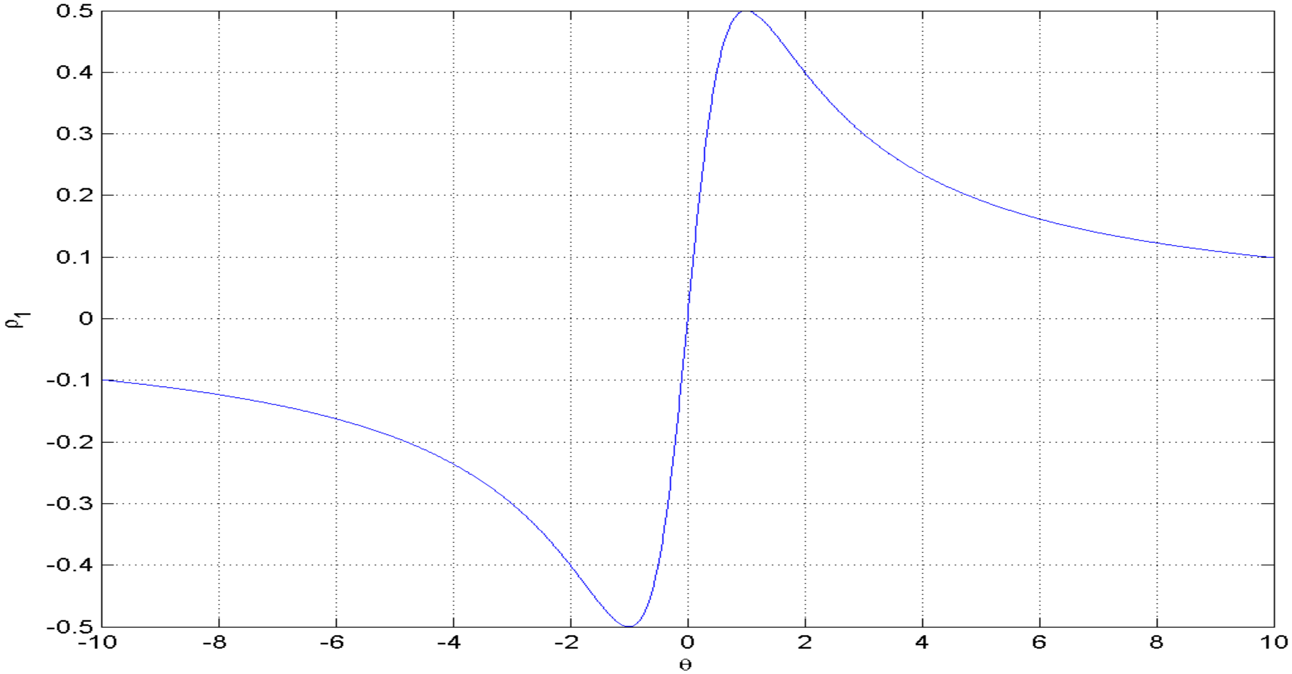


Figure 2

Note that:

1. positive value of θ induce positive autocorrelation, while negative value negative autocorrelations.
2. the largest possible value is 0.5 ($\theta = 1$) and the smallest one is -0.5 ($\theta = -1$)
3. for any value of ρ_1 between $[-0.5, 0.5]$ there are two values of θ that produce the same autocorrelation because ρ_1 is unchanged if we replace θ with $1/\theta$

$$\frac{1/\theta}{1 + 1/\theta^2} = \frac{\theta^2(1/\theta)}{\theta^2(1 + 1/\theta^2)} = \frac{\theta}{\theta^2 + 1} \tag{4}$$

So the processes

$$Y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$$

and

$$Y_t = \varepsilon_t + 2\varepsilon_{t-1}$$

generate the same autocorrelation functions.

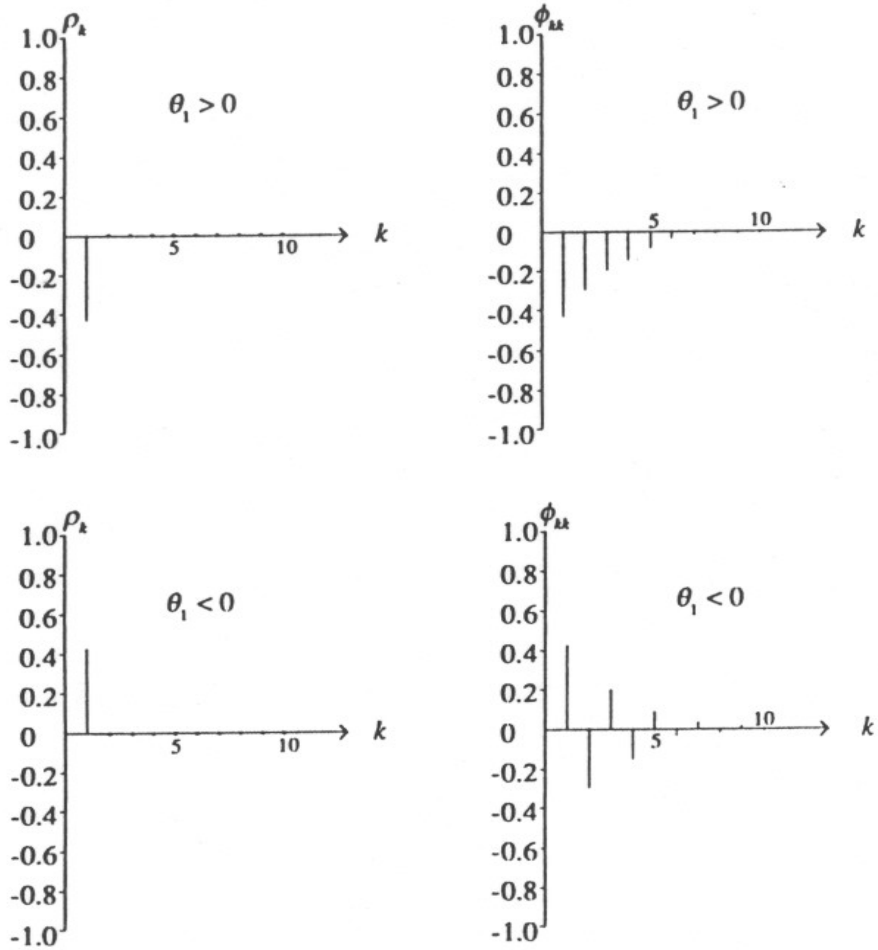


Fig. 3.10 ACF and PACF of MA(1) processes: $\dot{Z}_t = (1 - \theta B)a_t$.

Figure 3. Source: W.Wei ”Time Series Analysis: Univariate and Multivariate Methods”.

1.2 MA(q)

A *q-th order moving average process*, denoted MA(q), is characterized by

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} \tag{5}$$

where ε_t is WN and the θ_i's are any real number.

1.2.1 Moments

- The mean of (5) is

$$\begin{aligned} E(Y_t) &= E(\varepsilon_t) + \theta_1 E(\varepsilon_{t-1}) + \theta_2 E(\varepsilon_{t-2}) + ... + \theta_q E(\varepsilon_{t-q}) \\ &= 0 \end{aligned}$$

- The variance of (5) is

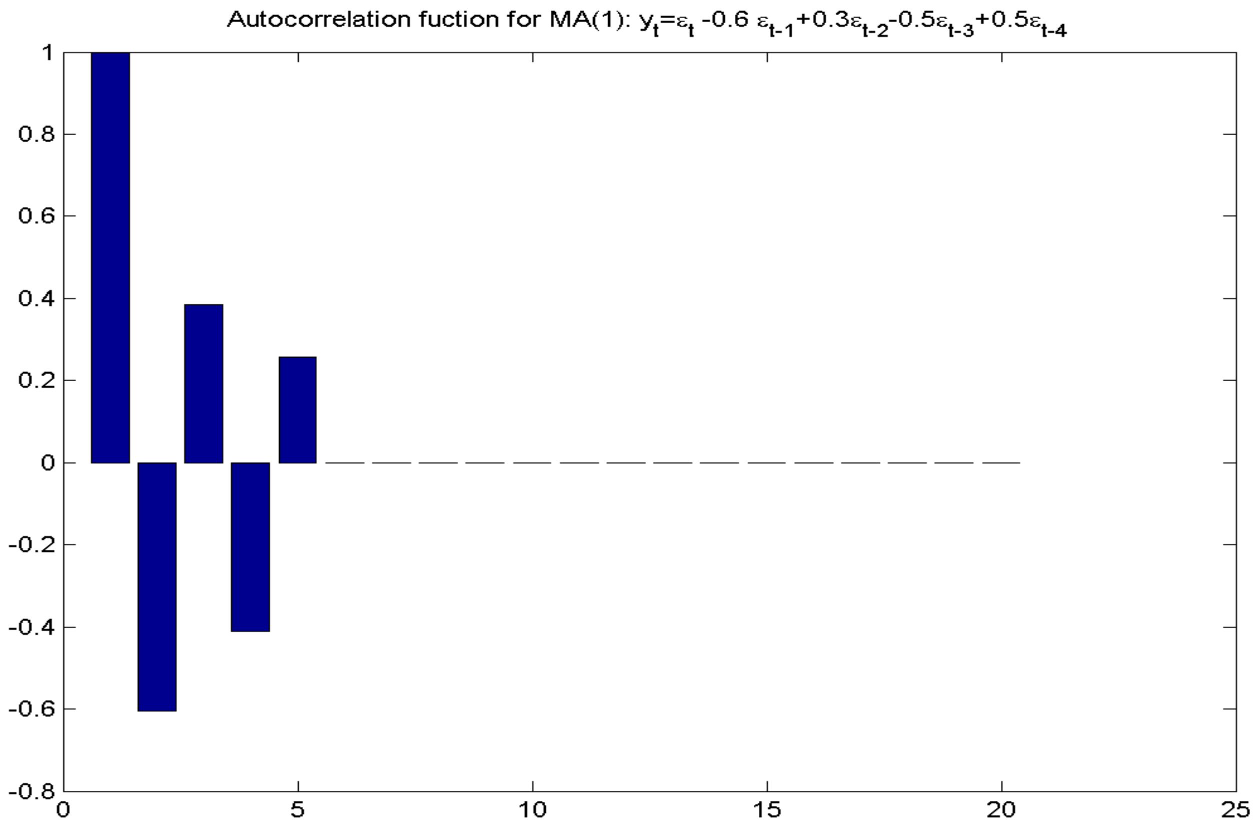
$$\begin{aligned} \gamma_0 = E(Y_t^2) &= E(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q})^2 \\ &= (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma^2 \end{aligned}$$

because all the terms involving the expected value of different ε_j's are zero because of the WN assumption.

- The autocovariances are

$$\gamma_j = (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 ... + \theta_q \theta_{q-j})\sigma^2 \quad j = 1, 2, ..., q$$

and zero for $j > q$.



Autocorrelation fuction for MA(1): y_t=ε_t -0.6 ε_{t-1}+0.3ε_{t-2}-0.5ε_{t-3}+0.5ε_{t-4}
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0
5
10
15
20
25

Figure 3

Example Consider an MA(2).

$$\begin{aligned}
\gamma_0 &= (1+\theta_1^2+\theta_2^2)\sigma^2 \\
\gamma_1 &= E(Y_t Y_{t-1}) \\
&= E(\varepsilon_t\varepsilon_{t-1}) + E(\theta_1\varepsilon_t\varepsilon_{t-2}) + E(\theta_2\varepsilon_t\varepsilon_{t-3} + E(\theta_1\varepsilon_{t-1}^2) + E(\theta_1^2\varepsilon_{t-1}\varepsilon_{t-2}) + \\
&\quad +E(\theta_1\theta_2\varepsilon_{t-1}\varepsilon_{t-3}) + E(\theta_2\varepsilon_{t-2}\varepsilon_{t-1}) + E(\theta_2\theta_1\varepsilon_{t-2}^2) + E(\theta_2^2\varepsilon_{t-2}\varepsilon_{t-3}) \\
&= (\theta_1+\theta_2\theta_1)\sigma^2 \\
\gamma_2 &= E(Y_t Y_{t-2}) \\
&= \theta_2 E(\varepsilon_{t-2}^2) \\
&= \theta_2\sigma^2
\end{aligned} \tag{6}$$

The MA(q) process is covariance stationary for any value of θ_i. Moreover the process is also ergodic for the first moment since $\sum_{j=-\infty}^{\infty}|\gamma_j| < \infty$. If ε_t is also Gaussian then the process is ergodic for all moments.

1.3 MA(∞)

The MA(∞) can be thought as the limit of a MA(q) process for $q \to \infty$

$$Y_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} = \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... \tag{7}$$

with absolute summable coefficients $\sum_{j=0}^{\infty} |\theta_j| < \infty$.

If the MA coefficients are square summable (implied by absolute summability), i.e. $\sum_{j=0}^{\infty} \theta_j^2 < \infty$ then the above infinite sum generates a mean square convergent random variable.

1.3.1 Moments

- The mean of the process is

$$E(Y_t) = \sum_{j=0}^{\infty} \theta_j E(\varepsilon_{t-j}) = 0 \tag{8}$$

- The variance is

$$\begin{aligned}\gamma_0 = E(Y_t^2) &= E(\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j})^2 \\ &= E(\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + ...)^2 \\ &= (\theta_1^2 + \theta_2^2 + \theta_3^2 + ...)\sigma^2\end{aligned}$$

$$= \sigma^2 \sum_{j=0}^{\infty} \theta_j^2 \tag{9}$$

Again all the terms involving the expected value of different ε_j's are zero because of the WN assumption.

- Autocovariances are

$$\begin{aligned}
\gamma_j = E(Y_t Y_{t-j}) &= E(\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + ...)(\theta_1 \varepsilon_{t-j-1} + \theta_2 \varepsilon_{t-j-2} + \theta_3 \varepsilon_{t-j-3} + ...) \\
&= (\theta_j \theta_0 + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \theta_{j+3}\theta_3 + ...)\sigma^2
\end{aligned}$$

The process is stationary finite and constant first and second moments.

Moreover an MA(∞) with absolutely summable coefficients has absolutely summable autocovariances, $\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$, so it is ergodic for the mean.

the ε's are Gaussian is ergodic for all moments.

1.4 Invertibility and Fundamentalness

Consider an MA(1)

$$Y_t = \varepsilon_t + \theta\varepsilon_{t-1} \tag{10}$$

where ε_t is $WN(0, \sigma^2)$. Provided that $|\theta| < 1$ both sides can be multiplied to obtain

$$(1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + ...)Y_t = \varepsilon_t$$

which could be viewed as a AR(∞) representation.

If a moving average representation can be rewritten in terms as an AR(∞) representation by inverting $(1 + \theta L)$, then the moving average representation is said to be *invertible*. For an MA(1) invertibility requires $|\theta| < 1$.

Let us investigate what invertibility means in terms of the first and second moments of the process. Consider the following MA(1)

$$\tilde{Y}_t = (1 + \tilde{\theta}L)\tilde{\varepsilon}_t \tag{11}$$

where ε_t is $WN(0, \tilde{\sigma}^2)$. Moreover suppose that the parameters in this new MA(1) are related to the other as follows:

$$\begin{aligned}\tilde{\theta} &= \theta^{-1}\\ \tilde{\sigma}^2 &= \theta^2\sigma^2\end{aligned}$$

Let us derive the first two moments of the two processes. $E(Y_t) = E(\tilde{Y}_t) = 0$. For Y_t

$$\begin{aligned} E(Y_t^2) &= \sigma^2(1+\theta^2) \\ E(Y_t Y_{t-1}) &= \theta\sigma^2 \end{aligned}$$

For $\tilde{Y}_t$

$$\begin{aligned} E(Y_t^2) &= \tilde{\sigma}^2(1+\tilde{\theta}^2) \\ E(Y_t Y_{t-1}) &= \tilde{\theta}\tilde{\sigma}^2 \end{aligned}$$

However note that given the above restrictions

$$\begin{aligned} (1+\theta^2)\sigma^2 &= \left(1+\frac{1}{\tilde{\theta}^2}\right)\tilde{\theta}^2\tilde{\sigma}^2 \\ &= \left(\frac{\tilde{\theta}^2+1}{\tilde{\theta}^2}\right)\tilde{\theta}^2\tilde{\sigma}^2 \\ &= \left(\tilde{\theta}^2+1\right)\tilde{\sigma}^2 \end{aligned}$$

and

$$\begin{aligned} \theta\sigma^2 &= \frac{\tilde{\theta}^2\tilde{\sigma}^2}{\tilde{\theta}} \\ &= \tilde{\theta}\tilde{\sigma}^2 \end{aligned}$$

That is the first two moments of the two processes are identical. Note that if $|\theta| < 1$ then $|\tilde{\theta}| > 1$. In other words for any invertible MA representation we can find a non invertible representation with identical first and second moments. The two process share the same autocovariance generating function.

The value of ε_t associated with the invertible representation is sometimes called the *fundamental innovation* for Y_t. For the borderline case $|\theta| = 1$ the process is non invertible but still fundamental.

Here we give a formal definition of invertibility

Invertibility A MA(q) process defined by the equation $Y_t = \theta(L)\varepsilon_t$ is said to be invertible is there exists a sequence of constants $\{\pi_j\}_{j=0}^{\infty}$ such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and $\sum_{j=0}^{\infty} \pi_j Y_{t-j} = \varepsilon_t$.

Proposition A MA process defined by the equation $Y_t = \theta(L)\varepsilon_t$ is invertible if and only if $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.

A similar concept is that of fundamentalness defined below.

Fundamentalness A MA $Y_t = \theta(L)\varepsilon_t$ is fundamental if and only if $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| < 1$.

1.5 Wold's decomposition theorem

Here is a very powerful result known as Wold'd Decomposition theorem.

Theorem Any zero-mean covariance stationary process Y_t can be represented in the form

$$Y_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} + k_t \tag{12}$$

where:

1. $\theta_0 = 1$,
2. $\sum_{j=0}^{\infty} \theta_j^2 < \infty$,
3. ε_t is the the error made in forecasting Y_t on the basis of a linear function of lagged Y_t (fundamental innovation),
4. the value of k_t is uncorrelated with ε_{t-j} for any j and can be perfectly predicted from a linear function of the past values of Y.

The term k_t is called the *linearly deterministic* component of Y_t. If $k_t = 0$ then the process is called purely non-deterministic.

The result is very powerful since holds for any covariance stationary process.

However the theorem does not implies that (12) is the *true* representation of the process.

- For instance the process could be stationary but non-linear or non-invertible. If the true system is generated by a nonlinear difference equation $Y_t = g(Y_{t-1}, ..., Y_{t-1}) + \eta_t$, obviously, when we fit a linear approximation, as in the Wold theorem, the shock we recover ε will be different from η_t.

- If the model is non invertible then the true shock will not be the Wold shock.

2 AR

2.1 AR(1)

A *first-order autoregression*, denoted AR(1), satisfies the following difference equation:

$$Y_t = \phi Y_{t-1} + \varepsilon_t \tag{13}$$

where again ε_t is a WN. When $|\phi| < 1$, the solution to (13) is

$$Y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + ... \tag{14}$$

(14) can be viewed as an MA(∞) with $\psi_j = \phi^j$. When $|\phi| < 1$ the autocovariances are absolutely summable since the MA coefficients are absolutely summable

$$\sum_{j=0}^{\infty} |\psi_j| = \sum_{j=0}^{\infty} |\phi|^j = \frac{1}{(1 - |\phi|)}$$

This ensures that the MA representation exists, the process is stationary and ergodic in mean.

Recall that $\sum_{j=0}^{\infty} \phi^j$ is a geometric series converging to $1/(1-\phi)$ if $|\phi| < 1$.

2.2 Moments

- The *mean* is

$$E(Y_t) = 0$$

- The *variance* is

$$\gamma_0 = E(Y_t^2) = \sigma^2 \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma^2}{1-\phi^2}$$

- The *jth autocovariance* is

$$\gamma_j = E(Y_t Y_{t-j}) = \frac{\phi^j \sigma^2}{1-\phi^2}$$

- The *jth autocorrelation* is

$$\rho_j = \phi^j$$

To find the moments of the AR(1) we can use a different strategy by directly working with the AR representation and the assumption of stationarity.

- Note that the *mean*

$$E(Y_t) = \phi E(Y_{t-1}) + E(\varepsilon_t)$$

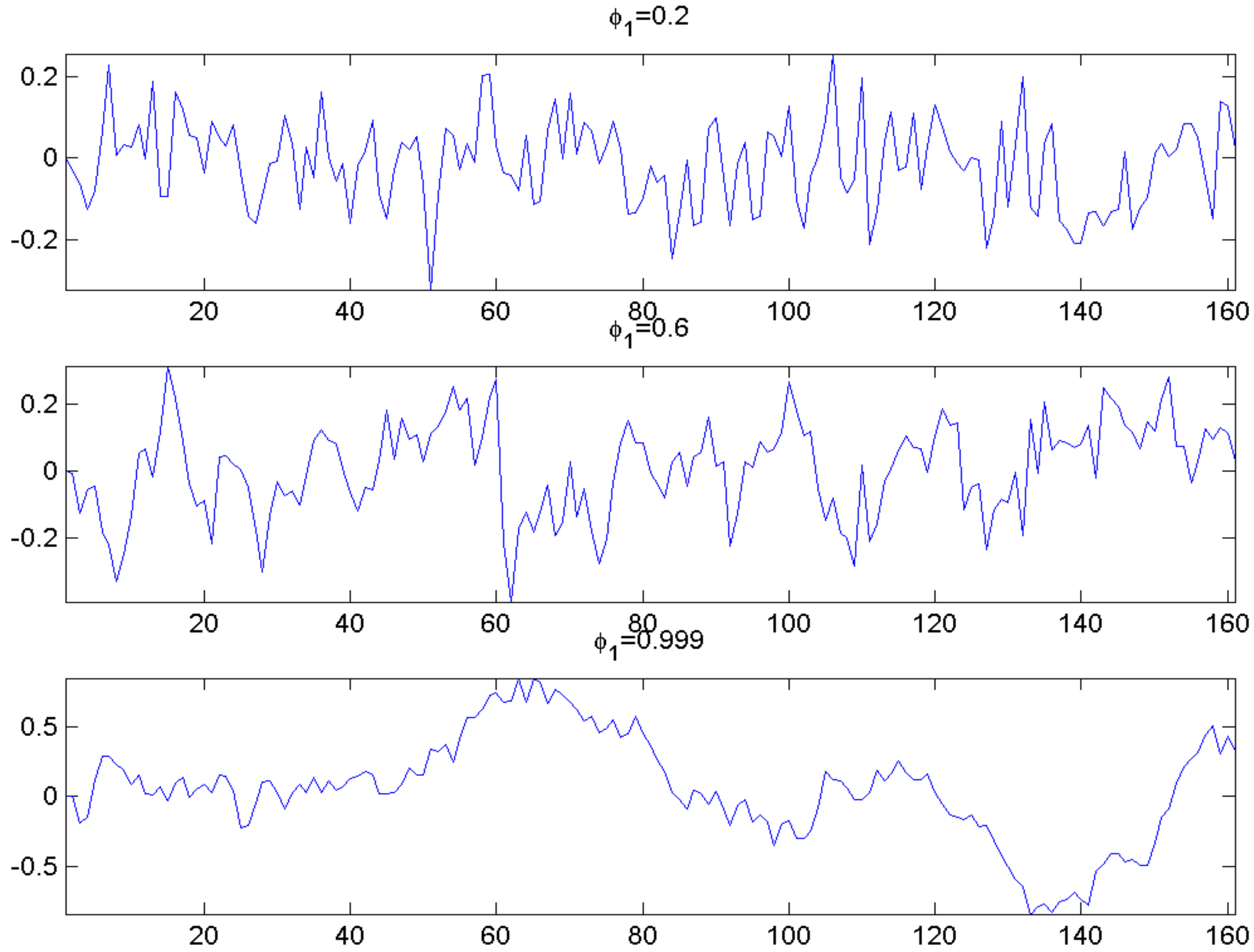
given the stationarity assumption $E(Y_t) = E(Y_{t-1})$ and therefore $(1-\phi)E(Y_t) = 0$.

- The *jth autocovariance* is

$$\begin{aligned}\gamma_j = E(Y_t Y_{t-j}) &= \phi E(Y_{t-1}Y_{t-j}) + E(\varepsilon_t Y_{t-j}) \\ &= \phi \gamma_{j-1} \\ &= \phi^j \gamma_0\end{aligned}$$

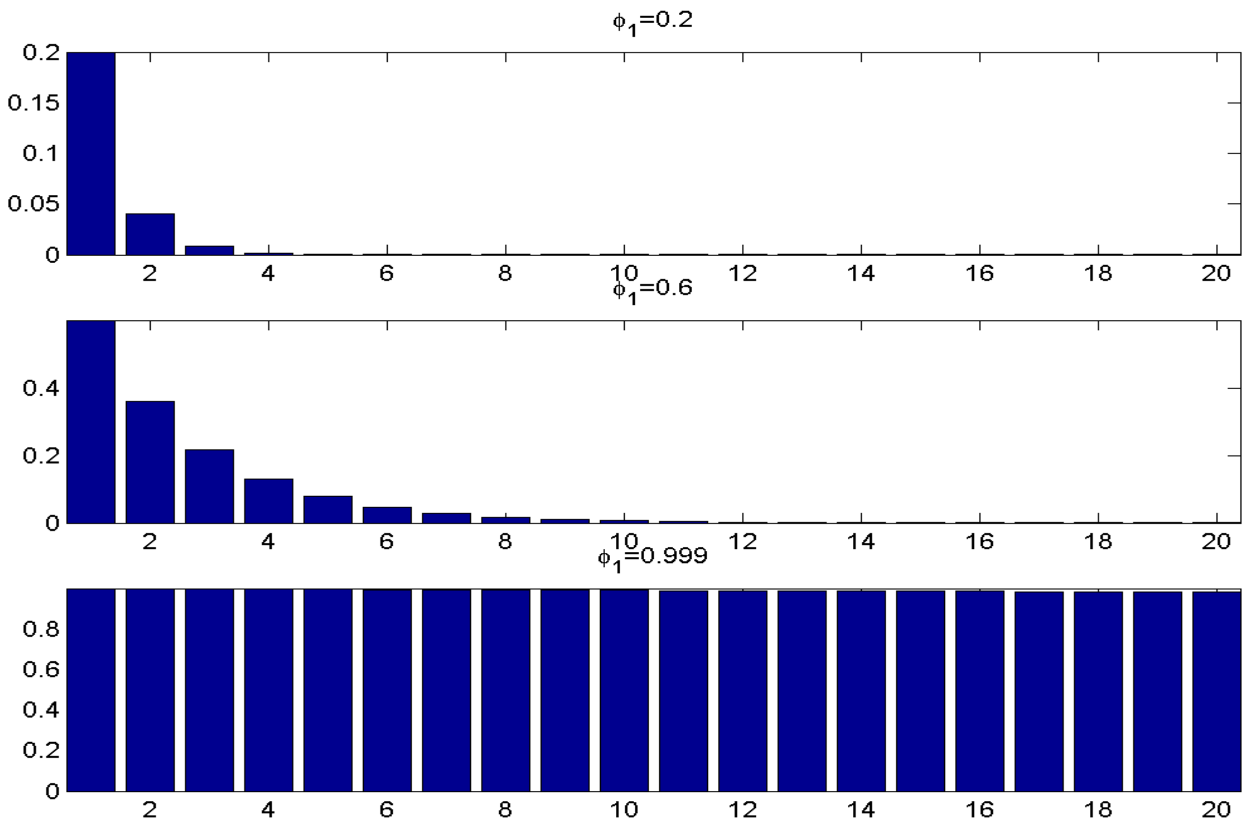
- Similarly the *variance*

$$\begin{aligned}E(Y_t^2) &= \phi E(Y_{t-1}Y_t) + E(\varepsilon_t Y_t) \\ &= \phi E(Y_{t-1}Y_t) + \sigma^2 \\ \gamma_0 &= \phi^2 \gamma_0 + \sigma^2 \\ &= \sigma^2/(1-\phi^2)\end{aligned}$$



φ1=0.2
0.2
0
-0.2
20
40
60
80
100
120
140
160
φ1=0.6
0.2
0
-0.2
20
40
60
80
100
120
140
160
φ1=0.999
0.5
0
-0.5
20
40
60
80
100
120
140
160

Figure 4



ϕ_1=0.2
0.2
0.15
0.1
0.05
0
2
4
6
8
10
12
14
16
18
20
ϕ_1=0.6
0.4
0.2
0
ϕ_1=0.999
0.8
0.6
0.4
0.2
0

Figure 5

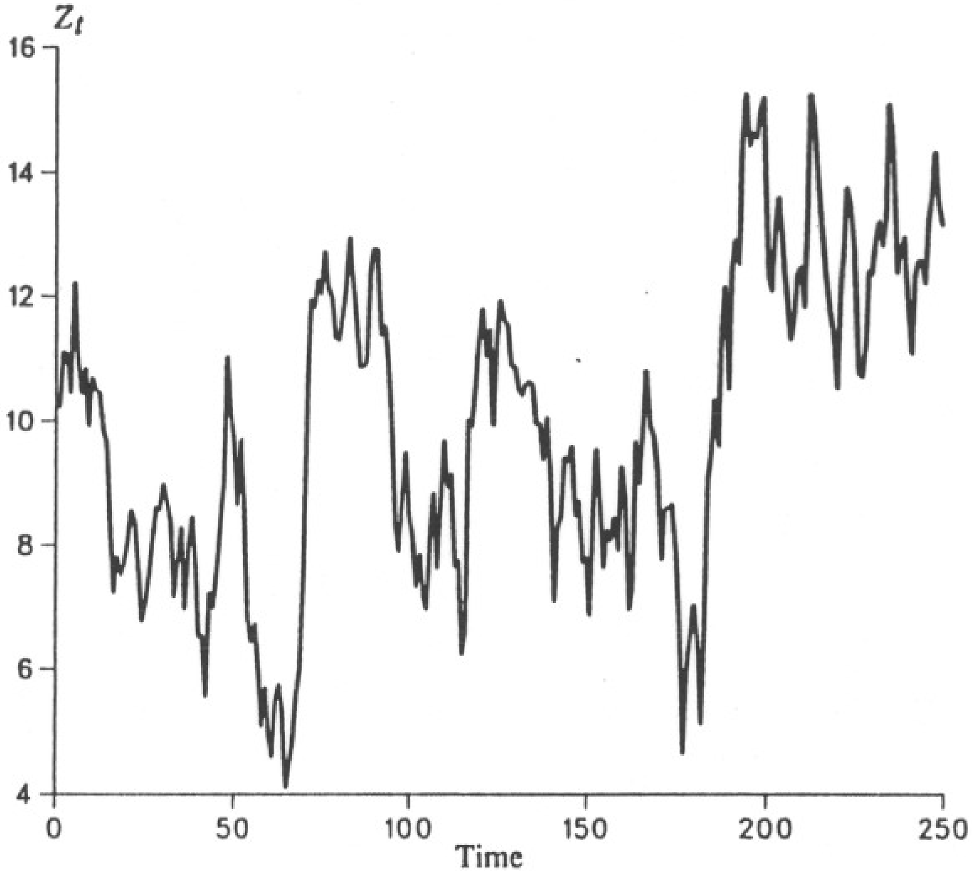


Fig. 3.2 A simulated AR(1) series, $(1 - .9B)(Z_t - 10) = a_t$.

Figure . Source: W.Wei "Time Series Analysis: Univariate and Multivariate Methods".

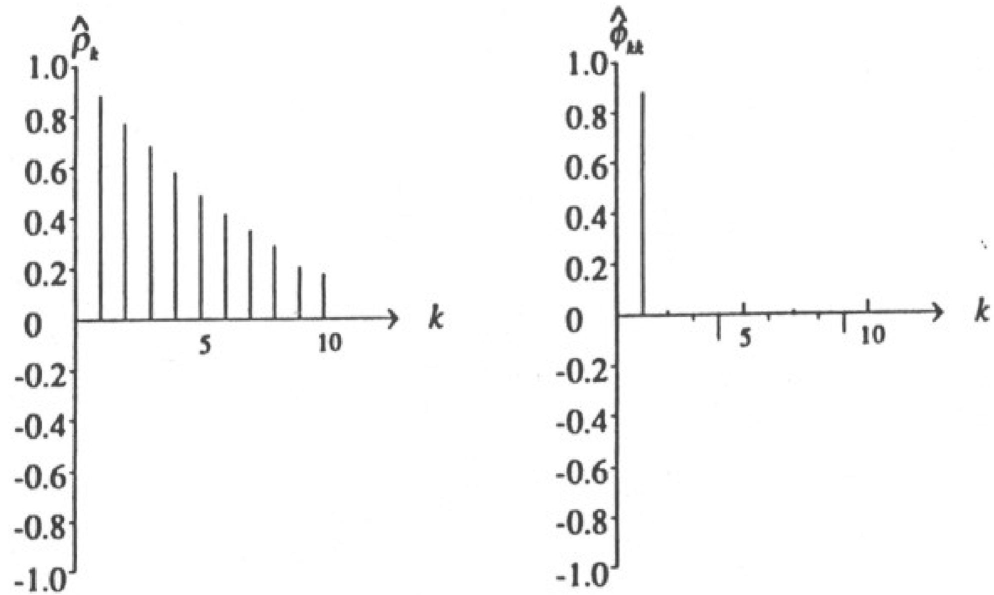


Fig. 3.3 Sample ACF and sample PACF of a simulated AR(1) series: $(1 - .9B)(Z_t - 10) = a_t$.

Figure . Source: W.Wei "Time Series Analysis: Univariate and Multivariate Methods".

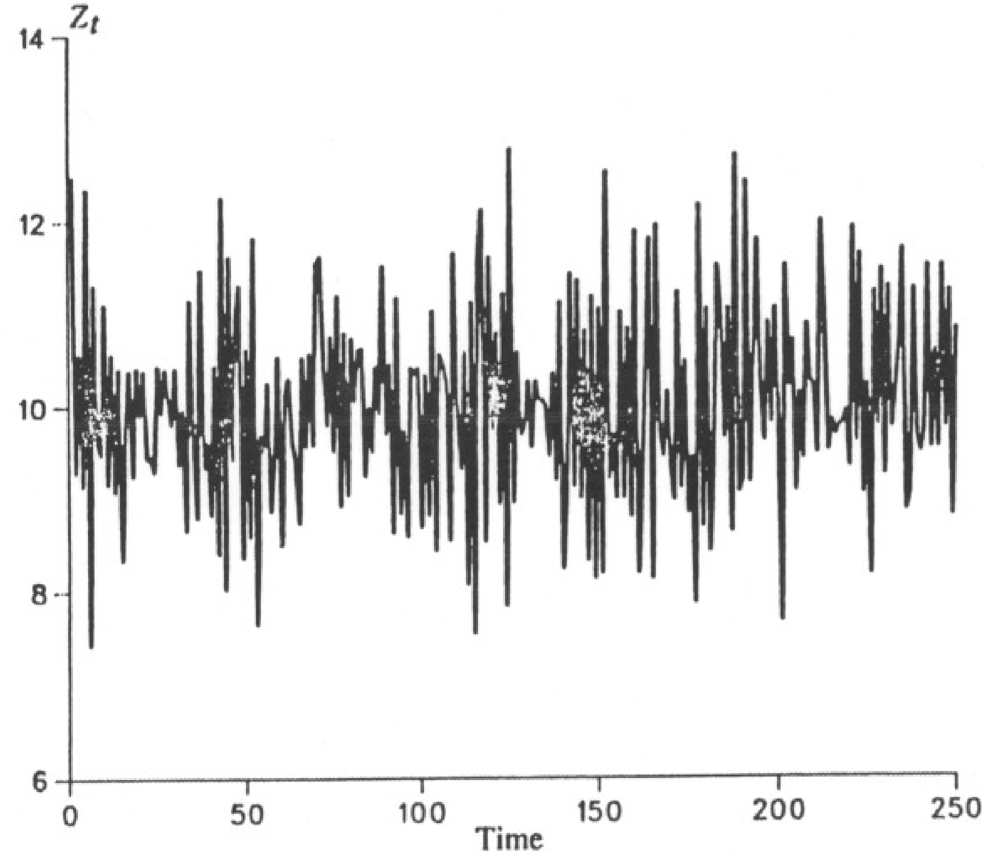


Fig. 3.4 A simulated AR(1) series $(1 + .65B)(Z_t - 10) = a_t$.

Figure . Source: W.Wei "Time Series Analysis: Univariate and Multivariate Methods".

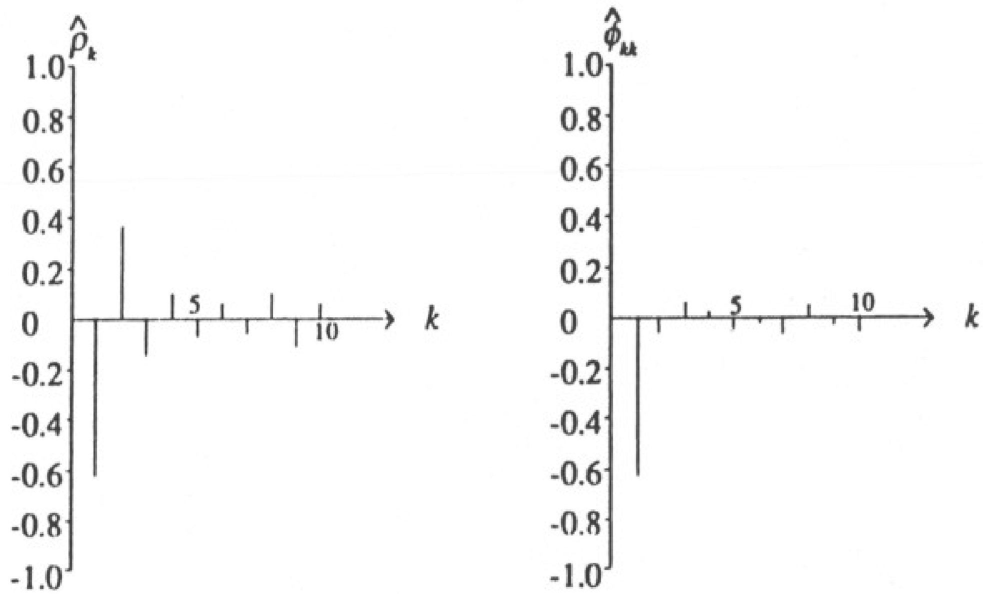


Fig. 3.5 Sample ACF and sample PACF of a simulated AR(1) series $(1 + .65B)(Z_t - 10) = a_t$.

Figure . Source: W.Wei "Time Series Analysis: Univariate and Multivariate Methods".

2.3 AR(2)

A *second-order autoregression*, denoted AR(2), satisfies

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{15}$$

where again ε_t is a WN. Using the lag operator

$$(1 - \phi_1 L - \phi_2 L^2) Y_t = \varepsilon_t \tag{16}$$

The difference equation is stable provided that the roots of the polynomial

$$1 - \phi_1 z - \phi_2 z^2$$

lie outside the unit circle. When this condition is satisfied the process is covariance stationary and the inverse of the autoregressive operator is

$$\psi(L) = (1 - \phi_1 L - \phi_2 L^2)^{-1} = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + ...$$

with $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

2.3.1 Moments

To find the moments of the AR(2) we can proceed as before.

- The *mean* is

$$E(Y_t) = \phi_1 E(Y_{t-1}) + \phi_2 E(Y_{t-2}) + E(\varepsilon_t)$$

again by stationarity $E(Y_t) = E(Y_{t-j})$ and therefore $(1 - \phi_1 - \phi_2)E(Y_t) = 0$.

- The *jth autocovariance* is given by

$$\begin{aligned}\gamma_j = E(Y_t Y_{t-j}) &= \phi_1 E(Y_{t-1} Y_{t-j}) + \phi_2 E(Y_{t-2} Y_{t-j}) + E(\varepsilon_t Y_{t-j}) \\ &= \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2}\end{aligned}$$

- Similarly the *jth autocorrelation* is

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2}$$

In particular setting $j = 1$

$$\begin{aligned}\rho_1 &= \phi_1 + \phi_2 \rho_1 \\ \rho_1 &= \phi_1/(1 - \phi_2)\end{aligned}$$

and setting $j = 2$,

$$\rho_2 = \phi_1 \rho_1 + \phi_2$$

$$\rho_2 = \phi_1 \frac{\phi_1}{1-\phi_2} + \phi_2$$

- The variance

$$E(Y_t^2) = \phi_1 E(Y_{t-1}Y_t) + \phi_2 E(Y_{t-2}Y_t) + E(\varepsilon_t Y_t)$$

The equation can be written as

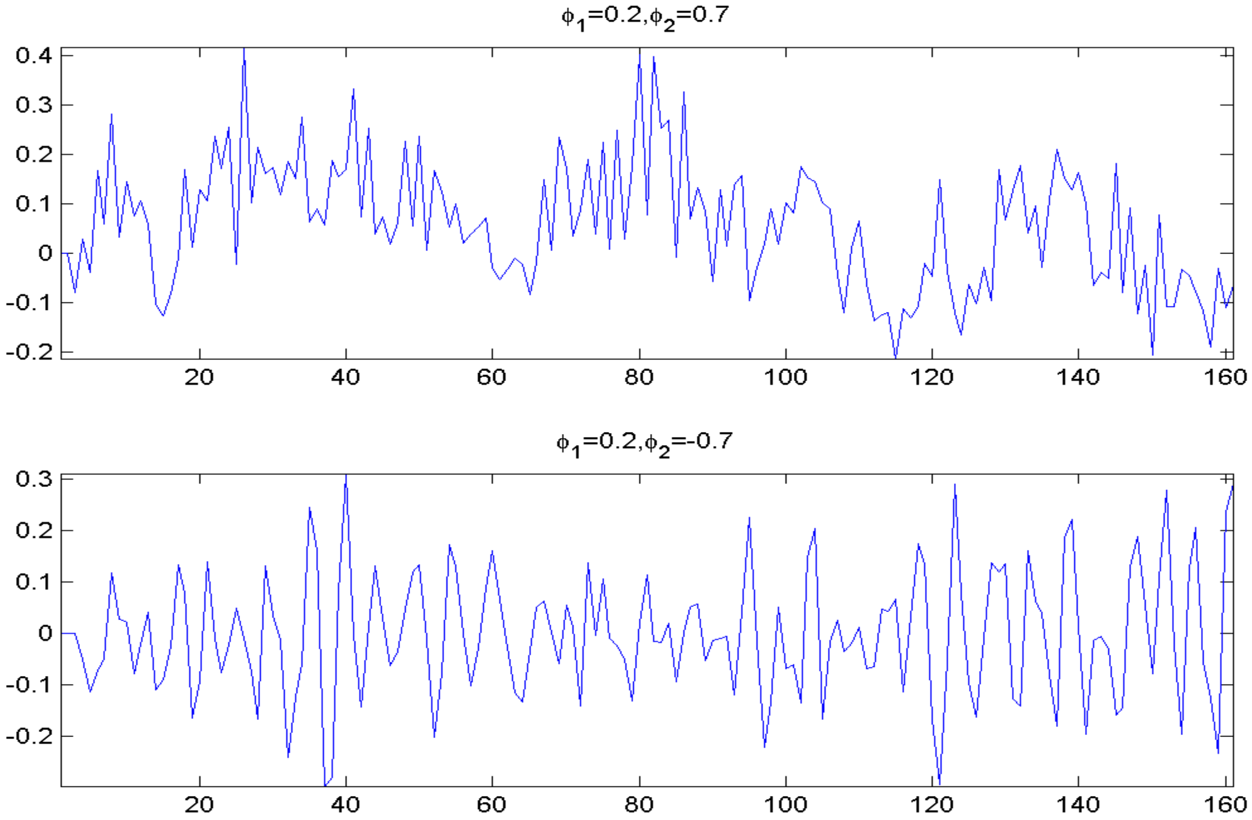
$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma^2$$

where the last term comes from the fact that

$$E(\varepsilon_t Y_t) = \phi_1 E(\varepsilon_t Y_{t-1}) + \phi_2 E(\varepsilon_t Y_{t-2}) + E(\varepsilon_t^2)$$

and that $E(\varepsilon_t Y_{t-1}) = E(\varepsilon_t Y_{t-2}) = 0$. Note that $\gamma_j/\gamma_0 = \rho_j$. Therefore the variance can be rewritten as

$$\begin{aligned}
\gamma_0 &= \phi\rho_1\gamma_0 + \phi_2\rho_2\gamma_0 + \sigma^2 \\
&= (\phi_1\rho_1 + \phi_2\rho_2)\gamma_0 + \sigma^2 \\
&= \left[\frac{\phi_1^2}{(1-\phi_2)} + \frac{\phi_2\phi_1^2}{(1-\phi_2)} + \phi_2^2\right]\gamma_0 + \sigma^2 \\
\gamma_0 &= \frac{(1-\phi_2)\sigma^2}{(1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]} \qquad (17)
\end{aligned}$$



φ1=0.2,φ2=0.7
0.4
0.3
0.2
0.1
0
-0.1
-0.2
20
40
60
80
100
120
140
160
φ1=0.2,φ2=-0.7
0.3
0.2
0.1
0
-0.1
-0.2
20
40
60
80
100
120
140
160

Figure 4

2.4 AR(p)

A *p-order autoregression*, denoted AR(2), satisfies

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} \varepsilon_t \tag{18}$$

where again ε_t is a WN. Using the lag operator

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p Y_{t-p}) Y_t = \varepsilon_t \tag{19}$$

The difference equation is stable provided that the roots of the polynomial

$$1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p$$

lie outside the unit circle. When this condition is satisfied the process is covariance stationary and the inverse of the autoregressive operator is

$$\psi(L) = (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)^{-1} = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + ...$$

with $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

2.4.1 Moments

- $E(Y_t) = 0$;

- The *jth autocovariance* is

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + ... + \phi_p \gamma_{j-p}$$

for $j = 1, 2, ...$

- The *variance* is

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + ... + \phi_p \gamma_p$$

- Dividing by γ_0 the autocovariances one obtains the Yule-Walker equations

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + ... + \phi_p \rho_{j-p}$$

2.4.2 Finding the roots of $(1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p)$

An easy way to find the roots of the polynomial is the following. Define two new vectors $Z_t = [Y_t, Y_{t-1}, ..., Y_{t-p+1}]'$, ${}_t = [\varepsilon_t, 0_{(1\times p-1)}]$ and a new matrix

$$F = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_p \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \vdots \\ 0 & \vdots & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

Then Z_t satisfies the AR(1)

$$Z_t = F Z_{t-1} + \epsilon_t$$

The roots of the polynomial $(1-\phi_1 z-\phi_2 z^2-...-\phi_p z^p)$ coincide with the reciprocal of the eigenvalues of F.

2.5 Causality and stationarity

Causality *An AR(p) process defined by the equation* $(1-\phi_1 L-\phi_2 L^2-\ldots-\phi_p L^p)Y_t=\phi(L)Y_t=\varepsilon_t$ *is said to be causal if there exists a sequence of constants* $\{\psi_j\}_{j=0}^{\infty}$ *such that* $\sum_{j=0}^{\infty}|\psi_j|<\infty$ *and* $Y_t=\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$.

Proposition An AR process $\phi(L)Y_t=\varepsilon_t$ is causal if and only if $\phi(z)\neq 0$ for all z such that $|z|\leq 1$.

Stationarity *The AR(p) is stationary if and only if* $\phi(z)\neq 0$ *for all the* z *such that* $|z|=1$

Here we focus on AR processes which are causal and stationary.

Example Consider the process

$$Y_t=\phi Y_{t-1}+\varepsilon_t$$

where ε_t is WN and $|\phi|>1$. Clearly the process is not causal. However we can rewrite the process as

$$Y_t=\frac{1}{\phi}Y_{t+1}-\frac{1}{\phi}\varepsilon_{t+1}$$

or using the forward operator $F=L^{-1}$

$$\begin{aligned}
Y_t &= \frac{1}{\phi}FY_t-\frac{1}{\phi}F\varepsilon_t\\
(1-\frac{1}{\phi}F)Y_t &= -\frac{1}{\phi}F\varepsilon_t
\end{aligned}$$

$$Y_t \;=\; -(1 - \frac{1}{\phi}F)^{-1}\frac{1}{\phi}F\varepsilon_t \tag{20}$$

which is a mean square convergent random variable. Using the lag operator it is easy to see that

$$(1 - \phi L) = ((\phi L)^{-1} - 1)(\phi L) = (1 - (\phi L)^{-1})(-\phi L)$$

3 ARMA

3.1 ARMA(p,q)

An ARMA(p,q) process includes both autoregressive and moving average terms:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} \tag{21}$$

where again ε_t is a WN. Using the lag operator

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) Y_t = (1 + \theta_1 L + \theta_2 1 L^2 + ... + \theta_p L^p) \varepsilon_t \tag{22}$$

Provided that the roots of the polynomial

$$1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p$$

lie outside the unit circle the ARMA process can bee written as

$$Y_t = \psi(L) \varepsilon_t$$

where

$$\psi(L) = \frac{(1 + \theta_1 L + \theta_2 1 L^2 + ... + \theta_p L^p)}{(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)}$$

Stationarity of the ARMA process depends on the AR parameters. Again a stationarity is implied by the roots being outside the unit circle in absolute value.

- The *variance* of the process is

$$\begin{aligned} E(Y_t^2) &= \phi_1 E(Y_{t-1}Y_t) + \phi_2 E(Y_{t-2}Y_t) + ... + \phi_p E(Y_{t-p}Y_t) + \\ &\quad + E(\varepsilon_t Y_t) + \theta_1 E(\varepsilon_{t-1}Y_t) + ... + \theta_q E(\varepsilon_{t-q}Y_t) \\ &= \phi_1[\sigma^2(\psi_1\psi_0 + \psi_2\psi_1 + ...)] + \phi_2[\sigma^2(\psi_2\psi_0 + \psi_3\psi_1 + ...)] \\ &\quad +... + \phi_p[\sigma^2(\psi_p\psi_0 + \psi_{p+1}2\psi_1 + ...)] + \\ &\quad + E(\psi_0\varepsilon_t^2) + \theta_1 E(\psi_1\varepsilon_t^2) + ... + \theta_q E(\psi_q\varepsilon_t^2) \\ &= \phi_1[\sigma^2(\psi_1\psi_0 + \psi_2\psi_1 + ...)] + \phi_2[\sigma^2(\psi_2\psi_0 + \psi_3\psi_1 + ...)] \\ &\quad +... + \phi_p[\sigma^2(\psi_p\psi_0 + \psi_{p+1}2\psi_1 + ...)] + \\ &\quad + \sigma^2(\psi_0 + \theta_1\psi_1 + ... + \theta_q\psi_q) \end{aligned}$$

and the j autocovariance is

$$\begin{aligned} \gamma_j = E(Y_tY_{t-j}) &= \phi_1 E(Y_{t-1}Y_{t-j}) + \phi_2 E(Y_{t-2}Y_{t-j}) + ... + \phi_p E(Y_{t-p}Y_{t-j}) + \\ &\quad + E(\varepsilon_t Y_{t-j}) + \theta_1 E(\varepsilon_{t-1}Y_{t-j}) + ... + \theta_q E(\varepsilon_{t-q}Y_{t-j}) \end{aligned}$$

for $j \leq q$

$$\begin{aligned} \gamma_j = E(Y_tY_{t-j}) &= \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + ... + \phi_p\gamma_{j-p} + \\ &\quad + \sigma^2(\theta_j\psi_0 + \theta_{j+1}\psi_1 ...) \end{aligned}$$

while for $j > q$ autocovariances are

$$\gamma_j = E(Y_tY_{t-j}) = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + ... + \phi_p\gamma_{j-p} +$$

There is a potential problem for redundant parametrization with ARMA processes. Consider a simple WN

$$Y_t = \varepsilon_t$$

and multiply both sides by $(1 - \rho L)$ to get

$$(1 - \rho L)Y_t = (1 - \rho L)\varepsilon_t$$

an ARMA(1,1) with $\theta = \phi = -\rho$. Both representations are valid however it is important to avoid such parametrization since we would get into trouble for estimating the parameter.

3.2 ARMA(1,1)

The ARMA(1,1) satisfies

$$(1 - \phi L)Y_t = (1 + \theta L)\varepsilon_t \tag{23}$$

where again ε_t is a WN, and

$$\psi(L) = \frac{(1 + \theta L)}{(1 - \phi L)}$$

Here we have that

$$\begin{aligned}
\gamma_0 &= \phi E(Y_{t-1}Y_t) + E(\varepsilon_t Y_t) + (\varepsilon_{t-1}Y_t) \\
\gamma_0 &= \phi\sigma^2(\psi_1\psi_0 + \psi_2\psi_1 + ...) + \sigma^2 + \theta\psi_1\sigma^2 \\
\gamma_1 &= \phi E(Y_{t-1}^2) + E(\varepsilon_t Y_{t-1}) + E(\theta\varepsilon_{t-1}Y_{t-1}) = \phi\gamma_0 + \theta\sigma^2 \\
\gamma_2 &= \phi\gamma_1
\end{aligned} \tag{24}$$

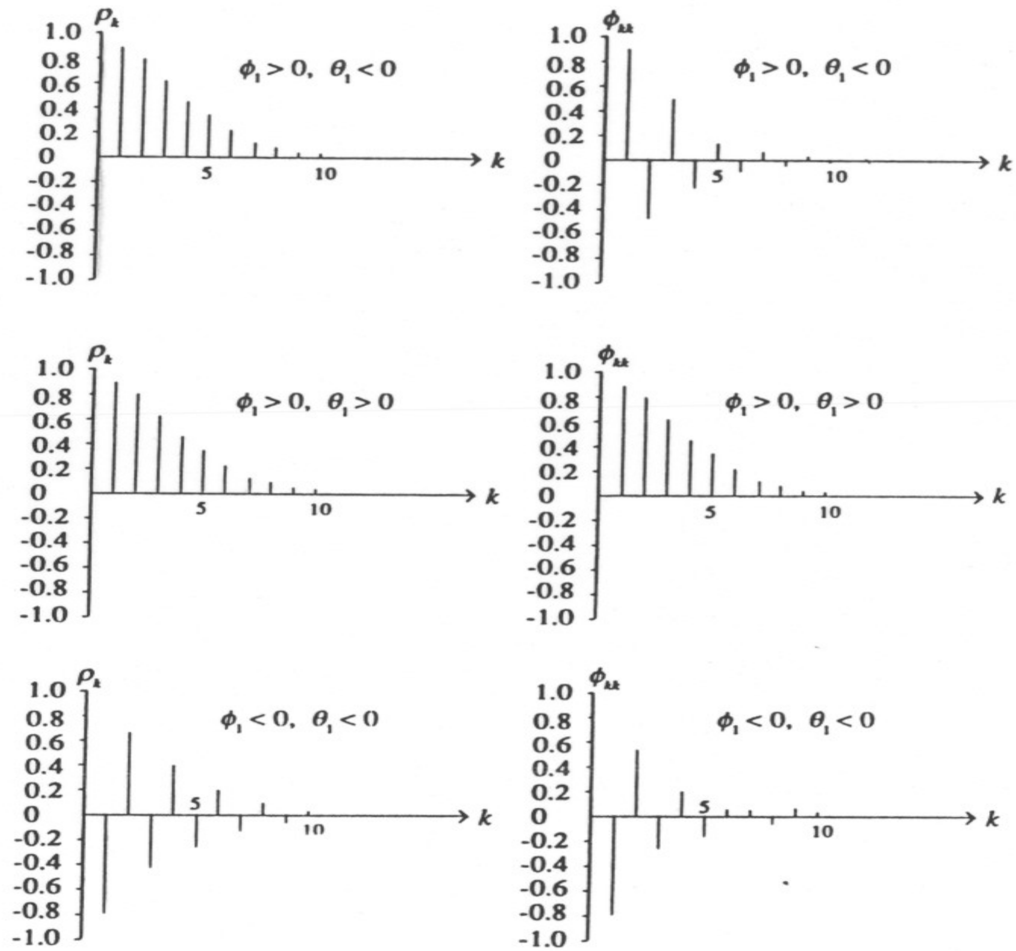


Fig. 3.14 ACF and PACF of ARMA(1,1) model $(1-\phi_1 B)\dot{Z}_t = (1-\theta_1 B)a_t$.

Figure . Source: W.Wei "Time Series Analysis: Univariate and Multivariate Methods".

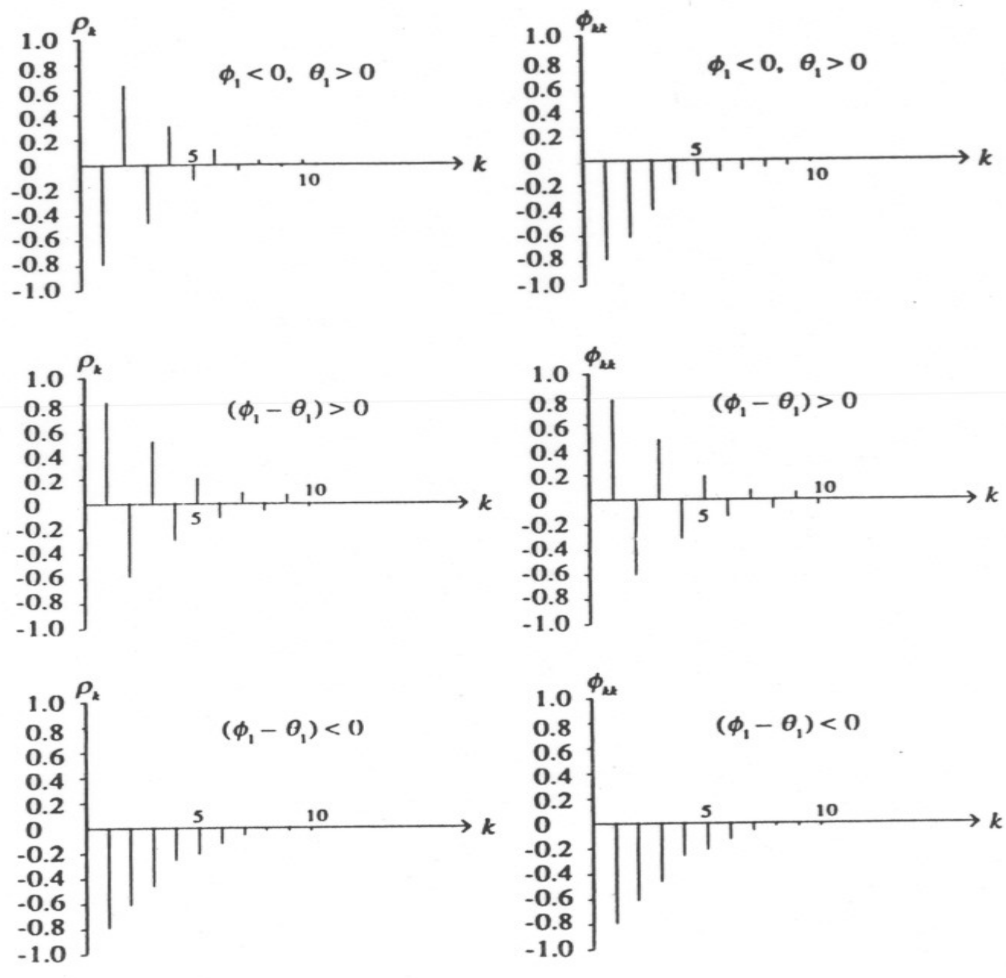


Fig. 3.14 *(continued)*

Figure . Source: W.Wei "Time Series Analysis: Univariate and Multivariate Methods".

3. ESTIMATION[1]

[1]This part is based on the Hamilton textbook.

1 Estimating an autoregression with OLS

Assumption: *The regression model is*

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$$

with roots of $(1 - \phi_1 z + \phi_2 z^2 + ... + \phi_p z^p) = 0$ *outside the unit circle and with* ε_t *an i.i.d. sequence with zero mean, variance* σ^2 *and finite fourth moment.*

The autoregression be be written in the standard form where $\mathbf{x}_t = [y_{t-1}, ..., y_{t-p}]'$.

Note that the autoregression cannot satisfy assumption A1 since u_t is not independent of y_{t+1}.

$\Rightarrow$ The estimator is biased

1.1 Asymptotic results for the estimator

The asymptotic results seen before hold. Suppose we have $T+p$ observations so that T observations can be used to estimate the model. Then:

Consistency

$$\begin{aligned}
\hat{\beta}-\beta &= \left[\sum_{t=1}^{T}(\mathbf{x}_t\mathbf{x}_t')\right]^{-1}\left[\sum_{t=1}^{T}(\mathbf{x}_t u_t)\right] \\
&= \left[(1/T)\sum_{t=1}^{T}(\mathbf{x}_t\mathbf{x}_t')\right]^{-1}\left[(1/T)\sum_{t=1}^{T}(\mathbf{x}_t u_t)\right]
\end{aligned}$$

Again $\mathbf{x}_t u_t$ is a martingale difference sequence with finite variance covariance matrix given by $E(\mathbf{x}_t u_t \mathbf{x}_t' u_t) = E(\mathbf{x}_t\mathbf{x}_t')\sigma^2$

$$\left[(1/T)\sum_{t=1}^{T}(\mathbf{x}_t u_t)\right]^{-1} \xrightarrow{p} 0$$

Moreover the first term is

$$\left[\sum_{t=1}^{T}(\mathbf{x}_t\mathbf{x}_t')\right]^{-1} = \begin{pmatrix}
1 & T^{-1}\sum y_{t-1} & T^{-1}\sum y_{t-2} & \dots & T^{-1}\sum y_{t-p} \\
T^{-1}\sum y_{t-1} & T^{-1}\sum y_{t-1}^2 & T^{-1}\sum y_{t-1}y_{t-2} & \dots & T^{-1}\sum y_{t-1}y_{t-p} \\
T^{-1}\sum y_{t-2} & T^{-1}\sum y_{t-2}y_{t-1} & T^{-1}\sum y_{t-2}^2 & \dots & T^{-1}\sum y_{t-2}y_{t-p} \\
\vdots & \vdots & \vdots & \dots & \vdots \\
T^{-1}\sum y_{t-p} & T^{-1}\sum y_{t-p}y_{t-1} & T^{-1}\sum y_{t-p}y_{t-2} & \dots & T^{-1}\sum y_{t-p}^2
\end{pmatrix}$$

The elements in the first row or column converge in probability to μ by proposition C.13L. The other elements by Theorem 7.2.1BD converge in probability to

$$E(y_{t-i}y_{t-j}) = \gamma_{|i-j|} + \mu^2$$

hence

$$\left[(1/T)\sum_{t=1}^{T}(\mathbf{x}_t\mathbf{x}_t')\right]^{-1} \xrightarrow{p} Q^{-1}$$

Therefore as before

$$\hat{\beta} - \beta \xrightarrow{p} Q^{-1}0$$

verifying that the estimator is consistent.

Asymptotic normality

Again this follows from the fact that $\mathbf{x}_t u_t$ is a martingale difference sequence with variance covariance matrix given by $E(\mathbf{x}_t u_t \mathbf{x}_t' u_t) = E(\mathbf{x}_t \mathbf{x}_t')\sigma^2$ so that

$$\left[(1/\sqrt{T}) \sum_{t=1}^{T} (\mathbf{x}_t u_t') \right] \xrightarrow{L} N(0, \sigma^2 Q)$$

and

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{L} N(0, Q^{-1}(\sigma^2 Q) Q^{-1}) = N(0, \sigma^2 Q^{-1})$$

2 Maximum likelihood: Introduction

Consider an ARMA model of the form

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} \tag{1}$$

with $\varepsilon_t \sim i.i.d. N(0, \sigma^2)$.

Here we explore how to estimate the values of $\phi_1, ..., \phi_p, \theta_1, ..., \theta_q$ on the basis of observations of y. The principle on which the estimation is based is *maximum likelihood.*

Let $\boldsymbol{\theta} = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \sigma^2)$ denote the vector of population parameters and suppose we have a sample of T observations $(y_1, y_2, ..., y_T)$.

The approach is to calculate the probability density or likelihood function

$$f(y_T, y_{T-1}, ..., y_1; \boldsymbol{\theta}) \tag{2}$$

The maximum likelihood estimates (ML) is the value of θ such that (2) is maximized.

2.1 Likelihood function for an AR(1)

Consider the Gaussian AR(1)

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t \tag{3}$$

with $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. In this case $\boldsymbol{\theta} = (c, \phi, \sigma^2)'$. Since ε_t is Gaussian y_1 is also Gaussian with

$$\mu = c/(1-\phi)$$

and variance

$$E(y_1 - \mu)^2 = \sigma^2/(1-\phi^2)$$

Hence the density of the first observation is

$$f(y_1; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi^2)}} exp\left\{-\frac{[y_1 - c/(1-\phi)]^2}{2\sigma^2/(1-\phi^2)}\right\}$$

Conditional on the first observation the density of y_2 is

$$\begin{aligned} f(y_2|y_1; \boldsymbol{\theta}) &= N(c + \phi y_1, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{[y_2 - c - \phi y_1]^2}{2\sigma^2}\right\} \end{aligned} \tag{4}$$

Similarly the conditional density for the third observation is

$$\begin{aligned} f(y_3|y_2, y_1; \theta) &= N(c + \phi y_2, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{[y_3 - c - \phi y_2]^2}{2\sigma^2}\right\} \end{aligned} \tag{5}$$

Note that for the tth observation, given the dependence only on $t-1$ the conditional density is

$$\begin{aligned} f(y_t|y_{t-1};\boldsymbol{\theta}) &= N(c+\phi y_{t-1},\sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{[y_t-c-\phi y_{t-1}]^2}{2\sigma^2}\right\} \end{aligned} \tag{6}$$

and the joint density of the first t observation can be written as

$$\begin{aligned} f(y_t,y_{t-1},...,y_1;\boldsymbol{\theta}) &= f(y_t|y_{t-1},...,y_1;\boldsymbol{\theta})f(y_{t-1},...,y_1;\theta) \\ &= f(y_t|y_{t-1},...,y_1;\theta)f(y_{t-1}|y_{t-2},...,y_1;\boldsymbol{\theta}) \\ &= f(y_t|y_{t-1},...,y_1;\theta)f(y_{t-1}|y_{t-2},...,y_1;\boldsymbol{\theta})...f(y_2|y_1;\boldsymbol{\theta})f(y_1;\boldsymbol{\theta}) \\ &= f(y_1;\theta)\prod_{t=2}^{T} f(y_t|y_{t-1},...,y_1;\boldsymbol{\theta}) \end{aligned}$$

and the log-likelihood is

$$\mathcal{L}(\boldsymbol{\theta}) = \log f(y_1;\boldsymbol{\theta}) + \sum_{t=2}^{T} \log f(y_t|y_{t-1},...,y_1;\boldsymbol{\theta})) \tag{7}$$

Substituting (6) into (7) we obtain the log-likelihood function for a sample of size T from a Gaussian AR(1)

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left[\sigma^2/(1-\phi^2))\right] - \frac{[y_1-c/(1-\phi)]^2}{2\sigma^2/(1-\phi^2)} - \\ & -\frac{(T-1)}{2}\log(2\pi) - \frac{(T-1)}{2}\log(\sigma^2) - \sum_{t=2}^{T}\left[\frac{(y_t-c-\phi y_{t-1})^2}{2\sigma^2}\right] \end{aligned} \tag{8}$$

(8) is the exact likelihood function. To find ML estimates numerical optimization must be used

2.2 Conditional maximum likelihood estimates

An alternative to numerical optimization of the exact likelihood function is to consider the first value y_1 to be deterministic and to minimize the likelihood conditional on the first observation. In this case the *conditional* log-likelihood is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=2}^{T} \log f(y_t|y_{t-1}, ..., y_1; \boldsymbol{\theta})) \tag{9}$$

that is

$$\mathcal{L}(\boldsymbol{\theta}) \;=\; -\frac{(T-1)}{2}\log(2\pi) - \frac{(T-1)}{2}\log(\sigma^2) - \sum_{t=2}^{T}\left[\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}\right] \tag{10}$$

Maximization of 10 is equivalent to minimization of

$$\sum_{t=2}^{T}(y_t - c - \phi y_{t-1})^2$$

which is achieved by an OLS regression of y_t on a constant and lagged values. The conditional ML estimates of c, ϕ are given by

$$\begin{pmatrix} \hat{c} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} T-1 & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_t \\ \sum y_t y_{t-1} \end{pmatrix} \tag{11}$$

The conditional ML estimates of the innovation variance is found by differentiating 10 with respect to σ^2 and setting the result equal to zero.

$$-\frac{T-1}{2\sigma^2} + \sum_{t=2}^{T} \frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^4} = 0$$

which gives

$$\hat{\sigma}^2 = \frac{\sum_{t=2}^{T}(y_t - c - \hat{\phi} y_{t-1})^2}{T-1} \tag{12}$$

2.3 Likelihood function for an AR(p)

Consider the Gaussian AR(p)

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} \varepsilon_t \tag{13}$$

with $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. In this case $\boldsymbol{\theta} = (c, \phi_1, \phi_2, ...\phi_p, \sigma^2)'$. Here we only study the conditional likelihood.

2.3.1 Conditional maximum likelihood estimates

The conditional likelihood can be derived as in the AR(1) case. In particular we have

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= -\frac{(T-p)}{2}\log(2\pi) - \frac{(T-p)}{2}\log(\sigma^2) - ... \\ &- \sum_{t=p+1}^{T}\left[\frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - ... - \phi_p y_{t-p})^2}{2\sigma^2}\right]\end{aligned} \tag{14}$$

Again maximization of 14 is equivalent to minimization of

$$\sum_{t=2}^{T}(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - ... - \phi_p y_{t-p})^2$$

which is achieved by an OLS regression of y_t on a constant and p lagged values of y.

The conditional ML estimates of the innovation variance is found by differentiating 10 with respect to σ^2 and setting the result equal to zero

$$\hat{\sigma}^2 = \frac{\sum_{t=p+1}^{T}(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - ... - \phi_p y_{t-p})^2}{T-p} \tag{15}$$

• Note that even if disturbances are not Gaussian ML estimates of the parameters are consistent estimates of the population parameters because they correspond to OLS estimates and we have seen in the previous class that consistency does not depend the assumption of normality.

• An estimate that maximizes a misspecified likelihood function is known a s a *quasi-maximum likelihood estimate.*

2.4 Conditional likelihood function for an MA(1)

Consider the MA(1) process

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

with $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. In this case $\boldsymbol{\theta} = (\mu, \theta, \sigma^2)'$.

If the value of ε_{t-1} were known with certainty then

$$y_t|\varepsilon_{t-1} \sim N(\mu + \theta\varepsilon_{t-1}, \sigma^2)$$

or

$$f(y_t|\varepsilon_{t-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-(y_t - \mu - \theta\varepsilon_{t-1})^2}{2\sigma^2}\right]$$

Suppose we knew $\varepsilon_0 = 0$. Then

$$y_1|\varepsilon_0 \sim N(\mu, \sigma^2)$$

Moreover given y_1, $\varepsilon_1 = y_1 - \mu$ we know

$$f(y_2|y_1, \varepsilon_0 = 0; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-(y_2 - \mu - \theta\varepsilon_1)^2}{2\sigma^2}\right]$$

Since ε_1 is known with certainty $\varepsilon_2 = y_2 - \mu - \theta\varepsilon_1$. Proceeding this way given the knowledge of ε_0 the full sequence of innovations can be calculated by iterating on $\varepsilon_t = y_t - \mu - \varepsilon_{t-1}$.

The conditional density for the tth observation can be calculated as

$$\begin{aligned} f(y_t|y_{t-1}, y_{t-2}, ..., y_1, \varepsilon_0 = 0; \theta) &= f(y_t|\varepsilon_{t-1}; \theta) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-\varepsilon_t^2}{2\sigma^2}\right] \end{aligned} \tag{16}$$

The sample likelihood would be the product of individual densities

$$f(y_t, y_{t-1}, y_{t-2}, ..., y_1|\varepsilon_0 = 0; \theta) = \prod_{t=1}^{T} f(y_t|y_{t-1}, y_{t-2}, ..., y_1|\varepsilon_0 = 0; \theta) \tag{17}$$

and the conditional log-likelihood is

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=1}^{T}\frac{\varepsilon_t^2}{2\sigma^2} \tag{18}$$

- For a particular numerical value of θ we calculate the sequence of ε's. The conditional likelihood is a function of the sum of the squares.

- ML estimates have to be found by numerical optimization.

2.5 Conditional likelihood function for an MA(q)

Consider the MA(q) process

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_w \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$

with $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. In this case $\boldsymbol{\theta} = (\mu, \theta_1, ..., \theta_q, \sigma^2)'$.

As before a simple approach is to condition on the assumption

$$\varepsilon_0 = \varepsilon_{-1} = ... = \varepsilon_{-q+1} = 0$$

Using these starting values we can iterate on

$$\varepsilon_t = y_t - \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}$$

The conditional likelihood is

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2} \tag{19}$$

2.6 Likelihood function for an ARMA(p,q)

A Gaussian ARMA(p,q) takes the form

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_w \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$

with $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. In this case $\boldsymbol{\theta} = (\mu, \phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \sigma^2)'$.

2.7 Conditional Likelihood function

Taking values of $y_0, ...y_{-p+1}, \varepsilon_0, ..., \varepsilon_{-q+1}$ as given the sequence of ε's can be calculated using

$$\varepsilon_t = y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - ... - \phi_p y_{t-p} - \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_w \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}$$

The conditional likelihood is then

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2} \tag{20}$$

Again numerical optimizxation has to be used to compute the ML estimates.

4. FORECASTING[1]

[1]This part is based on the Hamilton textbook.

1 Principles of forecasting

1.1 Forecast based on conditional expectations

• Suppose we are interested in forecasting the value of Y_{t+1} based on a set of variables X_t.

• Let $Y_{t+1|t}$ denote such a forecast.

• To evaluate the usefulness of the forecast we need to specify a *loss* function. Here we specify a quadratic loss function. A quadratic loss function means that $Y_{t+1|t}$ is chose to minimize $E(Y_{t+1} - Y_{t+1|t})^2$.

• $E(Y_{t+1} - Y_{t+1|t})^2$ is known as the *mean squared error* associated with the forecast $Y_{t+1|t}$ denoted

$$MSE(Y_{t+1|t}) = E(Y_{t+1} - Y_{t+1|t})^2$$

• Fundamental result: the forecast with the smallest MSE is the expectation of $Y_{t+1|t}$ conditional on X_t that is

$$Y_{t+1|t} = E(Y_{t+1}|X_t)$$

We now verify the claim. Let $g(X_t)$ be any other function and let $Y_{t+1|t} = g(X_t)$. The associated MSE is

$$\begin{aligned} E\left[Y_{t+1} - g(X_t)\right]^2 &= E\left[Y_{t+1} - E(Y_{t+1}|X_t) + E(Y_{t+1}|X_t) - g(X_t)\right]^2 \\ &= E\left[Y_{t+1} - E(Y_{t+1}|X_t)\right]^2 + \\ &\quad +2E\left\{\left[Y_{t+1} - E(Y_{t+1}|X_t)\right]\left[E(Y_{t+1}|X_t) - g(X_t)\right]\right\} + \\ &\quad +E\{[E(Y_{t+1}|X_t) - g(X_t)\}^2 \end{aligned} \tag{2}$$

Define

$$\eta_{t+1} \equiv \left[Y_{t+1} - E(Y_{t+1}|X_t)\right]\left[E(Y_{t+1}|X_t) - g(X_t)\right] \tag{3}$$

The conditional expectation is

$$\begin{aligned} E(\eta_{t+1}|X_t) &= E\left\{\left[Y_{t+1} - E(Y_{t+1}|X_t)\right]\left[E(Y_{t+1}|X_t) - g(X_t)\right]|X_t\right\} \\ &= \left[E(Y_{t+1}|X_t) - g(X_t)\right] E\left\{\left[Y_{t+1} - E(Y_{t+1}|X_t)\right]|X_t\right\} \\ &= \left[E(Y_{t+1}|X_t) - g(X_t)\right]\left[E(Y_{t+1}|X_t) - E(Y_{t+1}|X_t)\right] \\ &= 0 \end{aligned}$$

Therefore by law of iterated expectations

$$E(\eta_{t+1}) = E(E(\eta_{t+1}|X_t)) = 0$$

This means that

$$E\left[Y_{t+1} - g(X_t)\right]^2 \;=\; E\left[Y_{t+1} - E(Y_{t+1}|X_t)\right]^2 + E([E(Y_{t+1}|X_t) - g(X_t)])^2$$

Therefore the function that minimizes the MSE is

$$g(X_t) = E(Y_{t+1}|X_t)$$

$E(Y_{t+1}|X_t)$ is the optimal forecast of Y_{t+1} conditional of X_t under a quadratic loss function. The MSE of this optimal forecast is

$$E[Y_{t+1} - g(X_t)]^2 = E[Y_{t+1} - E(Y_{t+1}|X_t)]^2$$

1.2 Forecast based on linear projections

We now restrict the class of forecasts we consider to be linear function of X_t:

$$Y_{t+1|t} = \alpha' X_t$$

Suppose α' is such that the resulting forecast error is uncorrelated with X_t

$$E[(Y_{t+1} - \alpha' X_t) X_t'] = 0' \tag{4}$$

If (4) holds the we call $\alpha' X_t$ the linear projection of Y_{t+1} on X_t.

• The linear projection produces the smallest forecast error among the class of linear forecasting rules. To verify this let $g' X_t$ be any arbitrary forecasting rule.

$$\begin{aligned}
E\left(Y_{t+1} - g' X_t\right)^2 &= E\left(Y_{t+1} - \alpha' X_t + \alpha' X_t - g' X_t\right)^2 \\
&= E\left(Y_{t+1} - \alpha' X_t\right)^2 + \\
&\quad +2E\left[\left(Y_{t+1} - \alpha' X_t\right)\left(\alpha' X_t - g' X_t\right)\right] + \\
&\quad +E\left(\alpha' X_t - g' X_t\right)^2
\end{aligned}$$

the middle term

$$\begin{aligned}
E[\left(Y_{t+1} - \alpha' X_t\right)\left(\alpha' X_t - g' X_t\right)] &= E[\left(Y_{t+1} - \alpha' X_t\right) X_t' \left[\alpha - g\right]) \\
&= E[\left(Y_{t+1} - \alpha' X_t\right) X_t'] \left[\alpha - g\right] \\
&= 0'
\end{aligned}$$

by definition of linear projection. Thus

$$E\left(Y_{t+1} - g'X_t\right)^2 \;=\; E\left(Y_{t+1} - \alpha'X_t\right)^2 + E\left(\alpha'X_t - g'X_t\right)^2$$

The optimal linear forecast is the value $g'X_t = \alpha'X_t$. We use the notation

$$\hat{P}(Y_{t+1}|X_t) = \alpha'X_t$$

to indicate the linear projection of Y_{t+1} on X_t.

Notice that

$$MSE[\hat{P}(Y_{t+1}|X_t)] \geq MSE\left[E(Y_{t+1}|X_t)\right]$$

The projection coefficient α can be calculated in terms of moments of Y_{t+1} and X_t.

$$E(Y_{t+1}X_t') = \alpha'E(X_tX_t')$$

$$\alpha' = [E(X_tX_t')]^{-1}E(Y_{t+1}X_t')$$

Here we denote $\hat{Y}_{t+s|t} = \hat{E}(Y_{t+s}|X_t) = \hat{P}(1, Y_{t+s}|X_t)$ the best linear forecast of Y_{t+s} conditional on X_t.

1.3 Linear projections and OLS regression

There is a close relationship between OLS estimator and the linear projection coefficient. If Y_{t+1} and X_t are stationary processes and also ergodic for the second moments then

$$(1/T)\sum_{t=1}^{T} X_t X_t' \xrightarrow{p} E(X_t X_t')$$
$$(1/T)\sum_{t=1}^{T} X_t Y_{t+1} \xrightarrow{p} E(X_t Y_{t+1})$$

implying

$$\hat{\beta} \xrightarrow{p} \alpha$$

The OLS regression yields a consistent estimate of the linear projection coefficient.

2 Forecasting an AR(1)

For the covariance-stationary $AR(1)$ we have

$$\hat{Y}_{t+s|t} = \hat{E}(Y_{t+s}|Y_t, Y_{t-1}, ...) \ = \ \phi^s Y_t$$

The forecast decays geometrically toward zero as the forecast horizon increases.

The forecast error is

$$Y_{t+s} - \hat{Y}_{t+s|t} = \varepsilon_{t+s} + \phi\varepsilon_{t+s-1} + \phi^2\varepsilon_{t+s-2} + ... + \phi^{s-1}\varepsilon_{t+1}$$

The associated MSE will be

$$\begin{aligned} E(Y_{t+s} - \hat{Y}_{t+s|t})^2 &= E(\varepsilon_{t+s} + \phi\varepsilon_{t+s-1} + \phi^2\varepsilon_{t+s-2} + ... + \phi^{s-1}\varepsilon_{t+1})^2 \\ &= (1 + \phi^2 + \phi^4 + ... + \phi^{2(s-1)})\sigma^2 \\ &= \frac{1-\phi^{2s}}{1-\phi^2}\sigma^2 \end{aligned}$$

Notice that

$$\lim_{s\to\infty} \hat{Y}_{t+s|t} = 0$$

(in general it will converge to the mean of the process) and

$$\lim_{s\to\infty} E(Y_{t+s} - \hat{Y}_{t+s|t})^2 = \frac{\sigma^2}{1-\phi^2}$$

which is the variance of the process.

3 Forecasting an AR(p)

Now consider an $AR(p)$. Recall the $AR(p)$ can be written as

$$Z_t = FZ_{t-1} + \epsilon_t$$

where $Z_t = [Y_t, Y_{t-1}, ..., Y_{t-p+1}]\prime$, $_t = [\varepsilon_t, 0, ..., 0]'$ and

$$F = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

Therefore

$$Y_{t+s} = f_{11}^{(s)} Y_t + f_{12}^{(s)} Y_{t-1} + ... + f_{1p}^{(s)} Y_{t-p+1} + \varepsilon_{t+s} + f_{11}^1 \varepsilon_{t+s-1} + f_{11}^2 \varepsilon_{t+s-2} + ... + f_{11}^{s-1} \varepsilon_{t+1}$$

where f_{mn}^j denotes the (m, n) element of F^j.

The optimal s-step ahead forecast is thus

$$\hat{E}(Y_{t+s}|Y_t, Y_{t-1}, ...) \;=\; f_{11}^{(s)} Y_t + f_{12}^{(s)} Y_{t-1} + ... + f_{1p}^{(s)} Y_{t-p+1} \tag{5}$$

and the associated forecast error

$$Y_{t+s} - \hat{E}(Y_{t+s}|Y_t, Y_{t-1}, ...) \;=\; \varepsilon_{t+s} + f_{11}^1 \varepsilon_{t+s-1} + f_{11}^2 \varepsilon_{t+s-2} + ... + f_{11}^{s-1} \varepsilon_{t+1} \tag{6}$$

The forecast can be computed recursively. Let $\hat{Y}_{t+1|t}$ be one period step-ahead forecast of Y_{t+1} we have

$$\hat{Y}_{t+1|t} = \phi_1 Y_t + \phi_2 Y_{t-1} + ... + \phi_p Y_{t-p+1} \tag{7}$$

In general the j-step ahead forecast $\hat{Y}_{t+j|t}$ can be computed using the recursion

$$\hat{Y}_{t+j|t} = \phi_1 \hat{Y}_{t+j-1|t} + \phi_2 \hat{Y}_{t+j-2|t} + ... + \phi_p \hat{Y}_{t+j-p|t} \tag{8}$$

An easy way to see this is to use the AR(1) process

$$Z_t = FZ_{t-1} + \epsilon_t$$

We have

$$\begin{aligned} \hat{Z}_{t+1|t} &= FZ_t \\ \hat{Z}_{t+2|t} &= F^2 Z_t \\ &\vdots \\ \hat{Z}_{t+s|t} &= F^s Z_t \end{aligned}$$

which means

$$\hat{Z}_{t+s|t} = F\hat{Z}_{t+s-1|t}$$

The forecast $\hat{Y}_{t+s|t}$ will be the first element of $\hat{Z}_{t+s|t}$

The associated forecast errors will be

$$
\begin{aligned}
Z_{t+1} - \hat{Z}_{t+1|t} &= \epsilon_{t+1} \\
Z_{t+2} - \hat{Z}_{t+2|t} &= F\epsilon_{t+1} + \epsilon_{t+2} \\
&\vdots \\
Z_{t+s} - \hat{Z}_{t+s|t} &= F^{s-1}\epsilon_{t+1} + F^{s-2}\epsilon_{t+2} + ... + F\epsilon_{t+s-1} + \epsilon_{t+s}
\end{aligned}
$$

Let $E(\epsilon_t\epsilon_t') = \Sigma$. The mean squared errors will be

$$
\begin{aligned}
MSE(\hat{Z}_{t+1|t}) &= \Sigma \\
MSE(\hat{Z}_{t+2|t}) &= F\Sigma F' + \Sigma \\
&\vdots \\
MSE(\hat{Z}_{t+s|t}) &= F^{s-1}\Sigma F^{s-1\prime} + F^{s-2}\Sigma F^{s-2\prime} + ... + F\Sigma F' + \Sigma \\
&= \sum_{j=0}^{s-1} F^j \Sigma F^{j\prime}
\end{aligned}
$$

4 Forecasting an MA(1)

Consider the invertible MA(1)

$$Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

with $|\theta| < 1$. Replacing in the Wiener-Kolmogorov formula we obtain

$$\hat{Y}_{t+s|t} = \left[\frac{1+\theta L}{L^s}\right]_+ \frac{1}{1+\theta L} Y_t$$

For $s = 1$

$$\hat{Y}_{t+1|t} = \frac{\theta}{1+\theta L} Y_t = \theta\varepsilon_t \tag{9}$$

where ε_t is the outcome of the infinite recursion $\varepsilon_t = Y_t - \theta\varepsilon_{t-1}$ For $s = 2, 3, ...$

$$\hat{Y}_{t+1|t} = 0 \tag{10}$$

because $\left[\frac{\psi(L)}{L^s}\right]_+ = 0$

5 Forecasting an MA(q)

For an invertible MA(q) process

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

$$\hat{Y}_{t+s|t} = \left[\frac{1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q}{L^s} \right]_+ \frac{1}{1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q} Y_t$$

where

$$\left[\frac{1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L}{L^s} \right]_+ = \theta_s + \theta_{s+1} L^1 + ... + \theta_q L^{q-s} \quad \text{for} \quad s = 1, 2, ...q$$

and zero for $s = q + 1, q + 2, ...$

Therefore the forecast for horizons $s = 1, 2, ..., q$ is given by

$$\hat{Y}_{t+s|t} = (\theta_s + \theta_{s+1} L + ... + \theta_q L^{q-s}) \varepsilon_t$$

and zero for the other horizons.

6 Forecasting an ARMA(1,1)

Consider the ARMA(1,1) process

$$(1-\phi L)Y_t = (1+\theta L)\varepsilon_t$$

with $|\phi|, |\theta| < 1$.

$$\hat{Y}_{t+s|t} = \left[\frac{1+\theta L}{(1-\phi L)L^s}\right]_+ \frac{1-\phi L}{1+\theta L} Y_t$$

$$\begin{aligned}
\left[\frac{1+\theta L}{(1-\phi L)L^s}\right]_+ &= \left[\frac{1+\phi L+\phi^2 L^2+...}{L^s} + \frac{\theta L(1+\phi L+\phi^2 L^2+...)}{L^s}\right]_+ \\
&= (\phi^s + \phi^{s+1}L + \phi^{s+2}L^2 + ...) + \theta(\phi^{s-1} + \phi^s L + \phi^{s+1}L^2 + ...) \\
&= (\phi^s + \theta\phi^{s-1})(1+\phi L+\phi^2 L+^2+...) \\
&= \frac{\phi^s + \theta\phi^{s-1}}{1-\phi L}
\end{aligned}$$

Therefore

$$\begin{aligned}
\hat{Y}_{t+s|t} &= \left[\frac{\phi^s+\theta\phi^{s-1}}{1-\phi L}\right] \frac{1-\phi L}{1+\theta L} Y_t \\
&= \left[\frac{\phi^s+\theta\phi^{s-1}}{1+\theta L}\right] Y_t
\end{aligned}$$

for $s = 1$ the forecast is

$$\hat{Y}_{t+1|t} = \left[\frac{\phi+\theta}{1+\theta L}\right] Y_t$$

This can be written as

$$\hat{Y}_{t+1|t} = \frac{\phi(1+\theta L) + \theta(1-\phi L)}{1+\theta L} Y_t = \phi Y_t + \theta \varepsilon_t$$

where

$$\begin{aligned} \varepsilon_t &= \frac{(1-\phi L)}{(1+\theta L)} Y_t \\ &= Y_t - \phi Y_{t-1} - \theta \varepsilon_{t-1} \end{aligned}$$

For $s = 2, 3, ...$ the forecast obeys the recursion

$$\hat{Y}_{t+s|t} = \phi \hat{Y}_{t+s-1|t}$$

7 Direct forecast

An alternative is to compute the direct forecast by computing the projection of $\mathbf{Y}_{t+h}$ on $\mathbf{Y}_t$. To see this consider a bivariate VAR(p) with two variables, x_t and y_t. We want to forecast x_{t+h} given the incormation available at time t.

The direct forecast works as follows:

1. Estimate the projection equation

$$x_t = a + \sum_{i=0}^{p-1} \phi_i x_{t-h-i} + \sum_{i=0}^{p-1} \psi_i y_{t-h-i} + \varepsilon_t$$

2. Using the estimated coefficients, the predictor $x_{t+h|t}$ is obtain as

$$\hat{x}_{t+h|t} = a + \sum_{i=0}^{p-1} \phi_i x_{t-i} + \sum_{i=0}^{p-1} \psi_i y_{t-i}$$

8 Comparing Predictive Accuracy

Diebold and Mariano propose a procedure to formally compare the forecasting performance of two competing models. Let $\hat{Y}^1_{t+s|t}, \hat{Y}^2_{t+s|t}$ be two forecast based on the same information set but obtained using two different models (i.e. MA(1) and AR(1)). Let

$$\begin{aligned} w^1_{\tau+s|\tau} &= Y_{\tau+s} - \hat{Y}^1_{\tau+s|t} \\ w^2_{\tau+s|\tau} &= Y_{\tau+s} - \hat{Y}^2_{\tau+s|t} \end{aligned}$$

be the two forecast errors where $\tau = T_0, ...T - s$.

The accuracy of each forecast is measured by a particular loss function, say quadratic i.e. $L(w^i_{\tau+s|\tau}) = \left(w^1_{\tau+s|\tau}\right)^2$. The Diebold Mariano procedure is based on a test of the null hypothesis

$$H_0 : E(d_\tau) = 0, \quad H_1 : E(d_\tau) \neq 0$$

where $d_\tau = L(w^1_{\tau+s|\tau}) - L(w^2_{\tau+s|\tau})$. The Diebold-Mariano statistic is

$$S = \frac{\bar{d}}{\left(L\hat{R}V_{\bar{d}}\right)^{1/2}}$$

where

$$\bar{d} = (1/(T - T_0 - s)) \sum_{\tau=T_0}^{T-s} d_\tau$$

and $L\hat{R}V_{\bar{d}}$ is a consistent estimate of

$$LRV_{\bar{d}} = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j$$

and $\gamma_j = Cov(d_t, d_{t-j})$.

Under the null

$$S \xrightarrow{L} N(0,1)$$

- DM test is for forecast comparison not model comparison.

- When using forecast obtained from models one has to be careful. In nested models the distribution of the DM statistic is non-normal.

- For nested model an alternative is the bootstrapping procedure in Clark and McCracken (Advances in Forecast Evaluation, 2013).

9 Forecast in practice

• So far we assumed that the value of the coefficients is known. This is obviously not the case in practice. In real applications we will have to estimate the value of the parameters.

• For instance with an AR(p) we have to get estimate $\hat{\phi}_1, \hat{\phi}_2, ..., \hat{\phi}_p$ using OLS and then using the formulas seen before to produce forecasts of Y.

• Suppose we have data up to date T. We estimate the model using all the T observations for Y_t and we forecast Y_{T+s}, call it $\tilde{Y}_{T+s|T}$ to distinguish it from the forecast where coefficients were known with certainty $\hat{Y}_{T+s|T}$.

9.1 Forecast evaluation

• A key issue in forecasting is to evaluate the forecasting accuracy of a model of interest. In particular several times we will be interested in comparing the performance of competing forecasting model, e.g. AR(1) vs. ARMA(1,1).

• How can we perform such a forecast evaluation?

• Answer: we can compare the mean squared errors using pseudo-out of sample forecast exercises.

9.2 Pseudo out-of-sample exercises

Suppose we have a sample of T observations. Let $\tau = T_0$ and let $i = 1$. A pseudo out-of-sample exercise works as follows:

1. We use τ observations to estimate the parameters of the model.
2. We forecast $Y_{\tau+j}$ with obtaining $\tilde{Y}_{\tau+j|\tau}$ $j = 1, 2, ..., s$.
3. We compute the forecast error $w_{\tau+j|\tau} = Y_{\tau+j} - \tilde{Y}_{\tau+j|\tau}$ with $j = 1, 2, ..., s$.
4. We update $\tau = \tau + 1$ and repeat steps 1-3.

We repeat steps 1-4 up to the end of the sample and we compute the mean squared error

$$M\hat{S}E(\tilde{Y}_{T+j|T}) = \frac{1}{T - T_0 - j} \sum_{\tau=T_0}^{T-T_0-j} w_{\tau+j|\tau}^2$$

or the root mean squared error

$$RM\hat{S}E(\tilde{Y}_{T+j|T}) = \sqrt{\frac{1}{T - T_0 - j} \sum_{\tau=T_0}^{T-T_0-j} w_{\tau+j|\tau}^2}$$

5. FORECASTING: APPLICATIONS

“Why has U.S. inflation become harder to forecast?”
By Stock, J and M., Watson

10 “Why has U.S. inflation become harder to forecast?”

• The rate of price inflation in the United States has become both harder and easier to forecast.

• Easier: inflation is much less volatile than it was in the 1970s and the root mean squared error of inflation forecasts has declined sharply since the mid-1980s.

• Harder: standard multivariate forecasting models do not a better job than simple naive models. The point was made by Atkeson and Ohanian (2001) (henceforth, AO), who found that, since 1984 in the U.S., backwards-looking Phillips curve forecasts have been inferior to a forecast of twelve-month inflation by its average rate over the previous twelve months (naive or random walk forecast).

• Relevance of the topic. Change in terms of forecasting properties can signal changes in the structure of the economy. This can be taken as evidence that suggests that some relations have changed

• What relations? Structural models should be employed (next part of the course).

10.1 U.S. Inflation forecasts: facts and puzzles

10.1.1 Data

• GDP price index inflation (π).

• Robustness analysis done using personal consumption expenditure deflator for core items (PCEcore), the personal consumption expenditure deflator for all items (PCE-all), and the consumer price index (CPI, the official CPI-U).

• Real activity variables: the unemployment rate (u), log real GDP (y), the capacity utilization rate, building permits, and the Chicago Fed National Activity Index (CFNAI)

• Quarterly data. Quarterly values for monthly series are averages of the three months in the quarter.

• The full sample is from 1960:I through 2004:IV.

10.1.2 Forecasting models

- Two univariate models and one multivariate forecasting models.

- Let $\pi_t = 400\log(p_t/p_{t-1})$ where p_t is the quarterly price index and let the h-period average inflation be $\pi_t^h = (1/h)\sum_{i=0}^{h-1}\pi_{t-i}$. Let $\pi_{t+h|t}^h$ be the forecast of π_{t+h}^h using information up to date t.

10.1.3 AR(r)

• Forecasts made using a univariate autoregression with r lags. r is estimated using the Akaike Information Criterion (AIC).

• Multistep forecasts are computed by the direct method: projecting h-period ahead inflation on r lags

• The h-step ahead AR(r) forecast was computed using the model

$$\pi_{t+h}^{h} - \pi_t = \mu^h + \alpha^h(L)\Delta\pi_t + v_t^h \tag{11}$$

where

1. μ^h is a constant
2. $\alpha^h(L)$ is a polynomial in the lag operator
3. v_t^h is the h-step ahead error term

• The number of lags is chosen according to the Akaike Information Criterion (AIC) meaning that r is such that

$$AIC = T\log(\sum_{t=1}^{T}\hat{\varepsilon}_t^2) + 2r$$

is minimum. An alternative criterion is the Bayesian Information Criterion (BIC)

$$BIC = T \log(\sum_{t=1}^{T} \hat{\varepsilon}_t^2) + r \log(p)$$

10.1.4 AO. Atkeson-Ohanian (2001)

AO forecasted the average four-quarter rate of inflation as the average rate of inflation over the previous four quarters. The AO forecast is

$$\pi^h_{t+h|t} = \pi^4_t = \frac{1}{4}(\pi_t + \pi_{t-1} + \pi_{t-2} + \pi_{t-3})$$

10.1.5 Backwards-looking Phillips curve (PC)

The PC forecasts are computed by adding a predictor to (11) to form the autoregressive distributed lag (ADL) specification,

$$\pi_{t+h}^h - \pi_t = \mu^h + \alpha^h(L)\Delta\pi_t + \beta^h xgap_t + \delta^h(L)\Delta x_t + v_t^h \tag{12}$$

where

1. μ^h is a constant
2. $\alpha^h(L)$, $\delta^h(L)$, is a polynomial in the lag operator (lag length chosen using AIC)
3. $xgap_t$ is the gap variable (deviations from a low pass filter) based on the variable x_t
4. v_t^h is the h-step ahead error term

• The PC forecast using $u_t = xgap_t = x_t$ and $\Delta u_t = \Delta x_t$ is called $PC - u$.

• The forecasts $PC - \Delta u$, $PC - \Delta y$, $PC - \Delta CapUtil$, $PC - \Delta Permits$, $PC - CFNAI$ omit the gap variable and only include stationary predictors Δu, Δy, $\Delta CapUtil$, $\Delta Permits$, $CFNAI$.

10.2 Out-of-sample methodology

• The forecasts were computed using the pseudo out-of-sample forecast methodology: that is, for a forecast made at date t, all estimation, lag length selection, etc. was performed using only data available through date t.

• The forecasts are recursive, so that forecasts at date t are based on all the data (beginning in 1960:I) through date t.

• The period 1960-1970 was used for initial parameter estimation. The forecast period 1970:I 2004:IV was split into the two periods 1970:I 1983:IV and 1984:I 2004:IV.

10.3 Results

Table 1
Pseudo Out-of-Sample Forecasting Results for GDP Inflation

Multivariate forecasting model: $\pi_{t+h}^{h} - \pi_t = \mu^h + \alpha^h(\text{B})\Delta\pi_t + \beta^h xgap_t + \delta^h(\text{B})\Delta x_t + u_t^h$

	1970:I – 1983:IV				1984:I – 2004:IV				$\frac{RMSFE_{84-04}^{h=4}}{RMSFE_{70-83}^{h=4}}$
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8	
AR(AIC) RMSFE	1.72	1.75	1.89	2.38	0.78	0.68	0.62	0.73	
Relative MSFEs									
AR(AIC)	1.00	1.00	1.00	1.00	1.00	**1.00**	1.00	1.00	0.33
AO	1.95	1.57	1.06	1.00	1.22	1.10	**0.89**	**0.84**	0.30
PC-u	**0.85**	0.92	0.88	0.61	**0.95**	1.11	1.48	1.78	0.42
PC-Δu	0.87	**0.87**	0.86	0.64	1.06	1.27	1.83	2.21	0.48
PC-$ugap^{1\text{-}sided}$	0.88	0.99	0.98	0.87	1.06	1.29	1.84	2.39	0.45
PC-Δy	0.99	1.06	0.93	0.58	1.05	1.06	1.23	1.53	0.37
PC-$ygap^{1\text{-}sided}$	0.94	0.97	0.99	0.78	0.97	0.97	1.25	1.55	0.37
PC-CapUtil	0.85	0.88	**0.79**	**0.55**	0.95	1.01	1.35	1.52	0.43
PC-ΔCapUtil	1.02	1.00	0.87	0.64	1.03	1.10	1.30	1.51	0.40
PC-Permits	0.93	1.02	0.98	0.78	1.08	1.23	1.31	1.52	0.38
PC-ΔPermits	1.02	1.04	0.99	0.86	1.00	1.00	1.00	1.02	0.33
PC-CFNAI	.	.	.	.	1.11	1.27	1.86	2.25	.

• The RMSFE of forecasts of GDP inflation has declined and the magnitude of this reduction is striking. In this sense inflation has become easier to forecast

• The relative performance of the Phillips curve forecasts deteriorated substantially from the first period to the second. This deterioration of Phillips curve forecasts is found for all the activity predictors.

• The AO forecast substantially improves upon the AR(AIC) and Phillips curve forecasts at the four- and eight-quarter horizons in the 1984-2004 period, but not at shorter horizons and not in the first period.

$\Rightarrow$ Substantial changes in the univariate inflation process and in the bivariate process of inflation and its activity-based predictors.

“Unpredictability and Macreconomic Stability”
By D’Agostino, A., D. Giannone and P. Surico

11 Unpredictability and Macreconomic Stability

• D'Agostino Giannone and Surico extend the result for inflation to other economic activity variables: the ability to predict several measures of real activity declined remarkably, relative to naive forecasts, since the mid-1980s.

• The fall in the predictive ability is a common feature of many forecasting models including those used by public and private institutions.

• The forecasts for output (and also inflation) of the Federal Reserves Greenbook and the Survey of Professional Forecasters (SPF) are significantly more accurate than a random walk only before 1985. After 1985, in contrast, the hypothesis of equal predictive ability between naive random walk forecasts and the predictions of those institutions is not rejected for all horizons but the current quarter.

• The decline in predictive accuracy is far more pronounced for institutional forecasters and methods based on large information sets than for univariate specifications.

• The fact that larger models are associated with larger historical changes suggests that the main sources of the decline in predictability are the dynamic correlations between variables rather than the autocorrelations of output and inflation.

11.1 Data

• Forecasts for nine monthly key macroeconomic series: three price indices, four measures of real activity and two interest rates:

1. The three nominal variables are Producer Price Index (PPI), Consumer Price Index (CPI) and Personal Consumption Expenditure implicit Deflator (PCED).
2. The four forecasted measures of real activity are Personal Income (PI), Industrial Production (IP) index, Unemployment Rate (UR), and EMPloyees on non-farm Payrolls (EMP).
3. the interest rates are 3 month Treasury Bills (TBILL) and 10 year Treasury Bonds (TBOND).

• The data set consists of monthly observations from 1959:1 through 2003:12 on 131 U.S.macroeconomic time series including also the nine variables of interest.

11.2 Forecasting models

The model used are the following:

1. A Naive forecast model (N or RW).
2. Univariate AR, where the forecasts are based exclusively on the own past values of the variable of interest.
3. Factor augmented AR forecast (FAAR), in which the univariate models are augmented with common factors extracted from the whole panel of series.
4. Pooling of bivariate forecasts (POOL): for each variable the forecast is defined as the average of 130 forecasts obtained by augmenting the AR model with each of the remaining 130 variables in the data set.

11.3 Out-of-sample methodology

• Pseudo out-of-sample forecasts are calculated for each variable and method over the horizons h = 1, 3, 6, and 12 months.

• The pseudo out-of-sample forecasting period begins in January 1970 and ends in December 2003. Forecasts constructed at date T are based on models that are estimated using observations dated T and earlier.

• Forecast based on rolling samples using, at each point in time, observations over the most recent 10 years.

11.4 Results: full sample

Table 1: *Relative Mean Square Forecast Errors - Full Period*

Random Walk (absolute values)

hor(m)	PPI	CPI	PCED	PI	IP	UR	EMP	TBILL	TBOND
1	0.45	0.11	0.06	45.58	75.84	0.03	9.45	0.31	0.11
3	1.83	0.59	0.32	13.93	46.23	0.14	7.25	1.29	0.47
6	4.40	1.63	0.94	7.72	35.04	0.45	6.66	2.50	0.99
12	11.87	5.02	2.90	5.03	25.30	1.38	5.75	4.74	2.20

Method AR (relative to RW)

hor(m)	PPI	CPI	PCED	PI	IP	UR	EMP	TBILL	TBOND
1	0.96	0.83***	0.83***	1.22	0.86*	0.91	0.60***	0.98	0.92
3	1.03	0.88*	0.82**	1.09	0.86	0.81*	0.53***	1.10	1.10
6	1.00	0.84	0.82	1.08	0.94	0.88	0.61***	1.05	1.05
12	1.05	0.93	1.00	1.01	0.95	0.97	0.75***	1.20	1.03

Method FAAR (relative to RW)

hor(m)	PPI	CPI	PCED	PI	IP	UR	EMP	TBILL	TBOND
1	0.94	0.76***	0.78***	1.15	0.74***	0.72***	0.50***	0.93	0.95
3	0.91	0.71***	0.77**	0.93	0.64**	0.58***	0.39***	1.06	1.19
6	0.84	0.60***	0.75	0.90	0.63*	0.55***	0.43***	0.95	1.17
12	0.84	0.60*	0.83	0.94	0.63	0.64*	0.56***	1.05	1.26

Method POOL (relative to RW)

hor(m)	PPI	CPI	PCED	PI	IP	UR	EMP	TBILL	TBOND
1	0.94	0.80***	0.80***	1.18	0.80**	0.83***	0.56***	0.94	0.91
3	0.96	0.81***	0.78**	1.02	0.76**	0.73**	0.47***	1.08	1.12
6	0.92	0.72**	0.76*	1.00	0.80*	0.76*	0.54***	0.99	1.07
12	0.92	0.73*	0.85	0.93**	0.78**	0.84***	0.65	1.12	1.07

Notes: Asterisks denote model forecasts that are statistically more accurate than the Naive at 1% ($***$), 5% ($**$) and 10% ($*$) significance levels.

• For all prices and most real activity indicators, the forecasts based on large information are significantly more accurate than the Naive forecasts.

• The factor augmented model produces the most accurate predictions.

• Univariate autoregressive forecasts significantly improve on the naive models for EMP at all horizons and for CPI and PCED at one and three month horizons only. As far as interest rates are concerned, no forecasting model performs significantly better than the naive benchmark.

12 Results: sub samples - inflation

Table 2: *Relative MSFEs across Sub-Periods - Inflation*

PERIOD I: sub-sample 1971:1 - 1984:12					PERIOD II: sub-sample 1985:1 - 2002:12						CHANGE
Series: Producer Price Index											
hor	RW	AR	FAAR	POOL	hor	RW	AR	FAAR	POOL		Average
1	0.55	1.03	1.01	0.99	1	0.37	0.89*	0.87*	0.88***		7%
3	2.23	1.05	0.85	0.94	3	1.51	1.01	0.98	0.99**		20%
6	5.79	0.95	0.67	0.82**	6	3.31	1.08	1.08	1.07		34%
12	17.95	1.02	0.65	0.84	12	7.12	1.13	1.20	1.09		33%
Series: Consumer Price Index											
hor	RW	AR	FAAR	POOL	hor	RW	AR	FAAR	POOL		Average
1	0.17	0.83***	0.75***	0.78***	1	0.07	0.85*	0.77**	0.83***		5%
3	0.94	0.84*	0.61***	0.74***	3	0.31	0.99	0.93	0.96**		38%
6	2.85	0.78*	0.46***	0.65***	6	0.68	1.04	1.05	0.98*		83%
12	9.43	0.87	0.44***	0.64**	12	1.57	1.22	1.32	1.16		118%
Series: Personal Consumption Expenditure Deflator											
hor	RW	AR	FAAR	POOL	hor	RW	AR	FAAR	POOL		Average
1	0.08	0.73***	0.71***	0.71***	1	0.05	0.96	0.88**	0.93***		9%
3	0.50	0.72***	0.67**	0.68***	3	0.18	1.04	0.98	1.01		29%
6	1.63	0.72**	0.66*	0.66**	6	0.40	1.13	1.05	1.08		48%
12	5.52	0.92	0.75	0.77	12	0.85	1.37	1.27	1.27		59%

Notes: The column 'change' reads the percentage historical decline in predictability averaged across methods (excluding Naive). Asterisks denote model forecasts that are statistically more accurate than the Naive at 1% ($***$), 5% ($**$) and 10% ($*$) significance levels.

• For all lags except the first, result of AO confirmed, deterioration of the forecasting performance of inflation.

12.1 Results: sub samples - real activity

Table 3: *Relative MSFEs across Sub-Periods - Real Activity*

PERIOD I: sub-sample 1971:1 - 1984:12 | PERIOD II: sub-sample 1985:1 - 2002:12 | CHANGE

Series: Real Personal Income

hor	RW	AR	FAAR	POOL	hor	RW	AR	FAAR	POOL		Average
1	38.54	1.02	0.95	0.98	1	51.09	1.33	1.27	1.30		21%
3	17.15	1.01	0.86	0.94	3	11.41	1.19	1.01	1.12		14%
6	10.41	1.05	0.83	0.96	6	5.62	1.12	1.01	1.05		2%
12	6.92	0.97	0.84	0.87*	12	3.55	1.07	1.09	1.02		3%

Series: Industrial Production

hor	RW	AR	FAAR	POOL	hor	RW	AR	FAAR	POOL		Average
1	124.01	0.81*	0.65***	0.75**	1	38.14	0.97	0.95	0.92		14%
3	81.48	0.85	0.55**	0.73**	3	18.64	0.92	0.98	0.88		16%
6	61.42	0.94	0.49*	0.76*	6	14.41	0.97	1.11	0.95		34%
12	43.24	0.95	0.43**	0.72**	12	11.27	0.98	1.22	0.97		62%

Series: Unemployment Rate

hor	RW	AR	FAAR	POOL	hor	RW	AR	FAAR	POOL		Average
1	0.05	0.86	0.63***	0.78**	1	0.02	0.99	0.88*	0.94***		21%
3	0.25	0.79	0.52***	0.69**	3	0.06	0.91	0.79*	0.84**		18%
6	0.80	0.88	0.49***	0.75	6	0.17	0.85	0.75	0.80*		22%
12	2.42	0.99	0.56**	0.82**	12	0.56	0.93	0.90	0.89		41%

Series: Employees on Nonfarm Payrolls

hor	RW	AR	FAAR	POOL	hor	RW	AR	FAAR	POOL		Average
1	16.37	0.65***	0.51***	0.60***	1	4.04	0.42***	0.45***	0.40***		4%
3	12.39	0.60**	0.41***	0.53***	3	3.23	0.31***	0.34***	0.29***		-1%
6	11.16	0.70**	0.42***	0.60**	6	3.14	0.37**	0.44*	0.36*		-3%
12	9.21	0.82***	0.49***	0.69***	12	3.05	0.58**	0.72	0.56		8%

Notes: see Table 2.

- Little change in the structure of univariate models for real activity.

- The relative MSFEs of FAAR and POOL suggest that important changes have occurred in the relationship between output and other macroeconomic variables.

- The decline in predictability does not seem to extend to the labor market, especially at short horizons. The forecasts of the employees on nonfarm payrolls are associated with the smallest percentage changes across subsamples.

12.2 Results: sub samples - interest rates

Table 4: *Relative MSFEs across Sub-Periods - Interest Rates*

PERIOD I: sub-sample 1971:1 - 1984:12					PERIOD II: sub-sample 1985:1 - 2002:12					CHANGE
Series: 3 Months Treasury Bills										
hor	RW	AR	FAAR	POOL	hor	RW	AR	FAAR	POOL	Average
1	0.64	1.00	0.94	0.95	1	0.05	0.84	0.87	0.81***	-10%
3	2.59	1.12	1.05	1.10	3	0.27	0.98	1.16	0.94**	-5%
6	4.63	1.06	0.88	0.98	6	0.83	1.03	1.25	1.01	11%
12	7.63	1.27	0.93	1.14	12	2.47	1.04	1.34	1.06	8%
Series: 10 Years Treasury Bonds										
hor	RW	AR	FAAR	POOL	hor	RW	AR	FAAR	POOL	Average
1	0.17	0.95	0.96	0.94	1	0.07	0.88**	0.92	0.87***	-9%
3	0.68	1.17	1.21	1.18	3	0.31	1.00	1.15	1.02	-11%
6	1.28	1.07	1.12	1.09	6	0.77	1.02	1.23	1.05	3%
12	2.57	1.04	1.12	1.06	12	1.91	1.01	1.42	1.09	7%

Notes: see Table 2.

- In the second sample the AR, FAAR and P OOL methods produce more accurate forecasts than the RW at one month horizon.

- Possible interpretation: increased predictability of the FED due to a better communication strategy.

12.3 Results: private and institutional forecasters

• The predictions for output and its deflator from two large forecasters representing the private sector and the policy institutions are considered.

• The survey was introduced by the American Statistical Association and the National Bureau of Economic Research and is currently maintained by the Philadelphia Fed. The SPF refers to quarterly measures and is conducted in the middle of the second month of each quarter (here the median of the individual forecasts is considered)

• The forecasts of the Greenbook are prepared by the Board of Governors at the Federal Reserve for the meetings of the Federal Open Market Committee (FOCM), which takes place roughly every six weeks.

• Four forecast horizons ranging from 1 to 4 quarters.

• The measure of output is Gross National Product (GNP) until 1991 and Gross Domestic Product (GDP) from 1992 onwards.

• The evaluation sample begins in 1975 (prior to this date the Greenbook forecasts were not always available up to the fourth quarter horizon).

12.4 Results: private and institutional forecasters - inflation

Table 5: *Relative MSFEs of Institutional Forecasters - Inflation*

hor(q)	RW	Fed's Green Book(GB)/RW	Survey of Professional Forecasters(SPF)/RW
FULL SAMPLE: 1975:1 1999:4			
1	0.26	0.35***	0.37***
2	0.79	0.30**	0.36**
3	1.57	0.29*	0.37
4	2.51	0.32	0.46
PERIOD I: sub-sample 1975:1 - 1984:4			
1	0.54	0.30***	0.27***
2	1.72	0.21**	0.24**
3	3.51	0.21**	0.25*
4	5.69	0.23*	0.32*
PERIOD II: sub-sample 1985:1 - 1999:4			
1	0.08	0.58**	0.82
2	0.17	0.93	1.15
3	0.28	0.97	1.39
4	0.39	1.18	1.82

Notes: Asterisks denote rejection of the null hypothesis of equal predictive accuracy between each model and the RW at 1% ($***$), 5% ($**$) and 10% ($*$) significance levels.

12.5 Results: private and institutional forecasters - real activity

Table 6: *Relative MSFEs of Institutional Forecasters - Output*

FULL SAMPLE: 1975:1 1999:4

hor(q)	RW	Fed's Green Book(GB)/RW	Survey of Professional Forecasters(SPF)/RW
1	12.59	0.44**	0.51**
2	9.11	0.49**	0.46**
3	7.45	0.48**	0.50***
4	6.49	0.51**	0.51***

PERIOD I: sub-sample 1975:1 - 1984:4

hor(q)	RW	Fed's Green Book(GB)/RW	Survey of Professional Forecasters(SPF)/RW
1	25.82	0.37**	0.45**
2	19.01	0.44**	0.41**
3	15.39	0.40***	0.45***
4	13.18	0.42***	0.46***

PERIOD II: sub-sample 1985:1 - 1999:4

hor(q)	RW	Fed's Green Book(GB)/RW	Survey of Professional Forecasters(SPF)/RW
1	3.77	0.73	0.77
2	2.51	0.77	0.70
3	2.15	0.85	0.73
4	2.03	0.89	0.74

Notes: see Table 5.

"The Return of the Wage Phillips Curve"

13 The Return of the Wage Phillips Curve

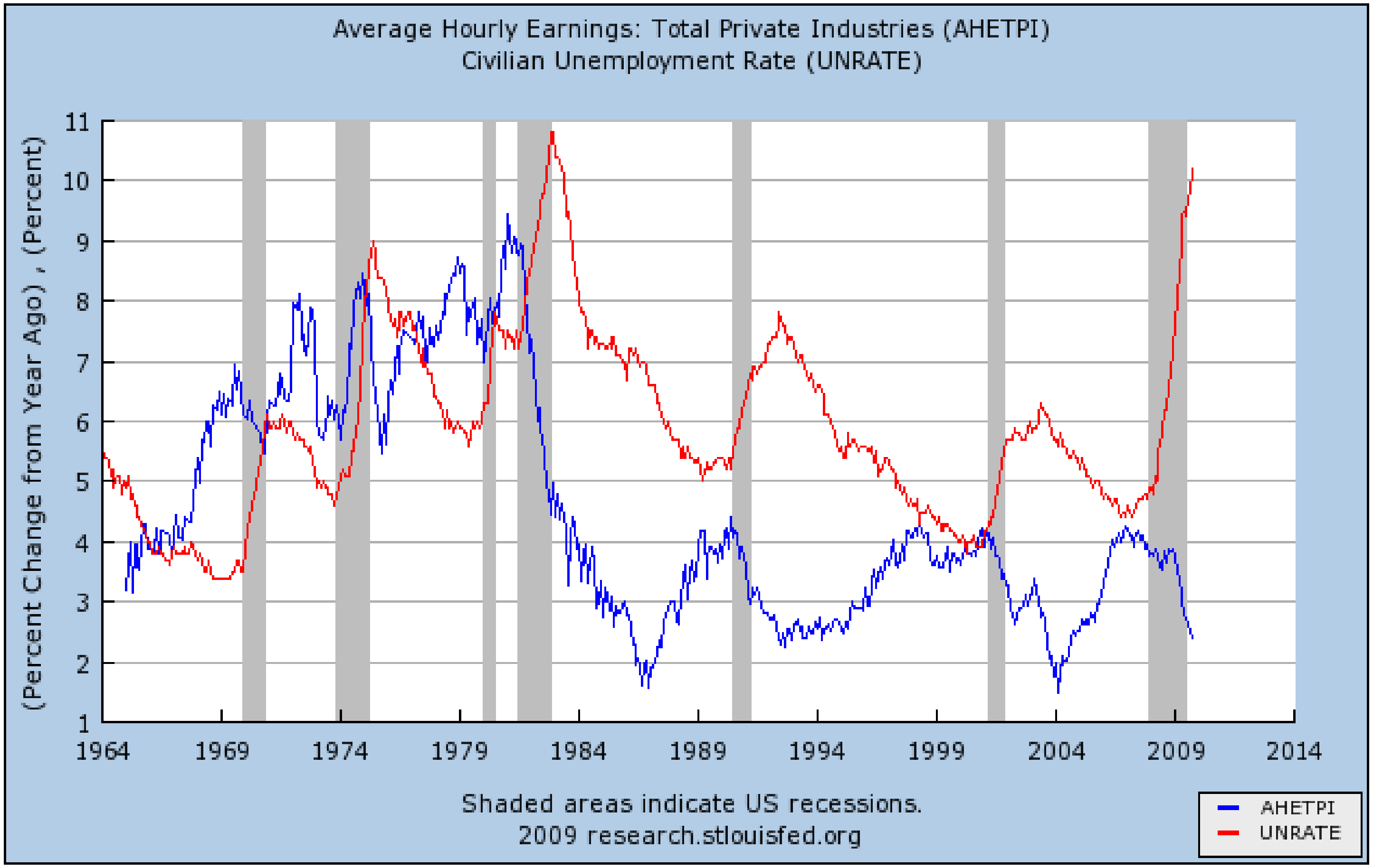
• Previous evidence has been taken as a motivation to dismiss the Phillips curve as a theoretical concept.

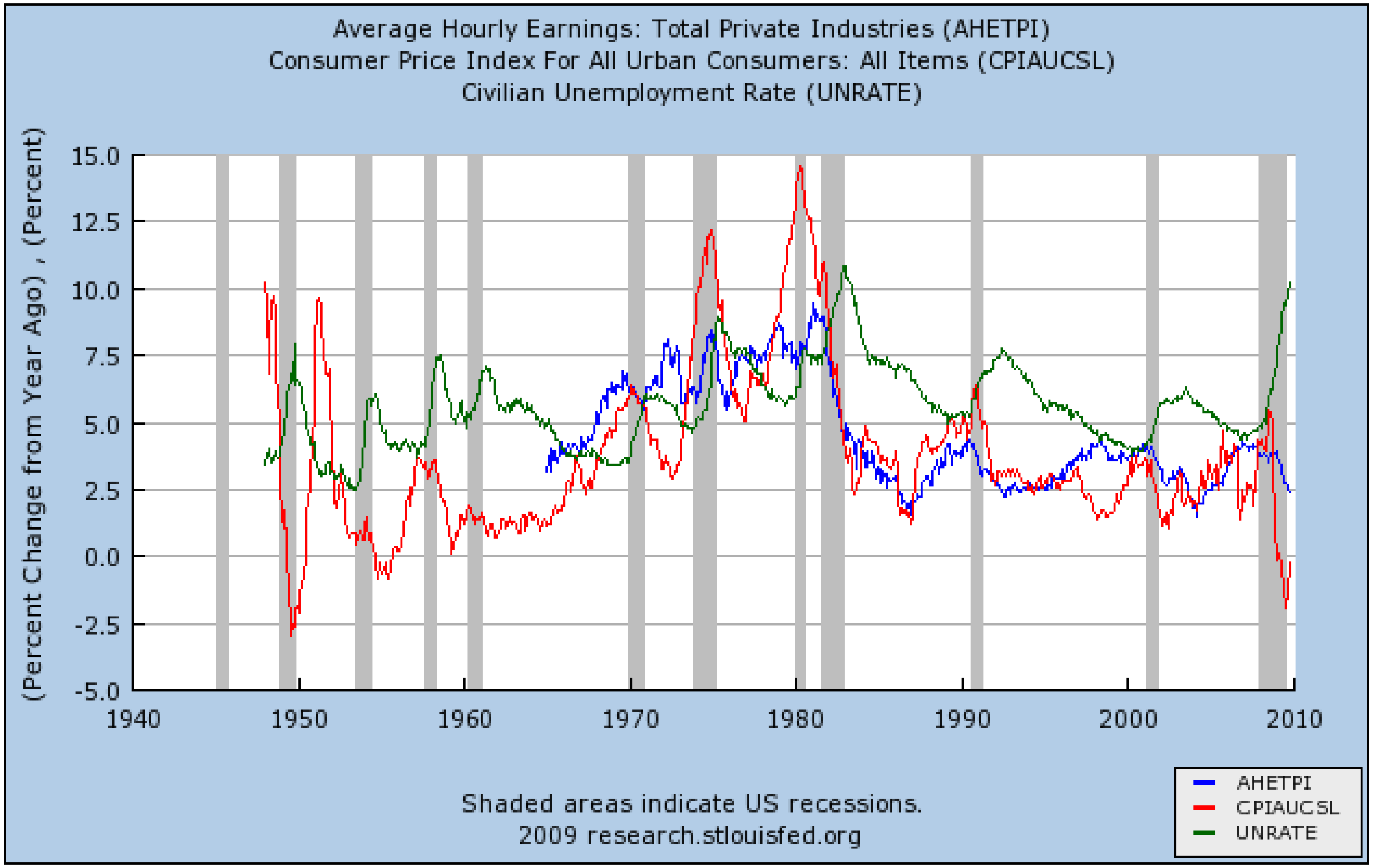
• Danger with that interpretation.

• In 1958 Phillips uncovered an inverse relation between wage rate inflation and unemployment.

• The focus however in recent years has been shifted to price inflation

13.1 Back to the origins




Average Hourly Earnings: Total Private Industries (AHETPI)
Consumer Price Index For All Urban Consumers: All Items (CPIAUCSL)
Civilian Unemployment Rate (UNRATE)
(Percent Change from Year Ago) , (Percent)
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
1940
1950
1960
1970
1980
1990
2000
2010
Shaded areas indicate US recessions.
2009 research.stlouisfed.org
AHETPI
CPIAUCSL
UNRATE

13.2 Results

forecast horizon	RMSE VAR/RW	RMSE AR/RW	% gain using var
1	0.2252	0.2048	9.976
4	0.3642	0.3976	-8.4208
8	0.4892	0.6110	-19.9406
12	0.544	0.6646	-18.1371
16	0.5356	0.6157	-13.0069
18	0.5259	0.5914	-11.0704

- Phillips curve still (now more than then) characterize dynamics of wage growth and unemployment.

- Crucial question: what has changed in the relation between prices and wages?

5: MULTIVARATE STATIONARY PROCESSES

1 Some Preliminary Definitions and Concepts

Random Vector: A vector $X = (X_1, ..., X_n)$ whose components are scalar-valued random variables on the same probability space.

Vector Random Process: A family of random vectors $\{X_t, t \in T\}$ defined on a probability space, where T is a set of time points. Typically $T = \mathbb{R}$, $T = \mathbb{Z}$ or $T = \mathbb{N}$, the sets or real, integer and natural numbers, respectively.

Time Series Vector: A particular realization of a vector random process.

Matrix of polynomial in the lag operator: $\Phi(L)$ if its elements are polynomial in the lag operator, i.e.

$$\Phi(L) = \begin{pmatrix} 1 & -0.5L \\ L & 1+L \end{pmatrix} = \Phi_0 + \Phi_1 L$$

where

$$\Phi_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Phi_1 = \begin{pmatrix} 0 & -0.5 \\ 1 & 1 \end{pmatrix}, \quad \Phi_{j>1} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

When applied to a vector X_t we obtain

$$\Phi(L)X_t \;=\; \begin{pmatrix} 1 & -0.5L \\ L & 1+L \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} X_{1t} - 0.5X_{2t-1} \\ X_{1t-1} + X_{2t} + X_{2t-1} \end{pmatrix}$$

The inverse is matrix such that $\Phi(L)^{-1}\Phi(L) = I$. Suppose $\Phi(L) = (I - \Phi_1 L)$. Its inverse $\Phi(L)^{-1} = A(L)$ is a matrix such that $(A_0 + A_1 L + A_2 L^2 + ...)\Phi = I$. That is

$$
\begin{aligned}
A_0 &= I \\
A_1 - \Phi_1 &= 0 \qquad \Rightarrow A_1 = \Phi_1 \\
A_2 - A_1\Phi_1 &= 0 \qquad \Rightarrow A_2 = \Phi^2 \\
&\vdots \\
A_k - A_{k-1}\Phi_1 &= 0 \qquad \Rightarrow A_k = \Phi^k
\end{aligned}
\tag{1}
$$

1.1 Covariance Stationarity

Let Y_t be a n-dimensional vector of time series, $Y_t' = [Y_{1t}, ..., Y_{nt}]$. Then Y_t is covariance (weakly) stationary if $E(Y_t) = \mu$, and the *autocovariance* matrix $\Gamma_j = E(Y_t - \mu)(Y_{t-j} - \mu)'$ for all t, j, that is are independent of t and both finite.

- Stationarity of each of the components of Y_t does not imply stationarity of the vector Y_t. Stationarity in the vector case requires that the components of the vector are stationary and costationary.
- Although $\gamma_j = \gamma_{-j}$ for a scalar process, the same is not true for a vector process. The correct relation is

$$\Gamma_j = \Gamma_{-j}'$$

Example: $n = 2$ and $\mu = 0$

$$\begin{aligned}
\Gamma_1 &= \begin{pmatrix} E(Y_{1t}Y_{1t-1}) & E(Y_{1t}Y_{2t-1}) \\ E(Y_{2t}Y_{1t-1}) & E(Y_{2t}Y_{2t-1}) \end{pmatrix} \\
&= \begin{pmatrix} E(Y_{1t+1}Y_{1t}) & E(Y_{1t+1}Y_{2t}) \\ E(Y_{2t+1}Y_{1t}) & E(Y_{2t+1}Y_{2t}) \end{pmatrix} \\
&= \begin{pmatrix} E(Y_{1t}Y_{1t+1}) & E(Y_{1t}Y_{2t+1}) \\ E(Y_{2t}Y_{1t+1}) & E(Y_{2t}Y_{2t+1}) \end{pmatrix}' = \Gamma_{-1}'
\end{aligned}$$

2 Vector Moving average processes

2.1 White Noise (WN)

A n-dimensional vector white noise $\epsilon_t' = [\epsilon_{1t}, ..., \epsilon_{nt}] \sim WN(0, \Omega)$ is such if $E(\epsilon_t) = 0$ and $\Gamma_k = \Omega$ (Ω a symmetric positive definite matrix) if $k = 0$ and 0 if $k \neq 0$. If $\epsilon_t,\ \epsilon_\tau$ are independent the process is an independent vector White Noise (i.i.d). If also $\epsilon_t \sim N$ the process is a Gaussian WN.

Important: A vector whose components are white noise is not necessarily a white noise. Example: let u_t be a scalar white noise and define $\epsilon_t = (u_t, u_{t-1})'$. Then $E(\epsilon_t\epsilon_t') = \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}$ and $E(\epsilon_t\epsilon_{t-1}') = \begin{pmatrix} 0 & 0 \\ \sigma_u^2 & 0 \end{pmatrix}$.

2.2 Vector Moving Average (VMA)

Given the n-dimensional vector White Noise ϵ_t a vector moving average of order q is defined as

$$Y_t = \mu + \epsilon_t + C_1\epsilon_{t-1} + ... + C_q\epsilon_{t-q}$$

where C_j are $n \times n$ matrices of coefficients.

VMA(1)

Let us consider the VMA(1)

$$Y_t = \mu + \epsilon_t + C_1\epsilon_{t-1}$$

with $\epsilon_t \sim WN(0, \Omega)$, μ is the mean of Y_t. The variance of the process is given by

$$\begin{aligned}\Gamma_0 &= E[(Y_t - \mu)(Y_t - \mu)'] \\ &= \Omega + C_1\Omega C_1'\end{aligned}$$

with autocovariances

$$\Gamma_1 = C_1\Omega, \quad \Gamma_{-1} = \Omega C_1', \quad \Gamma_j = 0 \ \text{ for } \ |j| > 1$$

The VMA(q)

Let us consider the VMA(q)

$$Y_t = \mu + \epsilon_t + C_1\epsilon_{t-1} + ... + C_q\epsilon_{t-q}$$

with $\epsilon_t \sim WN(0,\Omega)$, μ is the mean of Y_t. The variance of the process is given by

$$\begin{aligned}\Gamma_0 &= E[(Y_t-\mu)(Y_t-\mu)'] \\ &= \Omega + C_1\Omega C_1' + C_2\Omega C_2' + ... + C_q\Omega C_q'\end{aligned}$$

with autocovariances

$$\begin{aligned}\Gamma_j &= C_j\Omega + C_{j+1}\Omega C_1' + C_{j+2}\Omega C_2' + ... + C_q\Omega C_{q-j}' \quad for \quad j = 1,2,...,q \\ \Gamma_j &= \Omega C_{-j}' + C_1\Omega C_{-j+1}' + C_2\Omega C_{-j+2}' + ... + C_{q+j}\Omega C_q' \quad for \quad j = -1,-2,...,-q \\ \Gamma_j &= 0 \quad for \quad |j| > q\end{aligned}$$

The VMA(∞)

A useful process, as we will see, is the VMA(∞)

$$Y_t = \mu + \sum_{j=0}^{\infty} C_j \varepsilon_{t-j} \tag{2}$$

the process can be thought as the limiting case of a VMA(q) (for $q \to \infty$). Recall the previous result the process converges in mean square if $\{C_j\}$ is absolutely summable.

Proposition (10.2H). *Let Y_t be an $n \times 1$ vector satisfying*

$$Y_t = \mu + \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}$$

where ε_t is a vector WN with $E(\varepsilon_{t-j}) = 0$ and $E(\varepsilon_t \varepsilon'_{t-j} = \Omega$ for $j = 0$ and zero otherwise and $\{C_j\}_{j=0}^{\infty}$ is absolutely summable. Let Y_{it} denote the ith element of Y_t and μ_i the ith element of μ. Then

(a) The autocovariance between the ith variable at time t and the jth variable at time s periods earlier, $E(Y_{it} - \mu_i)(Y_{jt-s} - \mu_j)$ exists and is given by the row i column j element of

$$\Gamma_s = \sum_{v=0}^{\infty} C_{s+v} \Omega C'_s$$

for $s = 0, 1, 2, ...$

(b) The sequence of matrices $\{\Gamma_s\}_{s=0}^{\infty}$ *is absolutely summable.*

If furthermore $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ *is an i.i.d. sequence with* $E|\varepsilon_{i_1 t}\varepsilon_{i_2 t}\varepsilon_{i_3 t}\varepsilon_{i_4 t}| \leq \infty$ *for* $i_1, i_2, i_3, i_4 = 1, 2, ..., n$*, then also*

(c) $E|Y_{i_1 t_1} Y_{i_2 t_2} Y_{i_3 t_3} Y_{i_4 t_4}| \leq \infty$ *for all* t_1, t_2, t_3, t_4

(d) $(1/T)\sum_{t=1}^{T} y_{it} y_{jt-s} \xrightarrow{p} E(y_{it} y_{jt-s})$*, for* $i, j = 1, 2, ..., n$ *and for all* s

Implications:

1. Result (a) implies that the second moments of a $MA(\infty)$ with absolutely summable coefficients can be found by taking the limit of the autocovariance of an $MA(q)$.
2. Result (b) ensures ergodicity for the mean
3. Result (c) says that Y_t has bounded fourth moments
4. Result (d) says that Y_t is ergodic for second moments

Note: the vector $MA(\infty)$ representation of a stationary VAR satisfies the absolute summability condition so that assumption of the previous proposition hold.

2.3 Invertible and fundamental VMA

Invertibility The VMA is invertible i.e. it possesses a VAR representation, if and only if the determinant of $C(L)$ vanishes only outside the unit circle, i.e. if $det(C(z)) \neq 0$ for all $|z| \leq 1$.

Example Consider the process

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} 1 & L \\ 0 & \theta - L \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$det(C(z)) = \theta - z$ which is zero for $z = \theta$. The process is invertible if and only if $|\theta| > 1$.

Fundamentalness The VMA is fundamental if and only if the $det(C(z)) \neq 0$ for all $|z| < 1$. In the previous example the process is fundamental if and only if $|\theta| \geq 1$. In the case $|\theta| = 1$ the process is fundamental but noninvertible.

Provided that $|\theta| > 1$ the MA process can be inverted and the shock can be obtained as a combination of present and past values of Y_t. That is the VAR (Vector Autoregressive) representation can be recovered. The representation will entail infinitely many lags of Y_t with absolutely summable coefficients, so that the process converges in mean square.

Considering the above example

$$\begin{pmatrix} 1 & -\frac{L}{\theta-L} \\ 0 & \frac{1}{\theta-L} \end{pmatrix} \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

or

$$\begin{aligned} Y_{1t} + \frac{L}{\theta}\frac{1}{1-\frac{1}{\theta}L}Y_{2t} &= \varepsilon_{1t} \\ \frac{1}{\theta}\frac{1}{1-\frac{1}{\theta}L}Y_{2t} &= \varepsilon_{2t} \end{aligned} \tag{3}$$

2.4 Wold Decomposition

Any zero-mean stationary vector process Y_t admits the following representation

$$Y_t = C(L)\varepsilon_t + \mu_t \tag{4}$$

where $C(L)\epsilon_t$ is the stochastic component with $C(L) = \sum_{i=0}^{\infty} C_i L^i$ and μ_t the purely deterministic component, the one perfectly predictable using linear combinations of past Y_t.

If $\mu_t = 0$ the process is said *regular*. Here we only consider regular processes.

(4) represents the *Wold representation* of Y_t which is unique and for which the following properties hold:

(b) ϵ_t is the innovation for Y_t, i.e. $\epsilon_t = Y_t - \text{Proj}(Y_t|Y_{t-1}, Y_{t-1}, ...)$,i.e. the shock is fundamental.

(b) ϵ_t is White noise, $E\epsilon_t = 0$, $E\epsilon_t\epsilon'_\tau = 0$, for $t \neq \tau$, $E\epsilon_t\epsilon'_t = \Omega$

(c) The coefficients are square summable $\sum_{j=0}^{\infty} \|C_j\|^2 < \infty$.

(d) $C_0 = I$

- The result is very powerful since holds for any covariance stationary process.

- However the theorem does not implies that (4) is the *true* representation of the process. For instance the process could be stationary but non-linear or non-invertible.

2.5 Other fundamental MA(∞) Representations

• It is easy to extend the Wold representation to the general class of invertible MA(∞) representations. For any non singular matrix R of constant we define $u_t = R^{-1}\epsilon_t$ and we have

$$\begin{aligned} Y_t &= C(L)Ru_t \\ &= D(L)u_t \end{aligned}$$

where $u_t \sim WN(0, R^{-1}\Omega R^{-1\prime})$.

• Notice that all these representations obtained as linear combinations of the Wold representations are fundamental. In fact, $det(C(L)R) = det(C(L))det(R)$. Therefore if $det(C(L)R) \neq 0 \ \forall |z| < 1$ so will $det(C(L)R)$.

3 VAR: representations

• Every stationary vector process Y_t admits a Wold representation. If the MA matrix of lag polynomials is invertible, then a unique VAR exists.

• We define $C(L)^{-1}$ as an $(n \times n)$ lag polynomial such that $C(L)^{-1}C(L) = I$; i.e. when these lag polynomial matrices are matrix-multiplied, all the lag terms cancel out. This operation in effect converts lags of the errors into lags of the vector of dependent variables.

• Thus we move from MA coefficient to VAR coefficients. Define $A(L) = C(L)^{-1.}$ Then given the (invertible) MA coefficients, it is easy to map these into the VAR coefficients:

$$\begin{aligned} Y_t &= C(L)\epsilon_t \\ A(L)Y_t &= \epsilon_t \end{aligned} \tag{5}$$

where $A(L) = A_0 - A_1L^1 - A_2L^2 - ...$ and A_j for all j are $(n \times n)$ matrices of coefficients.

• To show that this matrix lag polynomial exists and how it maps into the coefficients in $C(L)$, note that by assumption we have the identity

$$(A_0 - A_1L^1 - A_2L^2 - ...)(I + C_1L^1 + C_2L^2 + ...) = I$$

After distributing, the identity implies that coefficients on the lag operators must be zero, which implies the following recursive solution for the VAR coefficients:

$$\begin{aligned} A_0 &= I \\ A_1 &= A_0 C_1 \\ A_k &= A_0 C_k + A_1 C_k + ... + A_{k-1} C_1 \end{aligned}$$

• As noted, the VAR is of infinite order (i.e. infinite number of lags required to fully represent joint density).

• In practice, the VAR is usually restricted for estimation by truncating the lag-length. Recall that the AR coefficients are absolutely summable and vanish at long lags.

pth-order vector autoregression VAR(p). A VAR(p) is given by

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \epsilon_t \tag{6}$$

Note: Here we are considering zero mean processes. In case the mean of Y_t is not zero we should add a constant in the VAR equations.

VAR(1) representation Any VAR(p) can be rewritten as a VAR(1). To form a VAR(1) from the general model we define: $\mathbf{e}_t' = [\epsilon, 0, ..., 0]$, $\mathbf{Y}_t' = [Y_t', Y_{t-1}', ..., Y_{t-p+1}']$

$$\mathbf{A} = \begin{pmatrix} A_1 & A_2 & ... & A_{p-1} & A_p \\ I_n & 0 & ... & 0 & 0 \\ 0 & I_n & ... & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & ... & ... & I_n & 0 \end{pmatrix}$$

Therefore we can rewrite the VAR(p) as a VAR(1)

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{e}_t$$

This is also known as the companion form of the VAR(p)

SUR representation The VAR(p) can be stacked as

$$\mathbf{Y} = \mathbf{X}\Gamma + \mathbf{u}$$

where $\mathbf{X} = [X_1, ..., X_T]'$, $X_t = [Y'_{t-1}, Y'_{t-2}..., Y'_{t-p}]'$ $\mathbf{Y} = [Y_1, ..., Y_T]'$, $\mathbf{u} = [\epsilon_1, ..., \epsilon_T]'$ and $\Gamma = [A_1, ..., A_p]'$

Vec representation Let vec denote the stacking columns operator, i.e $X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ X_{31} & X_{32} \end{pmatrix}$ then

$$vec(X) = \begin{pmatrix} X_{11} \\ X_{21} \\ X_{31} \\ X_{12} \\ X_{22} \\ X_{32} \end{pmatrix}$$

Let $\gamma = vec(\Gamma)$, then the VAR can be rewritten as

$$Y_t = (I_n \otimes X'_t)\gamma + \epsilon_t$$

4 VAR: Stationarity

4.1 Stability and stationarity

• Consider the VAR(1)

$$Y_t = \mu + AY_{t-1} + \varepsilon_t$$

Substituting backward we obtain

$$\begin{aligned}
Y_t &= \mu + AY_{t-1} + \varepsilon_t \\
&= \mu + A(\mu + AY_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= (I + A)\mu + A^2 Y_{t-2} + A\varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
Y_t &= (I + A + ... + A^{j-1})\mu + A^j Y_{t-j} + \sum_{i=0}^{j-1} A^i \varepsilon_{t-i}
\end{aligned}$$

• The eigenvalues of A, λ, solve $\det(A - I\lambda) = 0$. If all the eigenvalues of A are smaller than one in modulus the sequence A^i, $i = 0, 1, ...$ is absolutely summable. Therefore

1. the infinite sum $\sum_{i=0}^{j-1} A^i \varepsilon_{t-i}$ exists in mean square;
2. $(I + A + ... + A^{j-1})\mu \to (I - A)^{-1}$ and $A^j \to 0$ as j goes to infinity.

Therefore if the eigenvalues are smaller than one in modulus then Y_t has the following representation

$$Y_t = (I - A)^{-1}\mu + \sum_{i=0}^{\infty} A^i \varepsilon_{t-i}$$

• Note that the the eigenvalues correspond to the reciprocal of the roots of the determinant of $A(z) = I - Az$. A VAR(1) is called *stable* if

$$\det(I - Az) \neq 0 \;\text{ for }\; |z| \leq 1.$$

Stability A VAR(p) is stable if and aonly if that all the eigenvalues of $\mathbf{A}$ (the AR matrix of the companion form of Y_t) are smaller than one in modulus, or equivalently if and only if

$$\det(I - A_1 z - A_2 z^2, ..., A_p z^p) \neq 0 \;\text{ for }\; |z| \leq 1.$$

A condition for stationarity: A stable VAR process is stationary.

• Notice that the converse is not true. An unstable process can be stationary.

4.2 Back the Wold representation

• If the VAR is stationary Y_t has the following Wold representation

$$Y_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} C_j \epsilon_t$$

where the sequence $\{C_j\}$ is absolutely summable, $\sum_{j=0}^{\infty} |C_j| < \infty$

• How can we find it? Let us rewrite the VAR(p) as a VAR(1)

• We know how to find the MA(∞) representation of a stationary AR(1). We can proceed similarly for the VAR(1). Substituting backward in the companion form we have

$$\mathbf{Y}_t = \mathbf{A}^j \mathbf{Y}_{t-j} + \mathbf{A}^{j-1}\mathbf{e}_{t-j+1} + ... + \mathbf{A}^1 \mathbf{e}_{t-1} + ... + \mathbf{e}_t$$

If conditions for stationarity are satisfied, the series $\sum_{i=1}^{\infty} \mathbf{A}^j$ converges and Y_t has an VMA(∞) representation in terms of the Wold shock $\mathbf{e}_t$ given by

$$\begin{aligned} \mathbf{Y}_t &= (I - \mathbf{A}L)^{-1}\mathbf{e}_t \\ &= \sum_{i=1}^{\infty} \mathbf{A}^j \mathbf{e}_{t-j} \\ &= \mathbf{C}(L)\mathbf{e}_t \end{aligned}$$

where $\mathbf{C}_0 = \mathbf{A}_0 = I$, $\mathbf{C}_1 = \mathbf{A}_1, \mathbf{C}_2 = \mathbf{A}^2, ..., \mathbf{C}_k = \mathbf{A}^k$. C_j will be the $n \times n$ upper left matrix of $\mathbf{C}_j$.

5 VAR: second moments

Let us consider the companion form of a stationary (zero mean for simplicity) VAR(p) defined earlier

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{e}_t \tag{7}$$

The variance of $\mathbf{Y}_t$ is given by

$$\begin{aligned}\tilde{\mathbf{\Sigma}} &= E[(\mathbf{Y}_t)(\mathbf{Y}_t)'] \\ &= \mathbf{A}\tilde{\mathbf{\Sigma}}\mathbf{A}' + \tilde{\mathbf{\Omega}} \end{aligned} \tag{8}$$

a closed form solution to (7) can be obtained in terms of the *vec* operator. Let A, B, C be matrices such that the product ABC exists. A property of the *vec* operator is that

$$vec(ABC) = (C' \otimes A)vec(B)$$

Applying the vec operator to both sides of (7) we have

$$vec(\tilde{\mathbf{\Sigma}}) = (\mathbf{A} \otimes \mathbf{A})vec(\tilde{\mathbf{\Sigma}}) + vec(\tilde{\mathbf{\Omega}})$$

If we define $\mathcal{A} = (\mathbf{A} \otimes \mathbf{A})$ then we have

$$vec(\tilde{\mathbf{\Sigma}}) = (I - \mathcal{A})^{-1}vec(\tilde{\mathbf{\Omega}})$$

where

$$\tilde{\mathbf{\Gamma}}_0 = \tilde{\mathbf{\Sigma}} = \begin{pmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{p-1} \\ \Gamma_{-1} & \Gamma_0 & \cdots & \Gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{-p+1} & \Gamma_{-p+2} & \cdots & \Gamma_0 \end{pmatrix}$$

The variance $\Sigma = \Gamma_0$ of the original series Y_t is given by the first n rows and columns of $\tilde{\Sigma}$.

The jth autocovariance of $\mathbf{Y}_t$ (denoted $\tilde{\mathbf{\Gamma}}_j$) can be found by post multiplying (6) by $\mathbf{Y}_{t-j}$ and taking expectations:

$$E(\mathbf{Y}_t \mathbf{Y}_{t-j}) = \mathbf{A}E(\mathbf{Y}_t \mathbf{Y}_{t-j}) + E(\mathbf{e}_t \mathbf{Y}_{t-j})$$

Thus

$$\tilde{\mathbf{\Gamma}}_j = \mathbf{A}\tilde{\mathbf{\Gamma}}_{j-1}$$

or

$$\tilde{\mathbf{\Gamma}}_j = \mathbf{A}^j \tilde{\mathbf{\Gamma}}_0 = \mathbf{A}^j \tilde{\mathbf{\Sigma}}$$

The autocovariances Γ_j of the original series Y_t are given by the first n rows and columns of $\tilde{\Gamma}_j$ and are given by

$$\Gamma_h = A_1 \Gamma_{h-1} + A_2 \Gamma_{h-2} + ... + A_p \Gamma_{h-p}$$

known as Yule-Walker equations.

6. VAR: ESTIMATION AND HYPOTHESIS TESTING[1]

[1] This part is based on the Hamilton textbook.

1 Conditional Likelihood

Let us condsider the VAR(p)

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \epsilon_t \tag{1}$$

with $\epsilon_t \sim i.i.d N(0, \Omega)$. Suppose we have a sample of $T + p$ observations for such variables. Conditioning on the first p observations we can form the conditional likelihood

$$f(Y_T, Y_{T-1}, ..., Y_1 | Y_0, Y_{-1}, ..., Y_{-p+1}, \theta) \tag{2}$$

where θ is a vector containing all the parameters of the model. We refer to (2) as "conditional likelihood function".

The joint density of observations 1 through t conditioned on $Y_0, ..., Y_{-p+1}$ satisfies

$$\begin{aligned} f(Y_t, Y_{t-1}, ..., Y_1 | Y_0, Y_{-1}, ..., Y_{-p+1}, \theta) &= f(Y_{t-1}, ..., Y_1 | Y_0, Y_{-1}, ..., Y_{-p+1}, \theta) \\ &\quad \times f(Y_t | Y_{t-1}, ..., Y_1, Y_0, Y_{-1}, ..., Y_{-p+1}, \theta) \end{aligned}$$

Applying the formula recursively, the likelihood for the full sample is the product of the individual conditional densities

$$f(Y_t, Y_{t-1}, ..., Y_1 | Y_0, Y_{-1}, ..., Y_{-p+1}, \theta) = \prod_{t=1}^{T} f(Y_t | Y_{t-1}, Y_{t-2}, ..., Y_{-p+1}, \theta) \tag{3}$$

At each t, conditional on the values of Y through date $t-1$

$$Y_t|Y_{t-1}, Y_{t-2}, ..., Y_{-p+1} \sim N(c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p}, \Omega)$$

Recall

$$X_t = \begin{pmatrix} 1 \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{pmatrix}$$

is an $(np+1) \times 1$ vector and let $\Pi' = [c, A_1, A_2, ..., A_p]$ be an $(n \times np+1)$ matrix of coefficients. Using this notation we have that

$$Y_t|Y_{t-1}, Y_{t-2}, ..., Y_{-p+1} \sim N(\Pi' X_t, \Omega)$$

Thus the conditional density of the tth observation is

$$\begin{aligned} f(Y_t|Y_{t-1}, Y_{t-2}, ..., Y_{-p+1}, \theta) &= (2\pi)^{-n/2} \left|\Omega^{-1}\right|^{1/2} \\ &\exp\left[(-1/2)(Y_t - \Pi' X_t)' \Omega^{-1} (Y_t - \Pi' X_t)\right] \end{aligned} \tag{4}$$

The sample log-likelihood is found by substituting (4) into (3) and taking logs

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \log f(Y_t|Y_{t-1}, Y_{t-2}, ..., Y_{-p+1}, \theta)$$

$$
\begin{aligned}
= \; & -(Tn/2)\log(2\pi) + (T/2)\log|\Omega^{-1}| - \\
& (-1/2)\sum_{t=1}^{T}\left[(Y_t - \Pi' X_t)'\Omega^{-1}(Y_t - \Pi' X_t)\right] \qquad (5)
\end{aligned}
$$

2 Maximum Likelihood Estimate (MLE) of Π

The MLE estimate of Π are given by

$$\hat{\Pi}'_{MLE} = \left[\sum_{t=1}^{T} Y_t X_t'\right] \left[\sum_{t=1}^{T} X_t X_t'\right]^{-1}$$

$\hat{\Pi}'_{MLE}$ is $n \times (np+1))$. The jth row of $\hat{\Pi}'$ is

$$\hat{\pi}_j' = \left[\sum_{t=1}^{T} Y_{jt} X_t'\right] \left[\sum_{t=1}^{T} X_t X_t'\right]^{-1}$$

which is the estimated coefficient vector from an OLS regression of Y_{jt} on X_t. Thus the MLE estimates for equation j are found by an OLS regression of Y_{jt} on p lags of all the variables in the system.

We can verify that $\hat{\Pi}'_{MLE} = \hat{\Pi}'_{OLS}$. To verify this rewrite the last term in the log-likelihood as

$$\begin{aligned}
\sum_{t=1}^{T} \left[(Y_t - \Pi' X_t)'\Omega^{-1}(Y_t - \Pi' X_t)\right] &= \sum_{t=1}^{T}[(Y_t - \hat{\Pi}' X_t + \hat{\Pi}' X_t - \Pi' X_t)'\Omega^{-1} \\
&\quad \times (Y_t - \hat{\Pi}' X_t + \hat{\Pi}' X_t - \Pi' X_t)] \\
&= \sum_{t=1}^{T} \left[(\hat{\epsilon}_t + (\hat{\Pi}' - \Pi')X_t)'\Omega^{-1}(\hat{\epsilon}_t + (\hat{\Pi}' - \Pi')X_t)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=1}^{T} \hat{\epsilon}_t' \Omega^{-1} \hat{\epsilon}_t + 2 \sum_{t=1}^{T} \hat{\epsilon}_t' \Omega^{-1} (\hat{\Pi}' - \Pi')' X_t + \\
&\quad + \sum_{t=1}^{T} X_t' (\hat{\Pi}' - \Pi') \Omega^{-1} (\hat{\Pi}' - \Pi')' X_t
\end{aligned}
$$

The term $2 \sum_{t=1}^{T} \hat{\epsilon}_t' \Omega^{-1} (\hat{\Pi}' - \Pi')' X_t$ is a scalar so that

$$
\begin{aligned}
\sum_{t=1}^{T} \hat{\epsilon}_t' \Omega^{-1} (\hat{\Pi}' - \Pi')' X_t &= tr \left[\sum_{t=1}^{T} \hat{\epsilon}_t' \Omega^{-1} (\hat{\Pi}' - \Pi')' X_t \right] \\
&= tr \left[\sum_{t=1}^{T} \Omega^{-1} (\hat{\Pi}' - \Pi')' X_t \hat{\epsilon}_t' \right] \\
&= tr \left[\Omega^{-1} (\hat{\Pi}' - \Pi')' \sum_{t=1}^{T} X_t \hat{\epsilon}_t' \right]
\end{aligned}
$$

But $\sum_{t=1}^{T} X_t \hat{\epsilon}_t' = 0$ by construction since regressors are orthogonal to the residuals so that we have

$$
\sum_{t=1}^{T} \left[(Y_t - \Pi' X_t)' \Omega^{-1} (Y_t - \Pi' X_t) \right] = \sum_{t=1}^{T} \hat{\epsilon}_t' \Omega^{-1} \hat{\epsilon}_t + \sum_{t=1}^{T} X_t' (\hat{\Pi}' - \Pi') \Omega^{-1} (\hat{\Pi}' - \Pi')' X_t
$$

Given that Ω is positive definite, so is Ω^{-1}, thus the smallest values of

$$\sum_{t=1}^{T} X_t'(\hat{\Pi}' - \Pi')\Omega^{-1}(\hat{\Pi}' - \Pi')'X_t$$

is achieved by setting $\Pi = \hat{\Pi}$, i.e. the log-likelihood is maximized when $\Pi = \hat{\Pi}$. This establishes the claim that the MLE estimator coincides with the OLS estimator.

Recall the SUR representation

$$\mathbf{Y} = \mathbf{X}\mathbf{A} + \mathbf{u}$$

where $\mathbf{X} = [X_1, ..., X_T]'$, $X_t = [Y_{t-1}', Y_{t-2}'..., Y_{t-p}']'$ $\mathbf{Y} = [Y_1, ..., Y_T]'$, $\mathbf{u} = [\epsilon_1, ..., \epsilon_T]'$ and $\mathbf{A} = [A_1, ..., A_p]'$.The MLE estimator is given by

$$\hat{\mathbf{A}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

(notice that $\hat{\mathbf{A}} = \hat{\Pi}_{MLE}'$, different notation same estimator)

3 MLE of Ω

3.1 Some useful results

Let X be an $n \times 1$ vector and let A be a nonsymetric and unrestricted matrix. Consider the quadratic form $X'AX$.

(i) The first result says that

$$\frac{\partial X'AX}{\partial A} = XX'$$

(ii) The second result says that

$$\frac{\partial \log |A|}{\partial A} = (A')^{-1}$$

3.2 The estimator

We now find the MLE of Ω. When evaluated at $\hat{\Pi}$ the log likelihood is

$$\mathcal{L}(\theta) \;=\; -(Tn/2)\log(2\pi) + (T/2)\log|\Omega^{-1}| - (1/2)\sum_{t=1}^{T}\hat{\epsilon}_t'\Omega^{-1}\hat{\epsilon}_t \tag{6}$$

Taking derivatives and using results for matrix derivatives we have:

$$\begin{aligned}
\frac{\partial\mathcal{L}(\Omega,\hat{\Pi})}{\partial\Omega^{-1}} &= (T/2)\frac{\partial\log|\Omega^{-1}|}{\partial\Omega^{-1}} - (1/2)\frac{\sum_{t=1}^{T}\partial\hat{\epsilon}_t'\Omega^{-1}\hat{\epsilon}_t}{\partial\Omega^{-1}} \\
&= (T/2)\Omega' - (1/2)\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t' \qquad (7)
\end{aligned}$$

The likelihood is maximized when the derivative is set to zero, or when

$$\Omega' = (1/T)\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t' \tag{8}$$

$$\hat{\Omega} = (1/T)\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t' \tag{9}$$

4 Asymptotic distribution of $\hat{\Pi}$

Maximum likelihood estimates are consistent even if the true innovations are non-Gaussian. The asymptotic properties of the MLE estimator are summarized in the following proposition

Proposition (11.1H). *Let*

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \varepsilon_t$$

where ε_t is i.i.d. with mean zero and variance Ω and $E(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{lt}\varepsilon_{mt}) < \infty$ for all i, j, l, m and where the roots of

$$|I - A_1 z + A_2 z^2 + ... + A_p z^p| = 0$$

lie outside the unit circle. Let $k = np + 1$ and let X_t be the $1 \times k$ vector

$$X_t' = [1, Y_{t-1}', Y_{t-2}', ..., Y_{t-p}']$$

Let $\hat{\pi}_T = vec(\hat{\Pi}_T)$ denote the $nk \times 1$ vector of coefficients resulting from the OLS regressions of each of the element of Y_t on X_t for a sample of size T.

$$\hat{\pi}_T = \begin{pmatrix} \hat{\pi}_{1T} \\ \hat{\pi}_{2T} \\ \vdots \\ \hat{\pi}_{nT} \end{pmatrix}$$

where

$$\hat{\pi}_{iT} = \left[\sum_{t=1}^{T} X_t X_t'\right]^{-1} \left[\sum_{t=1}^{T} X_t Y_{it}\right]$$

and let π *denote the vector of corresponding population coefficients. Finally let*

$$\hat{\Omega} = (1/T) \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

where $\hat{\varepsilon}_t = [\hat{\varepsilon}_{1t}\ \hat{\varepsilon}_{2t}\ ...\ \hat{\varepsilon}_{nt}]$, *and* $\hat{\varepsilon}_{it} = Y_{it} - X_t' \hat{\pi}_{iT}$.

Then

(a) $(1/T) \sum_{t=1} X_t X_t' \xrightarrow{p} Q$ *where* $Q = E(X_t X_t')$

(b) $\hat{\pi}_T \xrightarrow{p} \pi$

(c) $\hat{\Omega} \xrightarrow{p} \Omega$

(d) $\sqrt{T}(\hat{\pi}_T - \pi) \xrightarrow{L} N(0, \Omega \otimes Q^{-1})$

Notice that result (d) implies that

$$\sqrt{T}(\hat{\pi}_{iT} - \pi_i) \xrightarrow{L} N(0, \sigma_i^2 Q^{-1})$$

where σ_i^2 is the variance of the error term of the ith equation. σ_i^2 is consistently estimated by $(1/T)\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2$ and that Q^{-1} is consistently estimated by $((1/T)\sum_{t=1} X_t X_t')^{-1}$. Therefore we can treat $\hat{\pi}_i$ approximately as

$$\hat{\pi}_i \approx N\left(\pi, \hat{\sigma}_i^2\left[\sum_{t=1} X_t X_t'\right]^{-1}\right)$$

5 Number of lags

As in the univariate case, care must be taken to account for all systematic dynamics in multivariate models. In VAR models, this is usually done by choosing a sufficient number of lags to ensure that the residuals in each of the equations are white noise.

AIC: Akaike information criterion Choosing the p that minimizes the following

$$AIC(p) = T \ln |\hat{\Omega}| + 2(n^2 p)$$

*BIC: Bayesian information criterion*Choosing the p that minimizes the following

$$BIC(p) = T \ln |\hat{\Omega}| + (n^2 p) \ln T$$

HQ: Hannan- Quinn information criterion Choosing the p that minimizes the following

$$HQ(p) = T \ln |\hat{\Omega}| + 2(n^2 p) \ln \ln T$$

$\hat{p}$ obtained using BIC and HQ are consistent while with AIC it is not.

AIC overestimate the true order with positive probability and underestimate the true order with zero probability.

Suppose a VAR(p) is fitted to $Y_1, ..., Y_T$ (Y_t not necessarily stationary). In small sample the following relations hold:

$\hat{p}_{BIC} \leq \hat{p}_{AIC}$ if $T \geq 8$

$\hat{p}_{BIC} \leq \hat{p}_{HQ}$ for all T

$\hat{p}_{HQ} \leq \hat{p}_{AIC}$ if $T \geq 16$

6 Testing: Wald Test

A general hypothesis of the form $R\pi = r$, involving coefficients across different equations can be tested using the Wald form of the χ^2 test seen in the first part of the course. Result (d) of Proposition 11.1H establishes that

$$\sqrt{T}(R\hat{\pi}_T - r) \xrightarrow{L} N(0, R(\Omega \otimes Q^{-1})R'), \tag{10}$$

The following proposition establishes a useful result for testing

Proposition (3.5L). *Suppose (10) holds,* $\hat{\Omega} \xrightarrow{p} \Omega$, $(1/T)\sum_{t=1} X_t X_t' \xrightarrow{p} Q$, *(*$Q, \Omega$ *both non-singular) and* $R\pi = r$ *is true.*
Then

$$(R\hat{\pi}_T - r)' \left\{ R \left[\hat{\Omega} \otimes \left(\sum_{t=1}^{T} X_t X_T' \right)^{-1} \right] R' \right\}^{-1} (R\hat{\pi}_T - r) \xrightarrow{d} \chi^2(m)$$

where m *is the number of restrictions, i.e. the number of rows of* R*.*

7 Testing: Likelihood ratio Test

First let consider the log Likelihood evaluated at the MLE

$$\mathcal{L}(\hat{\Omega},\hat{\Pi})) = -(Tn/2)\log(2\pi) + (T/2)\log|\hat{\Omega}^{-1}| - (1/2)\sum_{t=1}^{T}\hat{\epsilon}_t'\hat{\Omega}^{-1}\hat{\epsilon}_t \tag{11}$$

The last term is

$$\begin{aligned}
(1/2)\sum_{t=1}^{T}\hat{\epsilon}_t'\hat{\Omega}^{-1}\hat{\epsilon}_t &= (1/2)tr\left[\sum_{t=1}^{T}\hat{\epsilon}_t'\hat{\Omega}^{-1}\hat{\epsilon}_t\right] \\
&= (1/2)tr\left[\sum_{t=1}^{T}\hat{\Omega}^{-1}\hat{\epsilon}_t\hat{\epsilon}_t'\right] \\
&= (1/2)tr\left[\hat{\Omega}^{-1}T\hat{\Omega}\right] \\
&= (1/2)tr\left[TI_n\right] \\
&= Tn/2
\end{aligned}$$

Substituting in (18) we have

$$\mathcal{L}(\hat{\Omega},\hat{\Pi})) = -(Tn/2)\log(2\pi) + (T/2)\log|\hat{\Omega}^{-1}| - (Tn/2) \tag{12}$$

Suppose we want to test the null hypothesis that a set of variables was generated by a VAR with p_0 lags against the alternative specification with $p_1 > p_0$. Let $\hat{\Omega}_0 = (1/T)\sum_{t=1}^{T}\hat{\epsilon}_t(p_0)\hat{\epsilon}_t(p_0)'$ where

$\hat{\epsilon}(p_0)_t$ is the residual estimated in the VAR(p_0). the log likelihood is given by

$$\mathcal{L}_0 \;=\; -(Tn/2)\log(2\pi) + (T/2)\log|\hat{\Omega}_0^{-1}| - (Tn/2) \tag{13}$$

Let $\hat{\Omega}_1 = (1/T)\sum_{t=1}^{T}\hat{\epsilon}_t(p_1)\hat{\epsilon}_t(p_1)'$ where $\hat{\epsilon}(p_1)_t$ is the residual estimated in the VAR(p_1). the log likelihood is given by

$$\mathcal{L}_1 \;=\; -(Tn/2)\log(2\pi) + (T/2)\log|\hat{\Omega}_1^{-1}| - (Tn/2) \tag{14}$$

Twice the log ratio is

$$\begin{aligned} 2(\mathcal{L}_1 - \mathcal{L}_0) &= 2\{(T/2)\log|\hat{\Omega}_1^{-1}| - (T/2)\log|\hat{\Omega}_0^{-1}|\} \\ &= T\log(1/|\hat{\Omega}_1|) - T\log(1/|\hat{\Omega}_0|) \\ &= -T\log(|\hat{\Omega}_1|) + T\log(|\hat{\Omega}_0|) \\ &= T\{\log(|\hat{\Omega}_0|) - \log(|\hat{\Omega}_1|)\} \end{aligned} \tag{15}$$

Under the null hypothesis, this asymptotically has a χ^2 distribution with degrees of freedom equal to the number of restriction imposed under H_0. Each equation in the restricted model has $n(p_1 - p_0)$ restrictions, in total $n^2(p_1 - p_0)$. Thus is asymptotically χ^2 with $n^2(p_1 - p_0)$ degrees of freedom.

Example. Suppose $n = 2, p_0 = 3, p_1 = 4$, T=46. Let $\sum_{t=1}^{T} [\hat{\epsilon}(p_0)_{1t}]^2 = 2$, $\sum_{t=1}^{T} [\hat{\epsilon}(p_0)_{2t}]^2 = 2.5$ and $\sum_{t=1}^{T} \hat{\epsilon}(p_0)_{1t}\epsilon(p_0)_{2t} = 1$. Then

$$\hat{\Omega}_0 = \begin{pmatrix} 2.0 & 1.0 \\ 1.0 & 2.5 \end{pmatrix}$$

for which $\log|\hat{\Omega}_0| = \log 4 = 1.386$. Moreover

$$\hat{\Omega}_1 = \begin{pmatrix} 1.8 & 0.9 \\ 0.9 & 2.2 \end{pmatrix}$$

for which $\log|\hat{\Omega}_1| = 1.147$. Then

$$2(\mathcal{L}_1 - \mathcal{L}_0) = 46(1.386 - 1.147) = 10.99$$

The degrees of freedom for this test are $2^2(4-3) = 4$. Since $10.99 > 9.49$ (the 5% critical value for a χ_4^2 variable), the null hypothesis is rejected.

8 Granger Causality

Granger causality If a scalar y cannot help in forecasting x we say that y does not Granger cause x. *y fails to Granger cause x* if for all $s > 0$ the mean squared error of a forecast of x_{t+s} based on $(x_t, x_{t-1}, ...,)$ is the same as the MSE of a forecast of x_{t+s} based on $(x_t, x_{t-1}, ...,)$ and $(y_t, y_{t-1}, ...,)$. If we restrict ourselves to linear functions, y fails to Granger-cause x if

$$MSE\left[(\hat{E}(x_{t+s}|x_t, x_{t-1}, ...,)\right] = MSE\left[(\hat{E}(x_{t+s}|x_t, x_{t-1}, ..., y_t, y_{t-1}, ...,)\right] \tag{16}$$

where $\hat{E}(x|y)$ is the linear projection of vector x on the vector y, i.e. the linear function $\alpha' y$ satisfying $E[(x - \alpha' y)y] = 0$.

Also we say that x *is exogenous in the time series sense with respect to y* if (23) holds.

8.1 Granger Causality in Bivariate VAR

Let us consider a bivariate VAR

$$\begin{aligned}\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} &= \begin{pmatrix} A_{11}^{(1)} & A_{12}^{(1)} \\ A_{21}^{(1)} & A_{22}^{(1)} \end{pmatrix} \begin{pmatrix} Y_{1t-1} \\ Y_{2t-1} \end{pmatrix} + \begin{pmatrix} A_{11}^{(2)} & A_{12}^{(2)} \\ A_{21}^{(2)} & A_{22}^{(2)} \end{pmatrix} \begin{pmatrix} Y_{1t-2} \\ Y_{2t-2} \end{pmatrix} + \\ &+ ... + \begin{pmatrix} A_{11}^{(p)} & A_{12}^{(p)} \\ A_{21}^{(p)} & A_{22}^{(p)} \end{pmatrix} \begin{pmatrix} Y_{1t-p} \\ Y_{2t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \qquad (17)\end{aligned}$$

We say that Y_2 *fails to Granger cause* Y_1 if the elements $A_{12}^{(j)} = 0$ for $j = 1, ..., p$. We can check that if $A_{12}^{(j)} = 0$ the two MSE coincide. For $s = 1$ we have

$$\begin{aligned}\hat{E}(Y_{1t+1}|Y_{1t}, Y_{1t-1}, ..., Y_{2t}, Y_{2t-1}) &= A_{11}^{(1)} Y_{1t-1} + A_{12}^{(1)} Y_{2t-1} + A_{11}^{(2)} Y_{1t-2} + A_{12}^{(2)} Y_{2t-2} + \\ &+ ... + A_{11}^{(p)} Y_{1t-p} + A_{12}^{(p)} Y_{2t-p}\end{aligned}$$

clearly if $A_{12}^{(j)} = 0$

$$\begin{aligned}\hat{E}(Y_{1t+1}|Y_{1t}, Y_{1t-1}, ..., Y_{2t}, Y_{2t-1}) &= A_{11}^{(1)} Y_{1t-1} + A_{11}^{(2)} Y_{1t-2} + ... + A_{11}^{(p)} Y_{1t-p} \\ &= \hat{E}(Y_{1t+1}|Y_{1t}, Y_{1t-1}, ...) \qquad (18)\end{aligned}$$

An important implication of Granger causality in the bivariate context is that if Y_2 *fails to Granger cause* Y_1 then the Wold representation of Y_t is

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} C_{11}(L) & 0 \\ C_{21}(L) & C_{22}(L) \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \qquad (19)$$

that is the second Wold shock has no effects on the first variable. This it is easy to show by deriving the Wold representation by inverting the VAR polynomial matrix.

8.2 Econometric test for Granger Causality

The simplest approach to test Granger causality in an autoregressive framework is the following is to estimate the bivariate VAR with p lags by OLS and test the null hypothesis $H_0: \; A_{12}^{(1)} = A_{12}^{(2)} = \ldots A_{12}^{(p)} = 0$ using an F-test using

$$S_1 = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(T-2p-1)}$$

and reject if $S_1 > F_{(\alpha,p,T-2p-1)}$. An asymptotically equivalent test is

$$S_2 = \frac{T(RSS_0 - RSS_1)}{RSS_1}$$

and reject if $S_2 > \chi_{(\alpha,p)}$.

8.3 Application 1: Output Growth and the Yield Curve

• Many research papers have found that yield curve (difference in long and short yield) has been a good predictor, i.e. a variable that helps to forecast, for the real GDP growth in the US (Estrella, 2000,2005). However recent evidence suggests that its predictive power has reduced since the beginning of the 80s (see D'Agostino, Giannone and Surico, 2006). This means we should find that the yield curve Granger cuse output growth before mid 80's but not after.

• We estimate a bivariate VAR for the growth rates of the real GDP and the difference between the 10-year rate and the federal funds rate. Data are from FREDII StLouis Fed spanning from 1954:III-2007:III. The AIC criterion suggests $p = 6$.

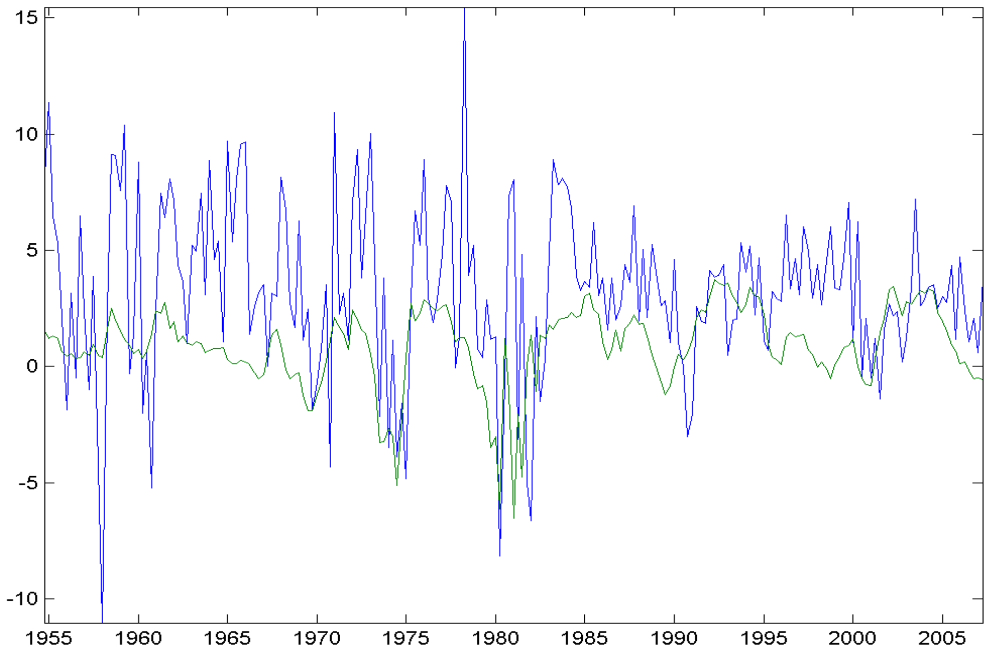


Figure 1: Blu: real gdp growth rates; green: spread long-short.

Table 1: F-Tests of Granger Causality

	1954:IV-2007:III	1954:IV-1990:I	1990:I-2007:III
S_1	5.4233	6.0047	0.9687
10%	1.8050	1.8222	1.8954
5%	2.1460	2.1725	2.2864
1%	2.8971	2.9508	3.1864

We cannot reject the hypothesis that the spread does not Granger cause real output growth in the last period, while we reject the hypothesis for all the other sample. This can be explained by a change in the information content of private agents expectations, which is the information embedded in the yield curve.

8.4 Application 2: Money, Income and Causality

In the 50's and 60's a big debate about the importance of money and monetary policy. Does money affect output? For Friedman and monetarist yes. For Keynesian (Tobin) no: output was affecting money, movement in money stock were reflecting movements in the money demand. Sims in 1972 run a test in order to distinguish between the two visions. He found that money was Granger-causing output but not the reverse, providing evidence in favor of the monetarist view.

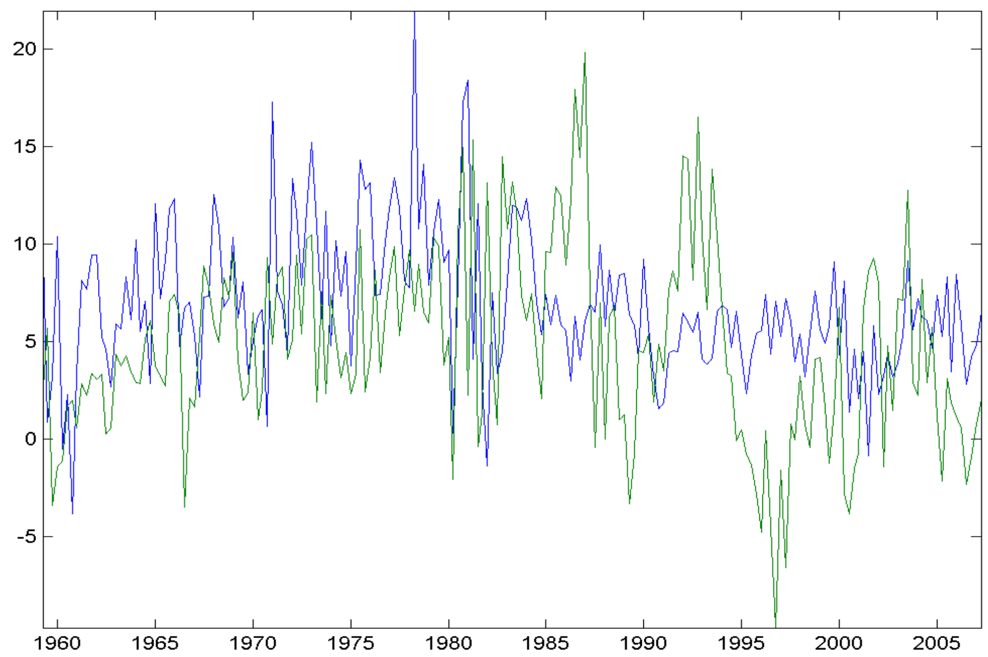


Figure 2: Blu: real gnp growth rates; green: M1 growth rates.

Table 2: F-Tests of Granger Causality

	1959:II-1972:I	1959:II-2007:III
$M \to Y$	4.4440	2.2699
$Y \to M$	0.5695	3.5776
10%	2.0948	1.7071
5%	2.6123	1.9939
1%	3.8425	2.6187

In the first sample money Granger cause (at 5%) output but not the converse (Sims(72)'s result). In the second sample at the 5% both output Granger cause money and money Granger cause output.

8.5 Caveat: Granger Causality Tests and Forward Looking Behavior

Let us consider the following simple model of stock price determination where P_t is the price of one share of a stock, D_{t+1} are dividends payed at $t+1$ and r is the rate of return of the stock

$$(1+r)P_t = E_t(D_{t+1} + P_{t+1})$$

According to the theory stock price incorporates the market's best forecast of the present value of the future dividends. Solving forward we have

$$P_t = E_t \sum_{j=1}^{\infty} \left[\frac{1}{1+r}\right]^j D_{t+j}$$

Suppose

$$D_t = d + u_t + \delta u_{t-1} + v_t$$

where u_t, v_t are Gaussian WN and d is the mean dividend. The forecast of D_{t+j} based on this information is

$$E_t(D_{t+j}) = \begin{cases} d + \delta u_t & \text{for} \quad j = 1 \\ d & \text{for} \quad j = 2, 3, ... \end{cases}$$

Substituting in the stock price equation we have

$$P_t = d/r + \delta u_t/(1+r)$$

Thus the price is white noise and could not be forecast on the basis of lagged stock prices or dividends. No series should Granger cause prices. The value of u_{t-1} can be uncovered from the lagged stock price

$$\delta u_{t-1} = (1+r)P_{t-1} - (1+r)d/r$$

The bivariate VAR takes the form

$$\begin{pmatrix} P_t \\ D_t \end{pmatrix} = \begin{pmatrix} d/r \\ -d/r \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ (1+r) & 0 \end{pmatrix} \begin{pmatrix} P_{t-1} \\ D_{t-1} \end{pmatrix} + \begin{pmatrix} \delta u_t/(1+r) \\ u_t + v_t \end{pmatrix} \tag{20}$$

Granger causation runs in the opposite direction from the true causation. Dividends fail to Granger cause prices even though expected dividends are the only determinant of prices. On the other hand prices Granger cause dividends even though this is not the case in the true model.

8.6 Granger Causality in a Multivariate Context

Suppose now we are interested in testing for Granger causality in a multivariate ($n > 2$) context. Let us consider the following representation of a VAR(p)

$$\begin{aligned}\tilde{Y}_{1t} &= \tilde{A}_1\tilde{X}_{1t} + \tilde{A}_2\tilde{X}_{2t} + \tilde{\epsilon}_{1t} \\ \tilde{Y}_{2t} &= \tilde{B}_1\tilde{X}_{1t} + \tilde{B}_2\tilde{X}_{2t} + \tilde{\epsilon}_{2t}\end{aligned} \tag{21}$$

where $\tilde{Y}_{1t}$ and $\tilde{Y}_{2t}$ are two vectors containing respectively n_1 and n_2 variables of Y_t. Let

$$\tilde{X}_{1t} = \begin{pmatrix} \tilde{Y}_{1t-1} \\ \tilde{Y}_{1t-2} \\ \vdots \\ \tilde{Y}_{1t-p} \end{pmatrix} \quad \tilde{X}_{2t} = \begin{pmatrix} \tilde{Y}_{2t-1} \\ \tilde{Y}_{2t-2} \\ \vdots \\ \tilde{Y}_{2t-p} \end{pmatrix}$$

$\tilde{Y}_{1t}$ is said *block exogenous in the time series sense* with respect to $\tilde{Y}_{2t}$ if the elements in $\tilde{Y}_{2t}$ are of no help in improving the forecast of any variable in $\tilde{Y}_{1t}$. $\tilde{Y}_{1t}$ is block exogenous if $\tilde{A}_2 = 0$.

In order to test block exogeneity we can proceed as follows. First notice that the log likelihood can be rewritten in terms of a conditional and a marginal log density

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \ell_{1t} + \sum_{t=1}^{T} \ell_{2t}$$

$$\begin{aligned}
\ell_{1t} &= \log f(\tilde{Y}_{1t}|\tilde{X}_{1t}, \tilde{X}_{2t}, \theta) \\
&= -(n/2)\log(2\pi) - (1/2)\log|\Omega_{11}| - \\
&\quad -(1/2)[(\tilde{Y}_{1t} - \tilde{A}_1\tilde{X}_{1t} - \tilde{A}_2\tilde{X}_{2t})'\Omega_{11}^{-1}(\tilde{Y}_{1t} - \tilde{A}_1\tilde{X}_{1t} - \tilde{A}_2\tilde{X}_{2t})] \\
\ell_{2t} &= \log f(\tilde{Y}_{2t}|\tilde{Y}_{1t}, \tilde{X}_{1t}, \tilde{X}_{2t}, \theta) \\
&= -(n/2)\log(2\pi) - (1/2)\log|H| - \\
&\quad -(1/2)[(\tilde{Y}_{2t} - \tilde{D}_0\tilde{Y}_{1t} - \tilde{D}_1\tilde{X}_{1t} - \tilde{D}_2\tilde{X}_{2t})'H^{-1}(\tilde{Y}_{2t} - \tilde{D}_0\tilde{Y}_{1t} - \tilde{D}_1\tilde{X}_{1t} - \tilde{D}_2\tilde{X}_{2t})]
\end{aligned}$$

where $\tilde{D}_0\tilde{Y}_{1t} + \tilde{D}_1\tilde{X}_{1t} + \tilde{D}_2\tilde{X}_{2t}$ and H represent the mean and the variance respectively of $\tilde{Y}_{2t}$ conditioning also on $\tilde{Y}_{1t}$.

Consider the the maximum likelihood estimation of the system subject to the constraint $\tilde{A}_2 = 0$ giving estimates $\hat{\tilde{A}}_1(0)$, $\hat{\Omega}_{11}(0), \hat{D}_0, \hat{D}_1, \hat{D}_2, \hat{H}$. Now consider the unrestricted maximum likelihood estimation of the system providing the estimates $\hat{\tilde{A}}_1, \hat{\Omega}_{11}, \hat{D}_0, \hat{D}_1, \hat{D}_2\hat{H}$. The likelihood functions evaluated at the MLE in the two cases are

$$\begin{aligned}
\mathcal{L}(\hat{\theta}(0)) &= -(T/2)\log(2\pi) + (T/2)\log|\hat{\Omega}_{11}^{-1}(0)| - (Tn/2) \\
\mathcal{L}(\hat{\theta}) &= -(T/2)\log(2\pi) + (T/2)\log|\hat{\Omega}_{11}^{-1}| - (Tn/2)
\end{aligned}$$

A likelihood ratio test of the hypothesis $\tilde{A}_1 = 0$ can be based on

$$2(\mathcal{L}(\hat{\theta}) - \mathcal{L}(\hat{\theta}(0)) = T(\log|\hat{\Omega}_{11}^{-1}| - \log|\hat{\Omega}_{11}^{-1}(0)|) \tag{22}$$

In practice we perform OLS regression of each of the elements in $\tilde{Y}_{1t}$ on p lags of all the elements in $\tilde{Y}_{1t}$ and all the elements in $\tilde{Y}_{2t}$. Let $\hat{\epsilon}_{1t}$ be the vector of sample residual and $\hat{\Omega}_{11}$ their estimated variance covariance matrix. Next perform OLS regressions of each element of $\tilde{Y}_{1t}$ on p lags of $\tilde{Y}_{1t}$. Let $\hat{\epsilon}_{1t}(0)$ be the vector of sample residual and $\hat{\Omega}_{11}(0)$ their estimated variance covariance matrix. If

$$T\{\log|\hat{\Omega}_{11}(0)| - \log|\hat{\Omega}_{11}|\}$$

is greater than the 5% critical values for a $\chi^2_{n_1 n_2 p}$, then the null hypothesis is rejected and the conclusion is that some elements in $\tilde{Y}_{2t}$ is important for forecasting $\tilde{Y}_{1t}$

7. STRUCTURAL VAR: THEORY

1 Structural Vector Autoregressions

Impulse response functions are interpreted under the assumption that *all the other shocks are held constant*. However in the Wold representation the shocks are not orthogonal. So the assumption is not very realistic!.

This is why we need Structural VAR in order to perform policy analysis. Ideally we would like to have

1) orthogonal shock

2) shocks with economic meaning (technology, demand, labor supply, monetary policy etc.)

1.1 Statistical Orthogonalizations

There are two easy way to orthogonalize shocks.

1) Cholesky decomposition

2) Spectral Decomposition

1.2 Cholesky decomposition

Let us consider the matrix Ω. The Cholesky factor, S, of Ω is defined as the unique lower triangular matrix such that $SS' = \Omega$. This implies that we can rewrite the VAR in terms of orthogonal shocks $\eta_t = S^{-1}\epsilon_t$ with identity covariance matrix

$$A(L)Y_t = S\eta_t$$

Impulse response to orthogonalized shocks are found from the MA representation

$$\begin{aligned} Y_t &= C(L)S\eta_t \\ &= \sum_{j=0}^{\infty} C_j S \eta_{t-j} \end{aligned} \tag{1}$$

where C_jS has the interpretation

$$\frac{\partial Y_{t+j}}{\partial \eta_t} = C_j S \tag{2}$$

That is, the row i, column k element of C_jS identifies the consequences of a unit increase in η_{kt} for the value of the ith variable at time $t + j$ holding all other shocks constant.

1.3 Spectral Decomposition

Let V and be a matrix containing the eigenvectors of Ω and Λ a diagonal matrix with the eigenvalues of Ω on the main diagonal. Then we have that $V\Lambda V' = \Omega$. This implies that we can rewrite the VAR in terms of orthogonal shocks $\xi_t = (VD^{1/2})^{-1}\epsilon_t$ with identity covariance matrix

$$A(L)Y_t = VD^{1/2}\xi$$

Impulse response to orthogonalized shocks are found from the MA representation

$$\begin{aligned} Y_t &= C(L)VD^{1/2}\xi_t \\ &= \sum_{j=0}^{\infty} C_j S\eta_{t-j} \end{aligned} \tag{3}$$

where $C_j VD^{1/2}$ has the interpretation

$$\frac{\partial Y_{t+j}}{\partial \xi_t} = C_j VD^{1/2} \tag{4}$$

That is, the row i, column k element of $C_j VD^{1/2}$ identifies the consequences of a unit increase in ξ_{kt} at date t for the value of the ith variable at time $t+j$ holding all other shocks constant.

Problem: what is the economic interpretation of the orthogonal shocks? What is the economic information contained in the impulse response functions to orthogonal shocks?

Except for special cases not clear.

1.4 The Class of Orthonormal Representations

From the class of invertible MA representation of Y_t we can derive the class of orthonormal representation, i.e. the class of representations of Y_t in term of orthonormal shocks. Let H any orthogonal matrix, i.e. $HH' = H'H = I$. Defining $w_t = (SH)^{-1}\epsilon_t$ we can recover the general class of the orthonormal representation of Y_t

$$\begin{aligned} Y_t &= C(L)SHw_t \\ &= F(L)w_t \end{aligned}$$

where $F(L) = C(L)SH$ and $w_t \sim WN$ with

$$\begin{aligned} E(w_t w_t') &= E((SH')^{-1}\epsilon_t\epsilon_t'(SH')^{-1\prime}) \\ &= HS^{-1}E(\epsilon_t\epsilon_t')H'(S')^{-1} \\ &= HS^{-1}\Omega H'(S')^{-1} \\ &= HS^{-1}SS'(S')^{-1}H' \\ &= I \end{aligned}$$

Problem: H can be any, so how should we choose one?

2 The Identification Problem

• Identifying the VAR means fixing a particular matrix H, i.e. choosing one particular representation of Y_t in order to recover the structural shocks from the VAR innovations

• Therefore structural economic shocks are linear combinations of the VAR innovations.

• In order to choose a matrix H we have to fix $n(n-1)/2$ parameters since there is a total of n^2 parameters and a total of $n(n+1)/2$ restrictions implied by orthonormality.

• Use economic theory in order to derive some restrictions on the effects of some shock on a particular variables to fix the remaining $n(n-1)/2$.

2.1 Zero restrictions: contemporaneous restrictions

• An identification scheme based on zero contemporaneous restrictions is a scheme which imposes restrictions to zero on the matrix F_0, the matrix of the impact effects.

Example. Let us consider a bivariate VAR. We have a total of $n^2 = 4$ parameters to fix. $n(n+1)/2 = 3$ are pinned down by the ortnonormality restrictions so that there are $n(n-1)/2 = 1$ free parameters. Suppose that the theory tells us that shock 2 has no effect on impact (contemporaneously) on Y_1 equal to 0, that is $F_0^{12} = 0$. This is the additional restriction that allows us to identify the shocks. In particular we will have the following restrictions:

$$HH' = I$$

$$S_{11}H_{12} + S_{12}H_{22} = 0$$

Since $S_{12} = 0$ the solution is $H_{11} = H_{22} = 1$ and $H_{12} = H_{21} = 0$.

• A common identification scheme is the Cholesky scheme (like in this case). This implies setting $H = I$. Such an identification scheme creates a recursive contemporaneous ordering ordering among variables since S^{-1} is triangular.

• This means that any variable in the vector Y_t does not depend contemporanously on the variables ordered after.

- Results depend on the particular ordering of the variables.

2.2 Zero restrictions: long run restrictions

• An identification scheme based on zero long run restrictions is a scheme which imposes restrictions on the matrix $F(1) = F_0 + F_1 + F_2 + ...$, the matrix of the long run coefficients.

Example. Again let us consider a bivariate VAR. We have a total of $n^2 = 4$ parameters to fix. $n(n+1)/2 = 3$ are pinned down by the ortnonormality restrictions so that there are $n(n-1)/2 = 1$ free parameters. Suppose that the theory tells us that shock 2 does not affect Y_1 in the long run, i.e. $F_{12}(1) = 0$. This is the additional restriction that allows us to identify the shocks. In particular we will have the following restrictions:

$$HH' = I$$
$$D_{11}(1)H_{12} + D_{12}(1)H_{22} = 0$$

where $D(1) = C(1)S$ represents the long run effects of the Cholesky shocks.

2.3 Signs restrictions

• The previous two examples yield just identification in the sense that the shockis uniquely identified, there exists a unique matrix H yielding the structural shocks.

• Sign identification is based on qualitative restriction involving the sign of some shocks on some variables. In this case we will have sets of consistent impulse response functions.

Example. Again let us consider a bivariate VAR. We have a total of $n^2 = 4$ parameters to fix. $n(n+1)/2 = 3$ are pinned down by the ortnonormality restrictions so that there are $n(n-1)/2 = 1$ free parameters. Suppose that the theory tells us that shock 2, which is the interesting one, produce a positive effect on Y_1 for k periods after the shock $F_j^{12} > 0$ for $j = 1, ..., k$. We will have the following restrictions:

$$
\begin{aligned}
HH' &= I \\
S_{11}H_{12} + S_{12}H_{22} &> 0 \\
D_{j,12}H_{12} + D_{j,22}H_{22} &> 0 \quad j = 1, ..., k
\end{aligned}
$$

where $D_j = C_j S$ represents the effects at horizon j.

• In a classical statistics approach this delivers not exact identification since there can be many H consistent with such a restriction. That is for each parameter of the impulse response functions we will have an admissible set of values.

• Increasing the number of restrictions can be helpful in reducing the number of H consistent with such restrictions.

2.4 Parametrizing H

• A useful way to parametrize the matrix H in order to include orthonormality restrictions is using rotation matrices. Let us consider the bivariate case. A rotation matrix in this case is the unity matrix

$$H = \begin{pmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{pmatrix}$$

• Note that such a matrix incorporates the orthonormality conditions. The parameter θ will be found by imposing the additional economic restriction.

• In general the rotation matrix will be found as the product of $n(n-1)/2$ rotation matrices. For the case of three shocks the rotation matrix can be found as the product of the following three matrices

$$\begin{pmatrix} cos(\theta_1) & sin(\theta_1) & 0 \\ -sin(\theta_1) & cos(\theta_1) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} cos(\theta_2) & 0 & sin(\theta_2) \\ 0 & 1 & 0 \\ -sin(\theta_2) & 0 & cos(\theta_2) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos(\theta_3) & sin(\theta_3) \\ 0 & -sin(\theta_3) & cos(\theta_3) \end{pmatrix}$$

Example. Suppose that $n = 2$ and the restriction we want to impose is that the effect of the first shock on the second variable has a positive sign, i.e.

$$S_{21}H_{11} + S_{22}H_{21} > 0$$

Using the parametrization seen before the restriction becomes

$$S_{21}\cos(\theta) - S_{22}\sin(\theta) > 0$$

Which implies

$$\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)} < \frac{S_{21}}{S_{22}}$$

If $S_{21} = 0.5$ and $S_{22} = 1$ then all the impulse response fcuntions obtained with $\theta < \text{atan}(0.5)$ satisfy the rstriction and should be kept.

2.5 Partial Identification

• In many cases we might be interested in identifying just a single shock and not all the n shocks.

• Since the shock are orthogonal we can also partially identify the model, i.e. fix just one (or a subset of) column of H. In this case what we have to do is to fix $n-1$ elements of H, all but one elements of a column of the identifying matrix. The additional restriction is provided by the norm of the vector equal one.

Example Suppose $n = 3$. We want to identify a single shock using the restriction that such shock has no effects on the first variable on impact a positive effect on the second variable and negative on the third variable.

First of all we notice that the first column of the product of orthogonal matrices seen before is

$$H_1 = \begin{pmatrix} \cos(\theta_1)\cos(\theta_2) \\ -\sin(\theta_1)\cos(\theta_2) \\ -\sin(\theta_2) \end{pmatrix}$$

therefore we have that the impact effects ofthe first shock are given by

$$\begin{pmatrix} S_{11} & 0 & 0 \\ S_{21} & S_{22} & 0 \\ S_{31} & S_{32} & S_{33} \end{pmatrix} \begin{pmatrix} \cos(\theta_1)\cos(\theta_2) \\ -\sin(\theta_1)\cos(\theta_2) \\ -\sin(\theta_2) \end{pmatrix}$$

To implement the first restriction we can set $\theta_1 = \pi/2$, i.e. $\cos(\theta_1) = 0$. This imples that

$$\begin{pmatrix} S_{11} & 0 & 0 \\ S_{21} & S_{22} & 0 \\ S_{31} & S_{32} & S_{33} \end{pmatrix} \begin{pmatrix} 0 \\ -\cos(\theta_2) \\ -\sin(\theta_2) \end{pmatrix}$$

The second restriction implies that

$$-S_{22}\cos(\theta_2) > 0$$

and the third

$$-S_{32}\cos(\theta_2) - S_{33}\sin(\theta_2) < 0$$

All the values of θ_2 satisfying the two restrictions yield impulse response functions consistent with the identification scheme.

2.6 Variance Decomposition

• The second type of analysis which is usually done in SVAR is the variance decomposition analysis.

• The idea is to decompose the total variance of a time series into the percentages attributable to each structural shock.

• Variance decomposition analysis is useful in order to address questions like "What are the sources of the business cycle?" or "Is the shock important for economic fluctuations?".

Let us consider the MA representation of an identified SVAR

$$Y_t = F(L)w_t$$

The variance of Y_{it} is given by

$$\begin{aligned}\mathrm{var}(Y_{it}) &= \sum_{k=1}^{n}\sum_{j=0}^{\infty} F_{ik}^{j2}\mathrm{var}(w_{kt}) \\ &= \sum_{k=1}^{n}\sum_{j=0}^{\infty} F_{ik}^{j2}\end{aligned}$$

where $\sum_{j=0}^{\infty} F_{ik}^{j2}$ is the variance of Y_{it} generated by the kth shock. This implies that

$$\frac{\sum_{j=0}^{\infty} F_{ik}^{j2}}{\sum_{k=1}^{n}\sum_{j=0}^{\infty} F_{ik}^{j2}}$$

is the percentage of variance of Y_{it} explained by the kth shock.

It is also possible to study the of the series explained by the shock at different horizons, i.e. short vs. long run. Consider the forecast error in terms of structural shocks. The horizon h forecast error is given by

$$Y_{t+h} - Y_{t+h|t} = F_0 w_{t+1} + F_2 w_{t+2} + ... + F_k w_{t+h}$$

the variance of the forecast error of the ith variable is thus

$$\begin{aligned} \text{var}(Y_{it+h} - Y_{it+h|t}) &= \sum_{k=1}^{n} \sum_{j=0}^{h} F_{ik}^{j2} \text{var}(w_{kt}) \\ &= \sum_{k=1}^{n} \sum_{j=1}^{h} F_{ik}^{j2} \end{aligned}$$

Thus the percentage of variance of Y_{it} explained by the kth shock is

$$\frac{\sum_{j=0}^{h} F_{ik}^{j2}}{\sum_{k=1}^{n} \sum_{j=1}^{h} F_{ik}^{j2}}$$

8. STRUCTURAL VAR: APPLICATIONS

1 Monetary Policy Shocks (Christiano Eichenbaum and Evans, 1999 HoM)

• Monetary policy shocks is the unexpected part of the equation for the monetary policy instrument (S_t).

$$S_t = f(\mathcal{I}_t) + w_t^{mp}$$

$f(\mathcal{I}_t)$ represents the systematic response of the monetary policy to economic conditions, $\mathcal{I}_t$ is the information set at time t and w_t^{mp} is the monetary policy shock.

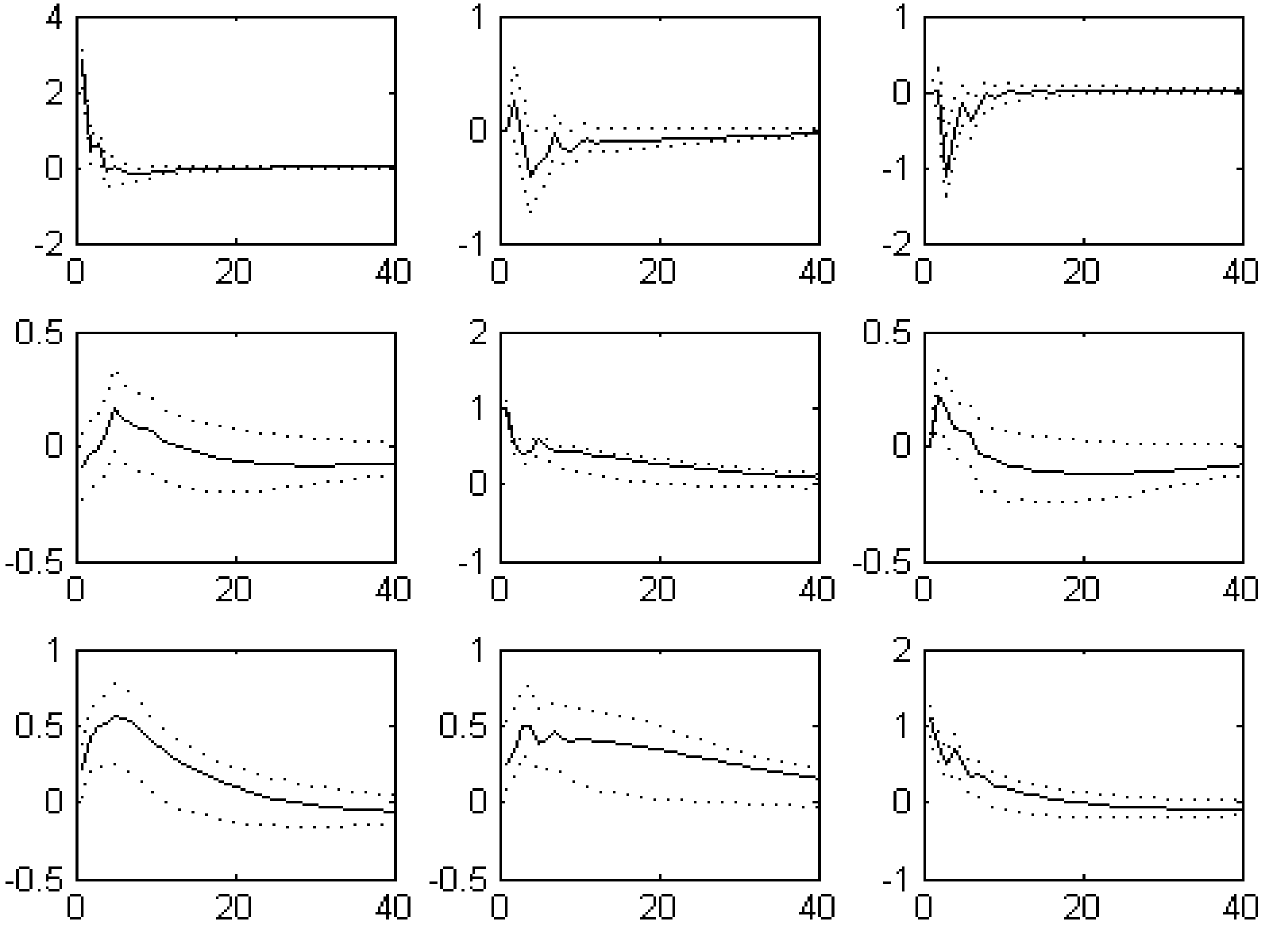
• The "standard" way to identify monetary policy shock is through zero contemporaneous restrictions. Using the standard monetary VAR (a simplified version of the CEE 98 VAR) including output growth, inflation and the federal funds rate we identify the monetary policy shock using the following restrictions:

1) Monetary policy shocks do not affect output within the same quarter

2) Monetary policy shocks do not affect inflation within the same quarter

• These two restrictions are not sufficient to identify all the shocks but are sufficient to identify the monetary policy shock.

- A simple way to implement the restrictions is to take simply the Cholesky decomposition of the variance covariance matrix in a system in which the federal funds rate is ordered last. The last column of the impulse response functions is the column of the monetary policy shock.



Cholesky impulse response functions of a system with GDP inflation and the federal funds rate. Monetary shock is in the third column.

• Notice that after a monetary tightening inflation goes up which is completely counterintuitive according to the standard transmission mechanism. This phenomenon if known as the *price puzzle*. Why is this the case?.

• *"Sims (1992) conjectured that prices appeared to rise after certain measures of a contractionary policy shock because those measures were based on specifications of* $\mathcal{I}_t$ *that did not include information about future inflation that was available to the Fed. Put differently, the conjecture is that policy shocks which are associated with substantial price puzzles are actually confounded with non-policy disturbances that signal future increases in prices."* (CEE 98)

• Sims shows that including commodity prices (signaling future inflation increases) may solve the puzzle.

2 Uhlig (JME 2006) monetary policy shocks

• Uhlig proposes a very different method to identify monetary policy shocks. Instead of using zero restrictions as in CEE he uses sign restrictions.

• He identifies the effects of a monetary policy shocks using restrictions which are implied by several economic models.

• In particular a contractionary monetary policy shock:

1. does not increase prices for k periods after the shock
2. does not increase money or monetary aggregates (i.e. reserves) for k periods after the shock
3. does not reduce short term interest rate for k periods after the shock.

• Since just one shock is identified only a column of H has to be identified, say column one.

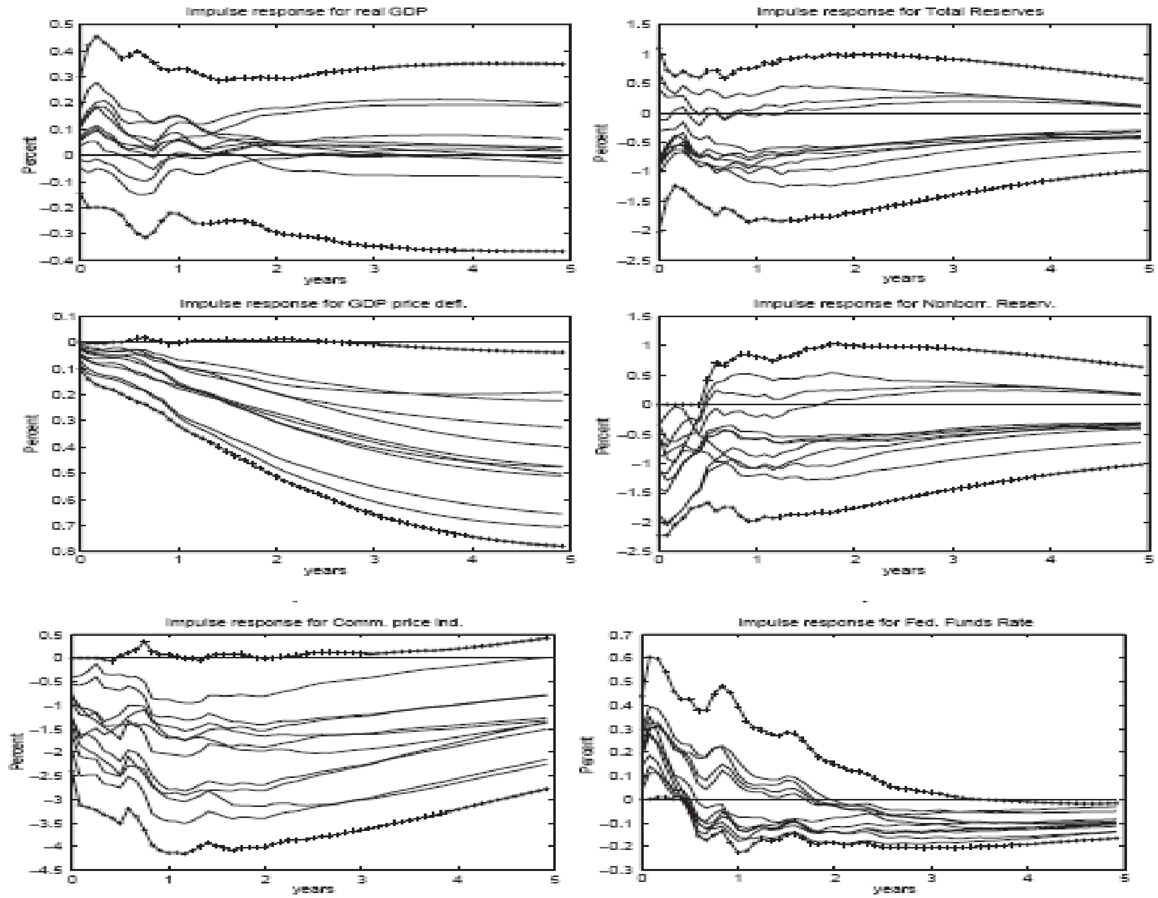
• If we order the variables in vector Y_t as follows: GDP inflation, money growth and the interest rate the restrictions imply $F_k^{i1} < 0$ for $i = 2, 3$ and $F_k^{41} > 0$.

- In order to draw impulse response functions he applies the following algorithm:

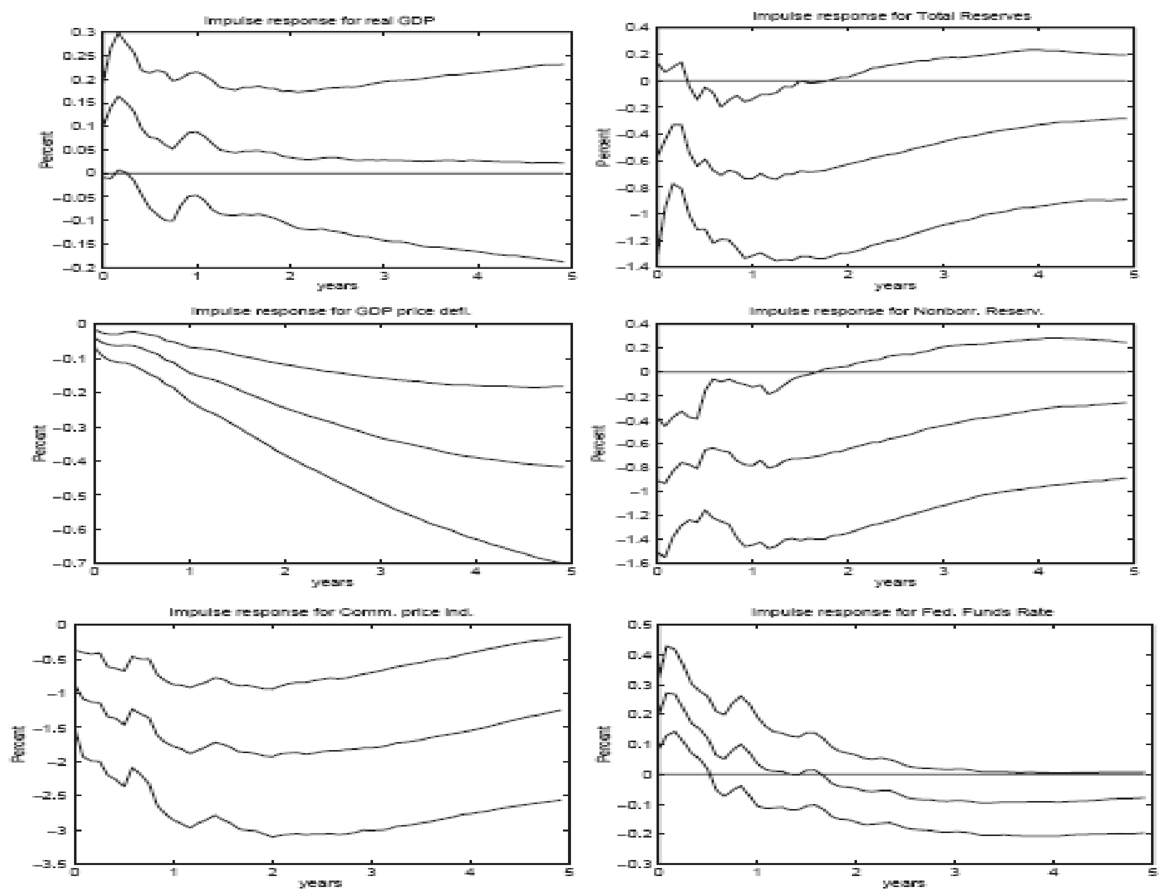
  1. He assumes that the column of H, H_1, represents the coordinate of a point uniformly distributed over the unit hypersphere (in case of bivariate VAR it represents a point in a circle). To draw such point he draws from a $N(0, I)$ and divide by the norm of the vector.

  2. Compute the impulse response functions $C_j S H_1$ for j=1,..,k.

  3. If the draw satisfies the restrictions keep it and go to 1), otherwise discard it and go to 1). Repeat 1)-3) a big number of timef L.



Source: What are the effects of a monetary policy shock... JME H. Uhlig (2006)



Source: What are the effects of a monetary policy shock... JME H. Uhlig (2006)

3 Blanchard Quah (AER 1989) aggregate demand and supply shocks

• Blanchard and Quah proposed an identification scheme based on long run restrictions.

• In their model there are two shocks: an aggregate demand and an aggregate supply disturbance.

• The restriction used to identify is that aggregate demand shocks have no effects on the long run levels of output, i.e. demand shocks are transitory on output. The idea behind of such a restriction is the existence of a vertical aggregate supply curve.

• Let us consider the following bivariate VAR

$$\begin{pmatrix} \Delta Y_t \\ U_t \end{pmatrix} = \begin{pmatrix} F_{11}(L) & F_{12}(L) \\ F_{21}(L) & F_{22}(L) \end{pmatrix} \begin{pmatrix} w_t^s \\ w_t^d \end{pmatrix}$$

where Y_t is output, U_t is the unemployment rate and w_t^s, w_t^d are two aggregate supply and demand disturbances respectively.

• The identification restriction is given by $F_{12}(1) = 0$.

- The restriction can be implemented in the following way. Let us consider the reduced form VAR

$$\begin{pmatrix} \Delta Y_t \\ U_t \end{pmatrix} = \begin{pmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

where $E(\epsilon_t \epsilon_t') = \Omega$.

Let $S = chol(A(1)\Omega A(1)')$ and $K = A(1)^{-1}S$. The identified shocks are

$$w_t = K^{-1}\epsilon_t$$

and the resulting impulse response to structural shocks are

$$F(L) = A(L)K$$

notice that the restrictions are satisfied

$$\begin{aligned} F(1) &= A(1)K \\ &= A(1)A(1)^{-1}S \\ &= S \end{aligned}$$

which is lower triangular implying that $F_{12}(1) = 0$.

Moreover we have that shocks are orthogonal since

$$\begin{aligned} KK' &= A(1)^{-1}SS'A(1)^{-1\prime} &(1)\\ &= A(1)^{-1}A(1)\Omega A(1)'A(1)^{-1\prime} \\ &= \Omega \\ & &(2)\end{aligned}$$

And

$$\begin{aligned} E(w_t w_t') &= E(K^{-1}\epsilon_t\epsilon_t' K^{-1\prime}) \\ &= K^{-1}\Omega K^{-1\prime} \\ &= K^{-1}KK'K^{-1\prime} \end{aligned}$$

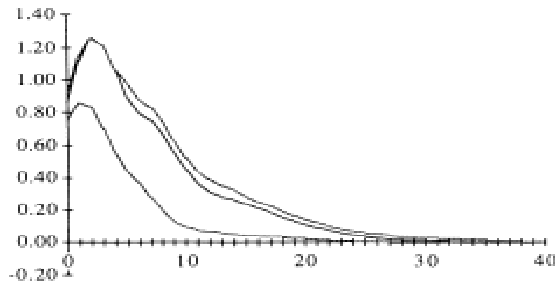


FIGURE 3. OUTPUT RESPONSE TO DEMAND

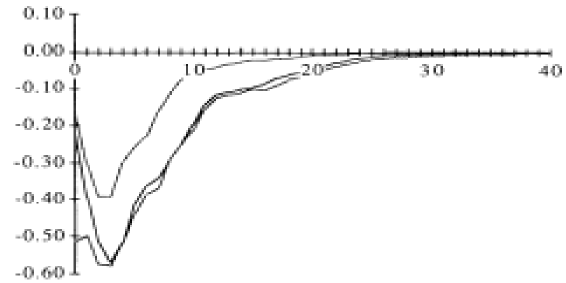


FIGURE 5. UNEMPLOYMENT RESPONSE TO DEMAND

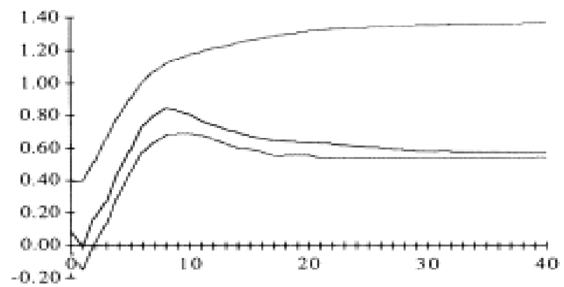


FIGURE 4. OUTPUT RESPONSE TO SUPPLY

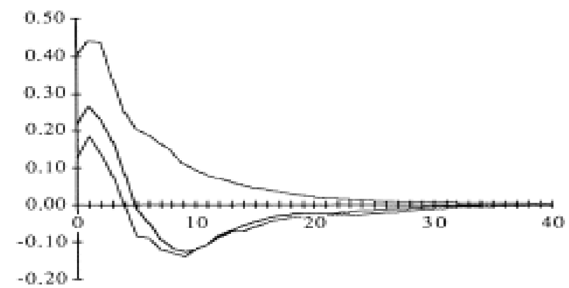


FIGURE 6. UNEMPLOYMENT RESPONSE TO SUPPLY

Source: The Dynamic Effects of Aggregate Demand and Supply Disturbances, (AER) Blanchard and Quah (1989):

TABLE 2—VARIANCE DECOMPOSITION OF OUTPUT AND UNEMPLOYMENT
(CHANGE IN OUTPUT GROWTH AT 1973/1974; UNEMPLOYMENT DETRENDED)

Percentage of Variance Due to Demand:		
Horizon (Quarters)	Output	Unemployment
1	99.0	51.9
	(76.9, 99.7)	(35.8, 77.6)
2	99.6	63.9
	(78.4, 99.9)	(41.8, 80.3)
3	99.0	73.8
	(76.0, 99.6)	(46.2, 85.6)
4	97.9	80.2
	(71.0, 98.9)	(49.7, 89.5)
8	81.7	87.3
	(46.3, 87.0)	(53.6, 92.9)
12	67.6	86.2
	(30.9, 73.9)	(52.9, 92.1)
40	39.3	85.6
	(7.5, 39.3)	(52.6, 91.6)

Source: The Dynamic Effects of Aggregate Demand and Supply Disturbances, (AER) Blanchard and Quah (1989):

TABLE 2A—VARIANCE DECOMPOSITION OF OUTPUT AND UNEMPLOYMENT
(NO DUMMY BREAK, TIME TREND IN UNEMPLOYMENT)

Percentage of Variance Due to Demand:		
Horizon (Quarters)	Output	Unemployment
1	83.8	79.7
	(59.4, 93.9)	(55,3, 92.0)
2	87.5	88.2
	(62.8, 95.4)	(58.9, 95.2)
3	83.4	93.5
	(58.8, 93.3)	(61.3, 97.5)
4	78.9	95.7
	(53.5, 90.0)	(63.9, 98.2)
8	52.5	88.9
	(31.4, 68.6)	(63.5, 94.5)
12	37.8	79.7
	(21.3, 51.4)	(58.8, 90.3)
40	18.7	75.9
	(7.4, 23.5)	(56.9, 88.6)

Source: The Dynamic Effects of Aggregate Demand and Supply Disturbances, (AER) Blanchard and Quah (1989):

4 The technology shocks and hours debate Gali (AER 1999), Christiano, Eichenbaum and Vigfusson (NBER WP, 2003)

This is a nice example of how SVAR models can be used in order to distinguish among competing models of the business cycles.

1) RBC technology important source of business cycles.

2) Other models (sticky prices) tech shocks not so important.

Response of hours worked very important in distinguish among theories

1) RBC hours increase.

2) Other hours fall

4.1 The model

- Technology shock: $z_t = z_{t-1} + \eta_t$ η_t = technology shock

- Monetary Policy: $m_t = m_{t-1} + \xi_t + \gamma\eta_t$ where ξ_t = monetary policy shock.

- Equilibrium:

$$\begin{aligned}\Delta x_t &= \left(1-\frac{1}{\varphi}\right)\Delta\xi_t + \left(\frac{1-\gamma}{\varphi}+\gamma\right)\eta_t + (1-\gamma)\left(1-\frac{1}{\varphi}\right)\eta_{t-1} \\ n_t &= \frac{1}{\varphi}\xi_t - \frac{(1-\gamma)}{\varphi}\eta_t\end{aligned}$$

or

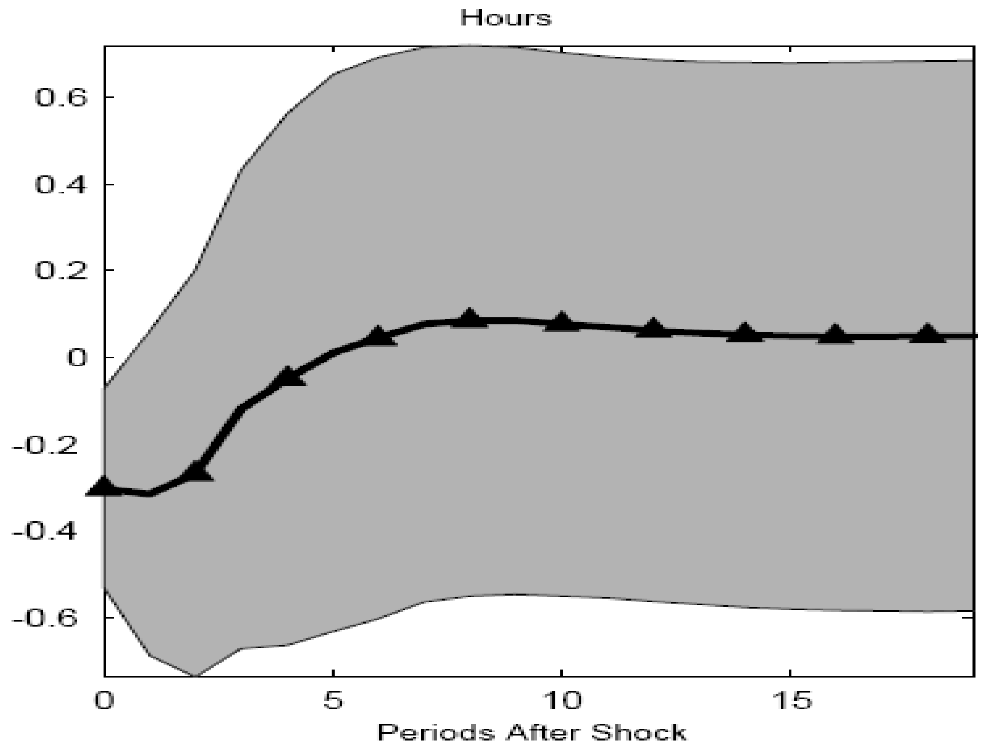
$$\begin{pmatrix}\Delta x_t \\ n_t\end{pmatrix} = \begin{pmatrix}\left(\frac{1-\gamma}{\varphi}+\gamma\right)+(1-\gamma)\left(1-\frac{1}{\varphi}\right)L & \left(1-\frac{1}{\varphi}\right)(1-L) \\ \frac{-(1-\gamma)}{\varphi} & \frac{1}{\varphi}\end{pmatrix}\begin{pmatrix}\eta_t \\ \xi_t\end{pmatrix} \tag{3}$$

In the long run $L = 1$

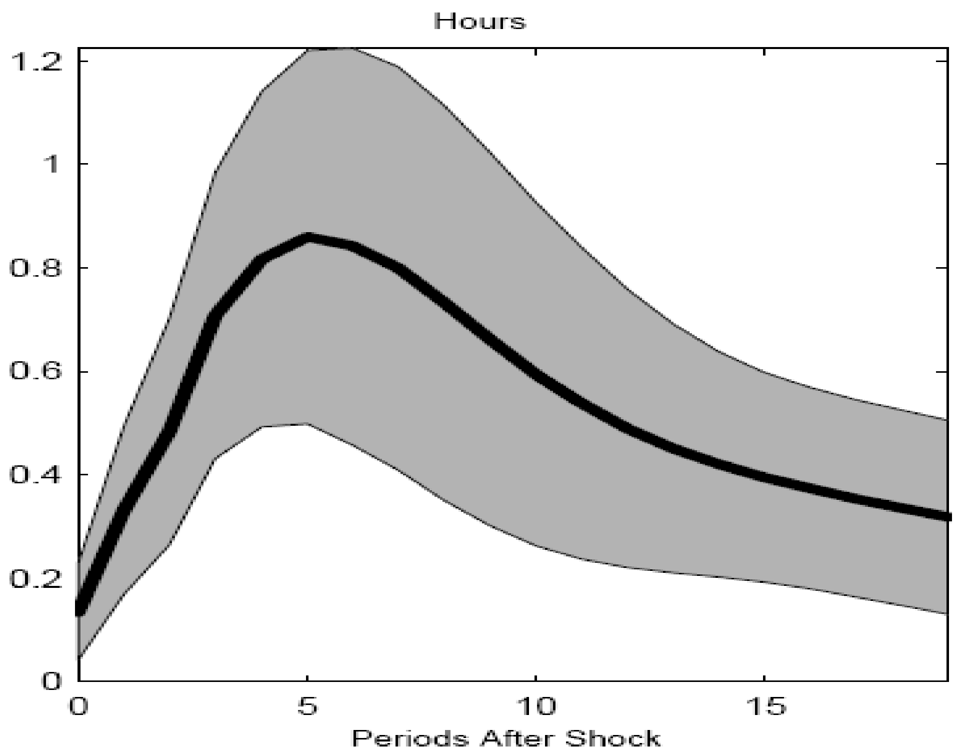
$$\begin{pmatrix}\Delta x_t \\ n_t\end{pmatrix} = \begin{pmatrix}\left(\frac{1-\gamma}{\varphi}+\gamma\right)+(1-\gamma)\left(1-\frac{1}{\varphi}\right) & 0 \\ \frac{-(1-\gamma)}{\varphi} & \frac{1}{\varphi}\end{pmatrix}\begin{pmatrix}\eta_t \\ \xi_t\end{pmatrix} \tag{4}$$

that is only the technology shocks affects labor productivity.

Note the model prediction. If monetary policy is not completely accommodative $\gamma < 1$ then the response of hours to a technology shock $\frac{-(1-\gamma)}{\varphi}$ is negative.



Source: What Happens After a Technology Shock?... Christiano Eichenbaum and Vigfusson NBER WK (2003)



Source: What Happens After a Technology Shock?... Christiano Eichenbaum and Vigfusson NBER WK (2003)

Table 1: Contribution of Technology Shocks to Variance, Bivariate System

Level Specification

	Forecast Variance at Indicated Horizon					
Variable	1	4	8	12	20	50
Output	81.1	78.1	86.0	89.1	91.8	96
Hours	4.5	23.5	40.7	45.4	47.4	48.3

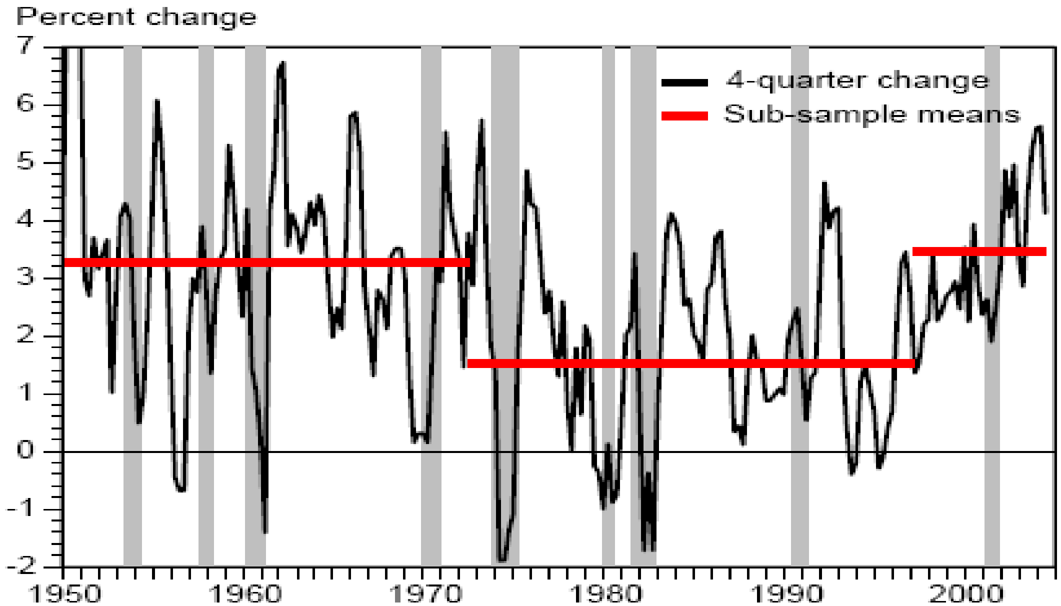
Difference Specification

	Forecast Variance at Indicated Horizon					
Variable	1	4	8	12	20	50
Output	16.5	11.7	17.9	20.7	22.3	23.8
Hours	21.3	6.4	2.3	1.6	1.0	0.5

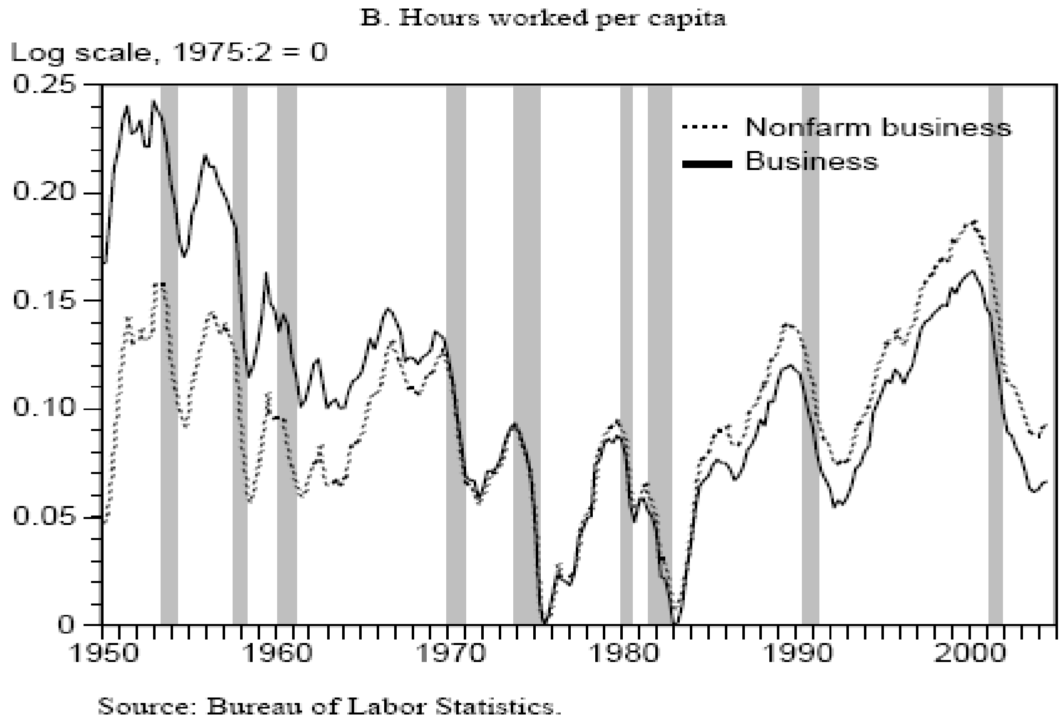
Source: What Happens After a Technology Shock?... Christiano Eichenbaum and Vigfusson NBER WK (2003)

Figure 1: Productivity and Hours

A. Labor productivity, business sector



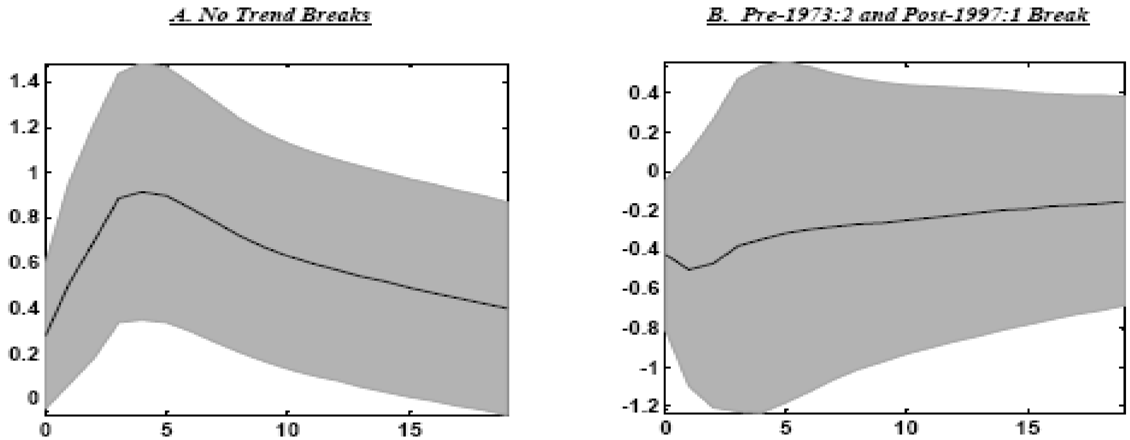
Source: Trend Breaks, Long-Run Restrictions, and Contractionary Technology Improvements, JME John Fernald (2007)



Source: Trend Breaks, Long-Run Restrictions, and Contractionary Technology Improvements, JME John Fernald (2007)

Figure 2. Impulse Responses from Bivariate Specification

Response of Hours to a Technology Shock



Source: Trend Breaks, Long-Run Restrictions, and Contractionary Technology Improvements, JME John Fernald (2007)

5 Government spending shocks

• Understanding the effects of government spending shocks is important for policy authorities but also to assess competing theories of the business cycle.

• Keynesian theory: $G \uparrow$, $Y \uparrow$, $C \uparrow$ because of the government spending multiplier.

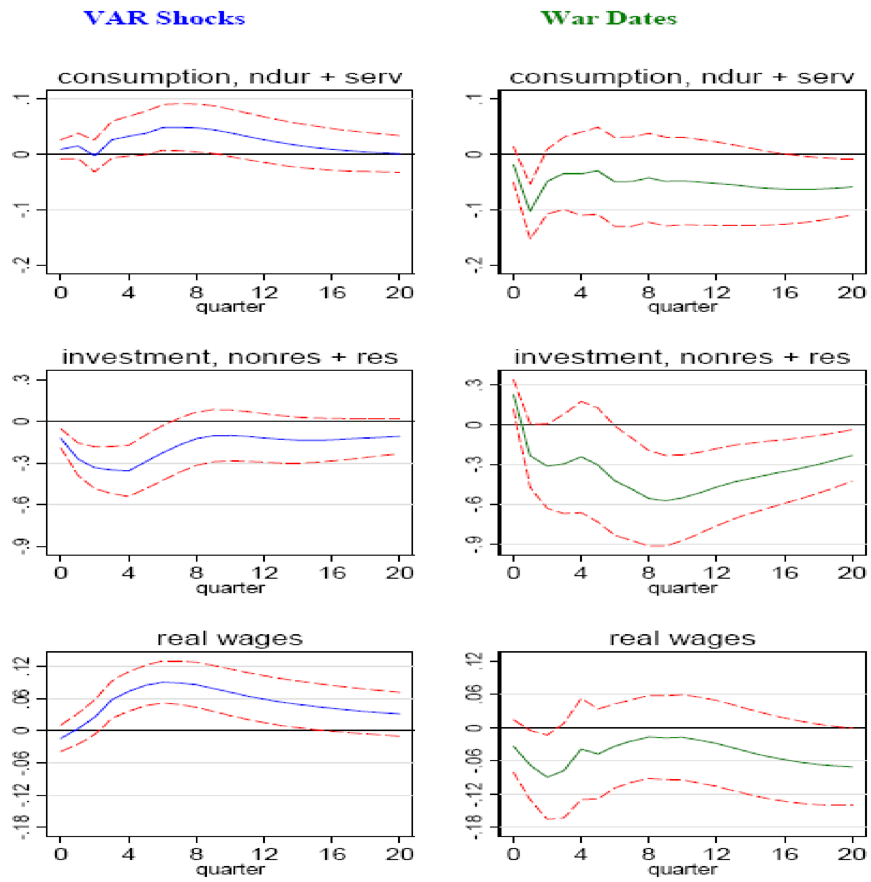
• RBC theory: $G \uparrow$, $C \downarrow$ because of a negative wealth effect.

• Empirical point of view disagreement.

5.1 Government spending shocks: Blanchard and Perotti (QJE 2002)

• BP (originally) use a VAR for real per capita taxes, government spending, and GDP with the restriction that government spending does not react to taxes and GDP contemporaneously, Cholesky identification with government spending ordered first. The government spending shock is the first one (quadratic trend four lags).

• When augmented with consumption consumption increases.



Source: IDENTIFYING GOVERNMENT SPENDING SHOCKS: IT'S ALL IN THE TIMING Valerie A. Ramey, QJE

5.2 Government spending shocks: Ramey and Shapiro (1998)

• Ramey and Shapiro (1998) use a narrative approach to identify shocks to government spending.

• Focus on episodes where Business Week suddenly forecast large rises in defense spending induced by major political events that were unrelated to the state of the U.S. economy (exogenous episodes of government spending).

• Three of such episodes: Korean War, The Vietnam War and the Carter-Reagan Buildup + 9/11.

• The military date variable takes a value of unity in 1950:3, 1965:1, 1980:1, and 2001:3, and zeros elsewhere.

• To identify government spending shocks, the military date variable is embedded in the standard VAR, but ordered before the other variables.

• Both methodologies have drawbacks.

• VARs: shocks are often anticipated (fiscal foresight shocks may be not invertible)

• War Dummy: few observations, subjective.

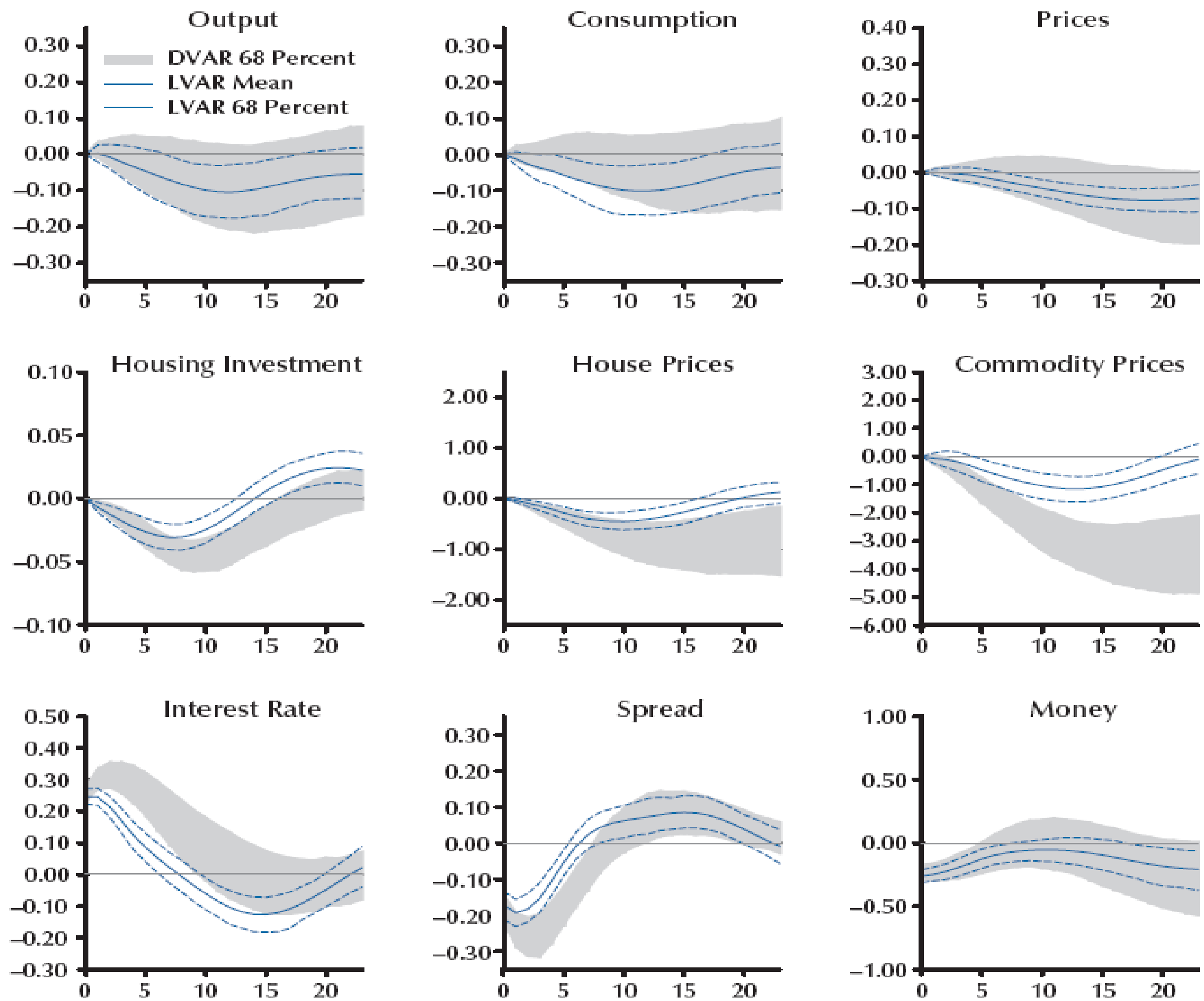
6 Monetary policy and housing

- Central question: how does monetary policy affects house prices?

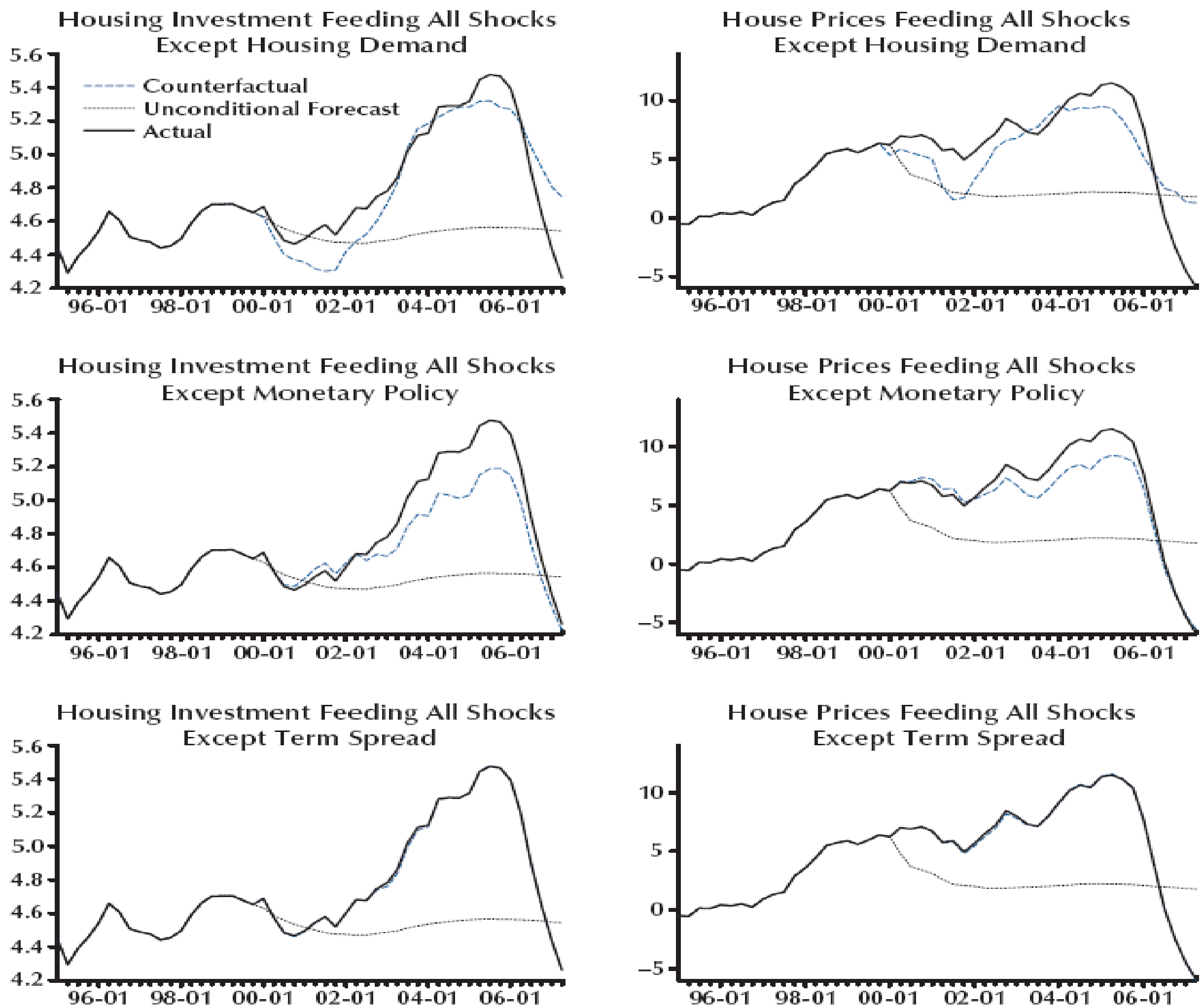
- Jarocinski and Smets (2008) addresses this question.

- Strategy:

1. Estimate a VAR nine variables (including: short term interest rate, interest rate spread, housing investment share of GDP, real GDP, real consumption, real hours prices, prices, commodity price index and a money indicator.
2. Identify the monetary policy shock using the restriction that the shock does not affect prices and output contemporaneously but affect the short term interest rate, the spread and the money stock and analyze the impulse response functions.
3. Shut down the identified shock and study the counterfactual path of housing prices over time.



Source: Jarocinki and Smets (2008)



Source: Jarocinki and Smets (2008)

Table 2A

Shares of Housing Demand, Monetary Policy, and Term Spread Shocks in Variance Decompositions, DVAR

Variable	Shock	Horizon			
		0	3	11	23
Output	Housing	0.016	0.034	0.052	0.062
	Monetary policy	0.000	0.004	0.021	0.039
	Term premium	0.000	0.003	0.015	0.028
Consumption	Housing	0.005	0.018	0.033	0.055
	Monetary policy	0.000	0.003	0.015	0.029
	Term premium	0.000	0.005	0.034	0.063
Prices	Housing	0.002	0.013	0.120	0.166
	Monetary policy	0.000	0.003	0.014	0.037
	Term premium	0.000	0.006	0.034	0.046
Housing investment	Housing	0.521	0.579	0.382	0.291
	Monetary policy	0.000	0.015	0.175	0.136
	Term premium	0.000	0.005	0.023	0.062
House prices	Housing	0.535	0.554	0.410	0.242
	Monetary policy	0.000	0.010	0.068	0.083
	Term premium	0.000	0.002	0.021	0.060
Commodity prices	Housing	0.027	0.028	0.041	0.085
	Monetary Policy	0.000	0.012	0.167	0.222
	Term premium	0.000	0.004	0.018	0.055
Interest rate	Housing	0.037	0.061	0.165	0.178
	Monetary policy	0.752	0.496	0.192	0.166
	Term premium	0.000	0.023	0.076	0.088
Spread	Housing	0.090	0.050	0.177	0.186
	Monetary policy	0.223	0.303	0.214	0.206
	Term premium	0.336	0.245	0.146	0.134
Money	Housing	0.060	0.044	0.062	0.099
	Monetary policy	0.204	0.141	0.044	0.045
	Term premium	0.013	0.042	0.129	0.135

NOTE: The reported shares are averages over the posterior distribution and relate to the (log) level variables.

Source: Jarocinki and Smets (2008)