2

THE PROBABILISTIC STRUCTURE OF TIME SERIES DATA

2.1 Introduction: what is a time-series ?

In the previous chapter we have introduced time-series to show that one of the fundamental properties necessary to perform valid estimation and inference in the linear model is generally violated by time-series. In this chapter we shall discuss this issue at greater depth and length by defining precisely time series and the fundamental concepts to analyze them, by illustrating how the problem introduced can be resolved in the context of stationary time-series, to finally extend our discussion to non-stationarity and cointegration.

We write a time-series as

$$\{x_1, x_2, ...x_T\} \ or \ \{x_t\}, \ t = 1, ...T$$

where t is an index denoting the period in time in which x occurs. We shall treat x_t as a random variables, hence a time-series is a sequence of random variables ordered in time. Such sequence is known as a stochastic process. The probability structure of a sequence of random variables is determined by the joint distribution of a stochastic process.

A possible probability model for such a joint distribution is :

$$x_t = \epsilon_t, \ \epsilon_t \sim n.i.d. \left(0, \sigma_\epsilon^2\right)$$

i.e. x_t is normally independently distributed over time with constant variance and zero mean. In other words x_t is a *white-noise* process. A white-noise process is not a proper model for most macroeconomic time-series because it does not feature their most common characteristic, namely persistence. To show the point consider the data-set USUK.XLS which contains, in EXCEL format, quarterly time series data for nominal and real personal disposable income and consumption in the UK and the US over the sample 1959:1-1998:1. The data-set, retrieved from DATASTREAM, contains nine variables:

TABLE 1: Dataset USUK.XLS

ukpdispid	personal disposable income in the UK at constant 1992 prices
uspdispid	personal disposable income in the US at constant 1992 prices
uscndurb	consumption of durable goods in the US at current prices
uscndurb	consumption of durable goods in the US at constant 1992 prices
uscnnondb	consumption of non-durable goods in the US at current prices
uscnnondd	consumption of non-durable goods in the US at constant 1992 prices
uscnservb	consumption of services in the US at current prices
uscnservd	consumption of services in the US at constant 1992 prices

All series are adjusted for seasonality. To assess the behaviour of an typical economic time series against the benchmark of the white-noise process, we have imported all series in an E-Views workfile and run the following routine:

```
smpl 1959:1 1998:1
genr lyus=log(uspdispid)
genr WN= 8.03+0.36*nrnd
plot WN lyus
```

The routine generates the log of US real disposable income and an artificial series defined as a constant (8.03) plus a normal random variable with zero mean and standard deviation of 0.36, where 8.03 and 0.36 are respectively the sample mean and the sample standard deviation of lyus. Having generated the series the program plots them to obtain the following result:

Figure 2.1 clearly shows that the white noise model does not capture the interesting property of persistence that motivates the study of time series. In order to construct more realistic models combinations of ϵ_t. We shall concentrate on a class of models created by taking linear combinations of white noise, the ARMA models:

$$\begin{aligned}
AR(1): &\quad x_t = \rho x_{t-1} + \epsilon_t \\
MA(1): &\quad x_t = \epsilon_t + \theta \epsilon_{t-1} \\
AR(p): &\quad x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + ... + \rho_p x_{t-p} + \epsilon_t \\
MA(q): &\quad x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} \\
ARMA(p,q): &\quad x_t = \rho_1 x_{t-1} + ... + \rho_p x_{t-p} + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}
\end{aligned}$$

In case it is not already clear, we shall show why ARMA models are obtained by taking linear combinations of white noise in the next section, where we discuss the strictly necessary fundamentals to analyze time series.

Note that each of the above models can be easily put to action to generate the equivalent time-series by modifying appropriately and running the following programme in Eviews, which generates an AR(1) series:

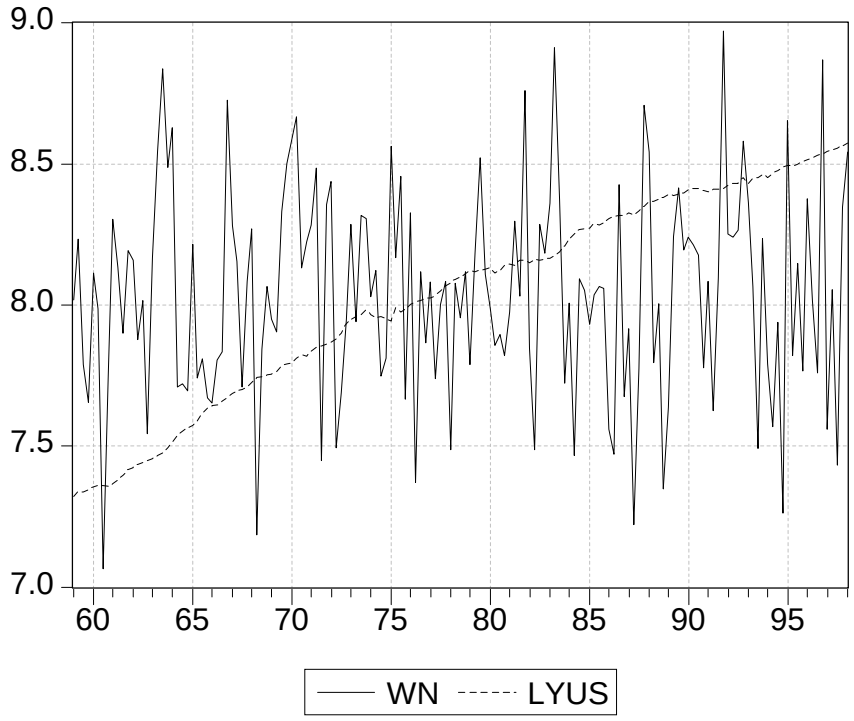

FIG. 2.1. A *white-noise* process and the log of US real disposable income

```
smpl 1 1
genr X=0
smpl 2 200
series x=0.5*x(-1) +NRND
```

The programme above generates a sample of 200 observations from an AR(1) model with $\rho = 0.5$. The series is first initialized for the first observations, the command series then generates the series for the specified process, each observation is 0.5 time the previous observation plus a random disturbance drawn from a serially independent standard normal distribution.

The time series behaviour of the generated X is plotted in Figure 2.2.

The following modified version of the programme will generate an ARMA(1,1) series:

```
smpl 1 1
genr X=0
smpl 1 200
genr u=NRND
```

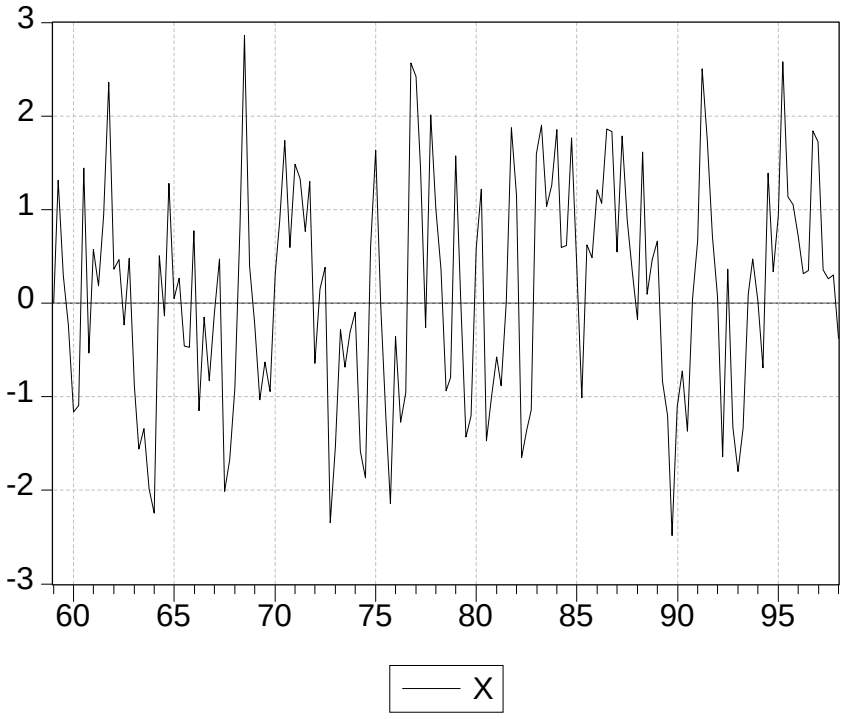

FIG. 2.2. A stationary ARMA(1,1) process

```
smpl 2 200
series x=0.5*x(-1) +u +0.4*u(-1)
```

2.2 Analyzing time-series: the fundamentals.

To illustrate empirically all the fundamentals we consider an interesting member of the the ARMA family: the AR model with drift :

$$\begin{aligned} x_t &= \rho_0 + \rho_1 x_{t-1} + \epsilon_t \\ \epsilon_t &\sim n.i.d.\left(0, \sigma_\epsilon^2\right) \end{aligned} \tag{2.1}$$

Given that each realization of our stochastic process is a random variable, the first relevant fundamental is the density of each observations. In particular, we distinguish between conditional and unconditional densities. Having introduced these two concepts we shall define and discuss stationarity, we then generalize form our specific member to the whole family of ARMA models, to end this

section with a discussion of deterministic and stochastic trends and de-trending methods. Note that at this introductory stage we concentrate almost exclusively on univariate models. We do so just for the sake of exposition. After the completion of our introductory tour, we shall concentrate on multivariate models, which are the focus of this book.

2.2.1 *Conditional and unconditional densities*

We distinguish between conditional and unconditional density of a time-series. The unconditional density is obtained under the hypothesis that no observation on the time series is available, while conditional densities are based on the observation of some realization of the random variables. In the case of the time series we derive unconditional by putting ideally ourselves at the moment in time preceeding the observation of any realization of the time series. At the moment the information set is given only by the knowledge of the process generating the observations. As observations become available conditional densities can be computed. As distributions are summarized by their moments, let us illustrate the difference between conditional and unconditional densities by looking at our AR(1) model.

The moments of the density of x_t conditional upon x_{t-1} are immediately obtained from (2.1)as follows:

$$\begin{aligned} E\left(x_t \mid x_{t-1}\right) &= \rho_0 + \rho_1 x_{t-1} \\ Var\left(x_t \mid x_{t-1}\right) &= \sigma_\epsilon^2 \\ Cov\left[\left(x_t \mid x_{t-1}\right), \left(x_{t-j} \mid x_{t-j-1}\right)\right] &= 0 \;\; for\; each\; j \end{aligned}$$

To derive the moments of the density of x_t conditional upon x_{t-2}, we need to substitute for x_{t-1} in terms of x_{t-2} from (2.1) to obtain:

$$\begin{aligned} E\left(x_t \mid x_{t-2}\right) &= \rho_0 + \rho_0\rho_1 + \rho_1^2 x_{t-2} \\ Var\left(x_t \mid x_{t-2}\right) &= \sigma_\epsilon^2\left(1+\rho_1^2\right) \\ Cov\left[\left(x_t \mid x_{t-2}\right), \left(x_{t-j} \mid x_{t-j-2}\right)\right] &= \rho_1\sigma_\epsilon^2 \quad for\; j = 1 \\ Cov\left[\left(x_t \mid x_{t-2}\right), \left(x_{t-j} \mid x_{t-j-2}\right)\right] &= 0 \quad for\; j > 1 \end{aligned}$$

Finally, unconditional moments are derived by substituing recursively from (2.1) to express x_t as a function of information available at time time t_0, the moment before we start observing realizations of our process.

$$
\begin{aligned}
E(x_t) &= \rho_0 \left(1 + \rho_1 + \rho_1^2 + ...\rho_1^{t-1}\right) + \rho_1^t x_0 \\
Var(x_t) &= \sigma_\epsilon^2 \left(1 + \rho_1^2 + \rho_1^4 + ...\rho_1^{2t-2}\right) \\
\gamma(j) &= Cov(x_t, x_{t-j}) = \rho_1^j \, Var(x_t) \\
\rho(j) &= \frac{Cov(x_t, x_{t-j})}{\sqrt{Var(x_t) \, Var(x_{t-1})}} = \frac{\rho_1^j \, Var(x_t)}{\sqrt{Var(x_t) \, Var(x_{t-1})}}
\end{aligned}
$$

Note that $\gamma(j)$ and $\rho(j)$ are function of j, known respectively as the autocovariance function and the autocorrelation function.

2.2.2 *Stationarity*

A stochastic process is said to be stricly stationary if its joint density function does not depend on time. More formally a stochastic process is stationary if ,for each $j_1, j_{2,...} j_n$, the joint distribution,

$$f(x_t, x_{t+j_1}, x_{t+j_2}, x_{t+j_n})$$

does not depend on t.

A stochastic process is said to be covariance stationary if its two first undiconditional moments do not depend on time, i.e. if the following relations are satisfied for each h,i,j:

$$
\begin{aligned}
E(x_t) &= E(x_{t+h}) = \mu \\
E\left(x_t^2\right) &= E\left(x_{t+h}^2\right) = \mu_2 \\
E(x_{t+i} x_{t+j}) &= \mu_{ij}
\end{aligned}
$$

In the case of our AR(1) process the condition for stationarity is that $|\rho_1| < 1$. In fact, when such condition is satisfied we have:

$$
\begin{aligned}
E(x_t) &= E(x_{t+h}) = \frac{\rho_0}{1 - \rho_1} \\
Var(x_t) &= Var(x_{t+h}) = \frac{\sigma_\epsilon^2}{1 - \rho_1^2} \\
Cov(x_t, x_{t-j}) &= \rho_1^j \, Var(x_t)
\end{aligned}
$$

on the other hand it easily shown that, when $|\rho_1| = 1$, the process is non stationary.

In fact we have:

$$
\begin{aligned}
E(x_t) &= \rho_0 t + x_0 \\
Var(x_t) &= \sigma_\epsilon^2 t \\
Cov(x_t, x_{t-j}) &= \sigma_\epsilon^2 \ (t - j)
\end{aligned}
$$

To illustrate graphically the properties of different AR process we generate, using the programme in E-views described above, we generate three AR process with ρ_1 set to 0.6 (series X1), 0.8 (series X2), and 1 (series X3) respectively. To allow direct comparison we do not include a drift in all process so for all of them we have $\rho_0 = 0$. The time-series behaviour of the three-processes is reported in Figure 2.3.

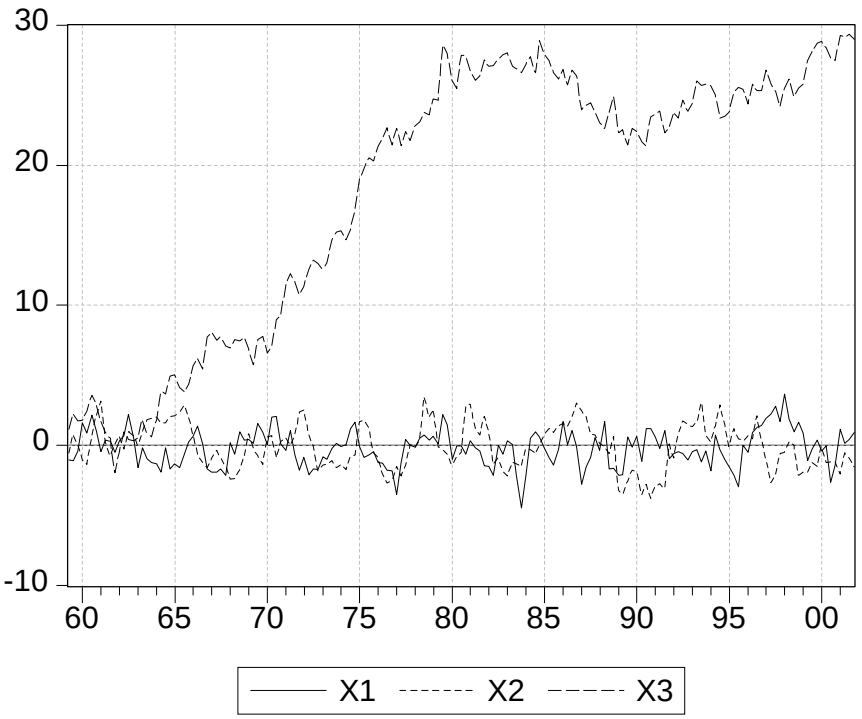

FIG. 2.3. First order autoregressive processes with $\rho_1 = 0.6$ (X1), $\rho_1 = 0.8$ (X2), $\rho_1 = 1$ (X3)

Note that X1 and X2 tend to revert towards their unconditional mean rather quickly. The unconditional mean of X3 is also zero but X3 does not show any tendency for reverting towards its mean, in fact, as the sample size grows, the variance of X3 increases without any bound.

2.2.3 *ARMA processes*

Before introducing the fundamentals of time-series we have asserted that white-noise processes were too simplistic to describe economic time series and that a closer fit could be obtained by considering combination of white-noises. we have

then introduced ARMA models and discussed the fundamentals to understand their properties, but we have not yet shown that ARMA models can be considered as combination of white-noise processes. The point is shown by considering a time-series as a polynomial distributed lag of a white-noise process:

$$\begin{aligned} x_t &= u_t + b_1 u_{t-1} + b_2 u_{t-2} + ...b_n u_{t-n} \\ &= \left(1 + b_1 L + b_2 L^2 + ... + b_n L^n\right) u_t \\ &= b(L) u_t \end{aligned}$$

where L is the lag operator. The Wald-decomposition theorem, which states that any stationary stochastic process could be expressed as the sum of a deterministic component and of a stochastic moving average component warrant generality of our representation. However in general, to describe successfully a time-series, a very high order in the polynomial b(L) is required. This feature can be problematic for estimation, given the usual limitations for sample sizes. This potential problem if the polynomial b(L) can be represented as the ratio of two polynomial of lower order. In this case we have:

$$\begin{aligned} x_t &= b\left(L\right) u_t \\ &= \frac{a\left(L\right)}{c\left(L\right)} u_t \\ c\left(L\right) x_t &= a\left(L\right) u_t \end{aligned} \tag{2.2}$$

(2.2) is an ARMA process. The process is stationary when the roots of $c\left(L\right) = 0$ lie outside the unit circle. The MA component is said to be invertible when the roots of $a\left(L\right) = 0$ lie outside the unit circle. Invertibility of the MA components allow to represent it as an autoregressive process.

To illustrate how the autocovariance and the autocorrelation functions of an ARMA model are derived, we consider the simplest case: the ARMA(1,1) process:

$$\begin{aligned} x_t &= c_1 x_{t-1} + \epsilon_t + a_1 \epsilon_{t-1} \\ \left(1 - c_1 L\right) x_t &= \left(1 + a_1 L\right) \epsilon_t \end{aligned} \tag{2.3}$$

(2.3) can be re-written as:

$$\begin{aligned} x_t &= \frac{1 + a_1 L}{1 - c_1 L} \epsilon_t \\ &= \left(1 + a_1 L\right) \left(1 + c_1 L + \left(c_1 L\right)^2 + ...\right) \epsilon_t \\ &= \left[1 + \left(a_1 + c_1\right) L + c_1 \left(a_1 + c_1\right) L^2 + c_1^2 \left(a_1 + c_1\right) L^3 + ...\right] \epsilon_t \end{aligned}$$

Then we have

$$Var(x_t) = \left[1 + (a_1 + c_1)^2 + c_1^2 (a_1 + c_1)^2 + ...\right] \sigma_\epsilon^2$$
$$= \left[1 + \frac{(a_1 + c_1)^2}{1 - c_1^2}\right] \sigma_\epsilon^2$$

$$Cov(x_t, x_{t-1}) = \left[(a_1 + c_1) + c_1 (a_1 + c_1) + c_1^2 (a_1 + c_1) + ...\right] \sigma_\epsilon^2$$
$$= \left[(a_1 + c_1) + \frac{c_1 (a_1 + c_1)^2}{1 - c_1^2}\right] \sigma_\epsilon^2$$

Hence

$$\rho(1) = \frac{Cov(x_t, x_{t-1})}{Var(x_t)}$$
$$= \frac{(1 + a_1 c_1)(a_1 + c_1)}{1 + c_1^2 + 2a_1 c_1}$$

Successive values for $\rho(j)$ are obtained from the recurrence relation $\rho(j) = c_1 \rho(j-1)$ for $j \geq 2$.

To illustrate the difference between an AR and an ARMA, we have generated an AR(0.7) process and an ARMA (0.7, 0.4) process in E-Views. The two autocorrelation functions (for lags up to 10) are reported in Table 2.

TABLE 2: Autocorrelation functions

AR (0.7)	ARMA (0.7,0.4)
0.712	0.836
0.561	0.639
0.437	0.491
0.304	0.364
0.254	0.305
0.270	0.305
0.270	0.313
0.298	0.326
0.279	0.323
0.296	0.316

Note that the autocorrelation of the ARMA(1,1) process is higher than the autocorrelation of the AR(1) process, this is because $a_1 > 0$.

2.2.4 *Deterministic and Stochastic Trends*

Figure 2.1 at the beginning of this chapter shows that macroeconomic time series, beside being persistent, feature (generally) upwarding trends. Non-stationarity of time-series is a possible manifestation of a trend. Consider for example the random walk with drift:

$$x_t = a_0 + x_{t-1} + \epsilon_t$$
$$\epsilon_t \sim n.i.d.\left(0, \sigma_\epsilon^2\right)$$

In this case recursive substitution yields:

$$x_t = x_0 + a_0 t + \sum_{i=0}^{t-1} \epsilon_{t-i} \tag{2.4}$$

which shows that the non-stationary series contains both a deterministic ($a_0 t$) and a stochastic $\left(\sum_{i=0}^{t-1} \epsilon_{t-i}\right)$ trend.

One of the easiest way to make a non-stationary series stationary is by differencing it:

$$\Delta x_t = x_t - x_{t-1} = (1 - L)\, x_t = a_0 + \epsilon_t$$

In general if a time series needs to be differenced k times to be stationary, then that series is said to be integrated of order k or $I(k)$. Our random walk is I(1).When the $d - th$ difference of a time-series x, $\Delta^d x_t$, can be represented by an ARMA(p, q) model we say that x_t is an integrated moving-average process of order p, d, q and we denote it as ARIMA(p, d, q).

It interesting to compare the behaviour of integrated process with that of trend stationary process. Trend stationary processes feature only a deterministic trend:

$$z_t = \alpha + \beta t + \epsilon_t \tag{2.5}$$

The z_t process is non-stationary, but the non-stationary is removed just by regressing z_t on a deterministic trend. This is not the case for integrated processes like (2.4) where the removal of the deterministic trend does not deliver a stationary time-series. Deterministic trend have no memory while integrated variables have infinite memory. Both integrated variable and deterministic trend exhibits systematic variations, but in one case the variation is predictable in the other case it is not. This point is easily seen in Figure 2.4 where we report three series for a sample of 200 observations. The series are generated in E-views by running the following programme:

```
smpl 1 1
genr ST1=0
genr ST2=0
smpl 2 200
series ST1= 0.1+ST1(-1) +nrnd
series ST2=0.1+ST2(-1)+nrnd
series DT= 0.1*@trend +nrnd
```

We have a deterministic trend (DT) generated by simulating equation (2.5) with $\alpha = 0, \beta = 0.1$, and a white-noise independently distributed as a standard normal (nrnd), and two integrated series (ST1 and ST2), which are random walks with a drift of 0.1. The only difference between ST1 and ST2 is in the realizations from the error terms, which are different drawings from the same serially independent standard normal distribution.

FIG. 2.4. Deterministic (DT) and stochastic (ST1 and ST2) trends

2.3 Persistence. A Monte-Carlo experiment

Persistence of time-series destroys one of the crucial properties to implement valid estimation and inference in the linear model. We have already seen that in the context of the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

The following property is required to implement valid estimation aand inference

$$E(\boldsymbol{\epsilon} \mid \mathbf{X}) = \mathbf{0} \tag{2.6}$$

Hypothesis (2.6) implies that

$$E(\boldsymbol{\epsilon}_i \mid \mathbf{x}_1, ...\mathbf{x}_i, ...\mathbf{x}_n) = 0 \quad (i = 1, ...n)$$

Think of the simplest time-series model for a generic variable y :

$$y_t = a_0 + a_1 y_{t-1} + \epsilon_t$$

It is clear that if $a_1 \neq 0$, then, although it is true that $E(\epsilon_t \mid y_{t-1}) = 0$, $E(\epsilon_{t-1} \mid y_{t-1}) \neq 0$ and (2.6)is destroyed.

The question is how serious is the problem. To assess intuitively the consequence of persistence we construct a small Monte-Carlo simulation on the short sample properties of the OLS estimator of the parameters in an AR(1) process.

A Monte-Carlo simulation is based on the generation of a sample from a known Data Generating Process(DGP). A set of random numbers from a given distribution is generated first (a normally independent white-noise disturbance in our case) for a sample size of interest (in our case 200 observations) and then the process of interest is constructed (in our case an AR(1) process). When a sample of observations on the process of interest is available, then the relevant parameters can be estimated and their fitted value can be compared with the known true value. For this reason the Monte-Carlo simulation is a sort of controlled experiment. The only potential problem with this procedure is that the set of random numbers drawn is just one possible outcome and the estimates are dependent on the sequence of simulated white-noise residuals. To overcome this problem in a Monte-Carlo study the DGP is replicated many times. For each replication a set of estimates is obtained and then averages across replications of the estimated parameters are computed to be assessed against the known true values.

Our Monte-Carlo simulation is performed by running the following programme in E-Views:

```
genr a1sum=0
```

```
for !i=1 to 500
smpl 1 1
genr y{!i}=10
smpl 2 200
series y{!i}=1+0.9*y{!i}(-1) +nrnd
equation eq.ls y{!i}= c(1)+c(2)*y{!i}(-1)
eq.rls(c,s)
genr a1sum=a1sum+R_c2
next
genr a1mean=a1sum/500
```

The first line of the programme generate a series to store the values of the estimated a_1 in each replication. In the next step we set a counter to keep track of the replications (in the specific case we have 500 of them).The loop for the five hundred replications is then set. In each replications a sample of two hundred observations from an AR(1) is generated and then the autoregressive parameters is estimated. Note that such estimation is performed recursively starting with a sample of five observations and then by adding one observation at the time until the last one. The series of these estimates is stored at each replications with the command eq.rls(c,s). At the end of all replications we have 500 hundred series each containing a series of 195 estimated parameters (the first being the parameter estimated on the sample 1-5, the second being the parameter estimated on the sample 1-6, the last one being the parameter estimated on the full sample). We report the average across replications in Figure 2.5.

From the Figure 2.5 we note that the estimate of a_1 is heavily biased in small samples, but the bias is reduced as the sample gets larger to eventually disappear. In fact, it can be shown analytically that the average of the OLS estimate of a_1 is $a_1\left(1-\frac{2}{T}\right)$. This is an interesting result, which could be generalized. For stationary time-series, the correlation, which destroys the orthogonality between residuals and regressors in the linear regression model, tends to disappear as the distance between observations increases. Therefore, as we shall show in the next section, the finite sample results can be extended to time-series by considering large samples. Such aim is obtained by introducing asymptotic theory.

2.4 The traditional solution: asymptotic theory

Stationary time-series feature time-independent distributions, as a consequence the effect of any specific innovation disappear as time elapses. We shall show in this section of the intuition given by the simple Monte-Carlo simultaion can be extended and asymtptotic theory can be used to perform valid estimation and inference when modelling *stationary* time-series.

2.4.1 *Basic elements of asymptotic theory*

In this section we shall introduce the elements of asymptotyc theory necessary to illustrate how all the results in estimation and inference for the linear model

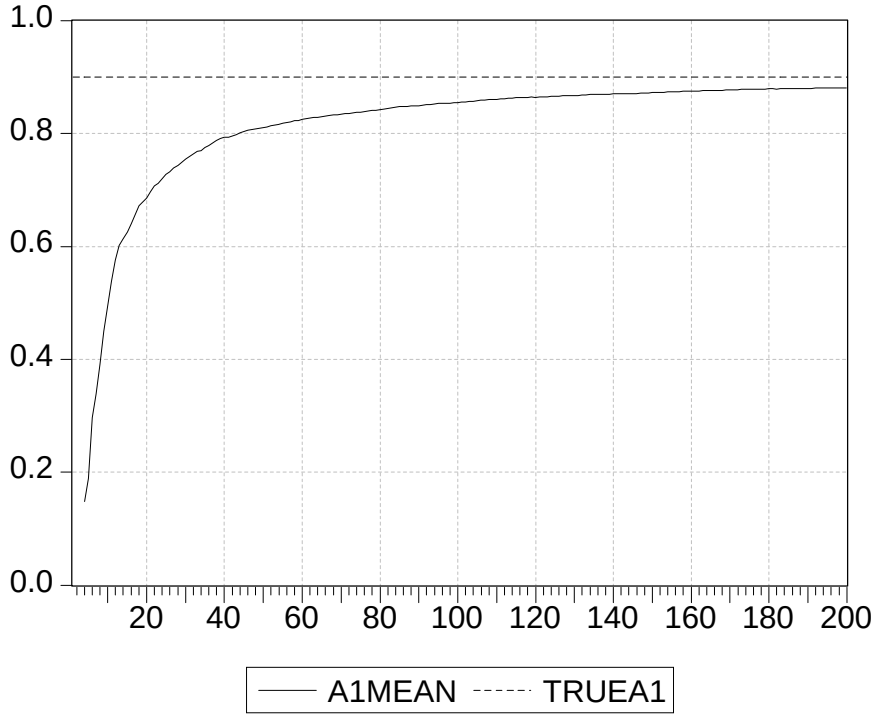

FIG. 2.5. Small sample bias

applied to cross-sectional data in Chapter 1 can be extended to time-series models[1].

Consider a sequence $\{X_T\}$ of random variables with the associated sequence of distribution functions $\{F_T\} = F_1, ..., F_T$, we give the following definitions of convergence for X_T

2.4.1.1 *Convergence in distribution* Given a random variable X with distribution function F, X_T converges in distribution to X if the following equality is satisfied:

$$\lim_{T\to\infty} pr\{X_T < x_0\} = pr\{X < x_0\}$$

for all x_0,where the function $F(x)$ is continuous.

[1]For a formal treatment of all these topics see White([60])

2.4.1.2 *Convergence in probability* Given a random variable X with distribution function F, X_T converges in probability to X if , for each $\epsilon > 0$,the following relation holds:

$$\lim_{T\to\infty} pr\left\{|X_T - X| < \epsilon\right\} = 1$$

Note that convegence in probability implies convergence in distribution.

2.4.1.3 *Central limit theorem (formulation of Lindeberg-Levy)* Given a sequence $\{X_T\}$ of identically and independently distributed random variables with mean μ and finite variance σ^2, defining

$$\bar{X} = \frac{1}{T}\sum_{i=1}^{T} X_i$$

$$\omega = \sqrt{T}\frac{\left(\bar{X} - \mu\right)}{\sigma}$$

ω converges in distribution to a standard normal.

2.4.1.4 *Slutsky's Theorem* For any random variable X_T such that $plim X_T = a$, where a is a constant, given a function $g(\cdot)$ continuous in a, we have that $p\lim g\left(X_T\right) = g\left(a\right)$.

2.4.1.5 *Cramer's Theorem* Given two random variables X_T and Y_T such that Y_T converges in distribution to Y and X_T converges in probability to a constant a, the two following relationships hold:

- $X_T + Y_T$ converges in distribution to $(a + Y)$
- Y_T/a_T converges in distribution to (Y/a)
- $Y_T \cdot a_T$ converges in distribution to $(Y \cdot a)$

Note that all theorems introduced so far are extended to vectors of random variables.

2.4.1.6 *Mann-Wald Theorem* Consider a vector $\mathbf{z}_t$ $(kx1)$ of random variables which satisfies the following property:

$$p\lim T^{-1}\sum_{i=1}^{T}\mathbf{z}_t\mathbf{z}_t' = \mathbf{Q}$$

where $\mathbf{Q}$ is a positive definite matrix. Consider also a sequence ϵ_t of random variables identically and independently distributed with zero mean and finite

variance σ^2, for which finite moments of each order are defined. If $E(\mathbf{z}_t\epsilon_t) = 0$,then we have

$$p\lim T^{-1}\sum_{i=1}^{T}\mathbf{z}_t\epsilon_t = \mathbf{0}, \sqrt{\frac{1}{T}}\sum_{i=1}^{T}\mathbf{z}_t\epsilon_t \xrightarrow{d} N\left(0, \sigma^2\mathbf{Q}\right)$$

2.4.2 *Application to models for stationary time-series*

Consider the following time-series model:

$$y_t = \alpha y_{t-1} + \beta x_t + u_t$$

where x_t is a stationary variable and $|\alpha| < 1$. As already shown $E(y_t u_{t-i}) \neq 0$ and the OLS estimator of α is biased.

Re-write the model as :

$$\begin{aligned} y_t &= \mathbf{z}_t\boldsymbol{\gamma} + u_t \\ \mathbf{z}_t &= \left[\, y_{t-1}\ x_t \,\right] \\ \boldsymbol{\gamma} &= \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \end{aligned}$$

By applying the Mann-Wald results we can derive the asymptotic distribution of the OLS estimator of $\gamma, \widehat{\gamma}$:

$$\widehat{\gamma} \xrightarrow{d} N\left[\gamma, \sigma^2\mathbf{Q}^{-1}\right]$$

and all the finite sample results available for cross-section can be extended to stationary time-series just by considering large-sample theory.

2.5 Stochastic-trends and spurious regressions

From what we have discussed so far it should be clear that most econometric analysis is based on the variance and covariance among variables. In the case of indepedent sampling (cross-section) we can use finite sample moments for estimation and inference, in the case of stationary time-series the consideration of moments in large samples can solve the problems peculiar to time-series in small samples. Within this framework it should be immediately clear that non-stationary causes problems. In fact, we know unconditional moments are not defined for non-stationary time-series. Consider, for the sake of illustration, an OLS regression of an I(0) variable y_t on an I(1) variable x_t. The OLS estimator of the regression y_t on x_t converges to zero as the sample size increases, in fact the variance of x_t, being divergent, dominates the covariance between the two

variables. In general asymptotic theory is not applicable to non-stationary time-series (see, for example, Hatanaka([23]) and Maddala-Kim([39]) . So, unless all the trends observed in time-series are deterministic, the solution of reverting to asymptotic theory is not directly accessible.

To give an intuition of the importance of non-stationarity in time-series and to illustrate the problems related to non-stationarity, consider the results of a "crazy" regression, obtained by relating the log of consumption in the US to the log of personal disposable income in the UK :

TABLE 3: Regressing US consumption on UK disposable income
Sample : 1959 : 1 1998 : 1 , *Dependent Variable LCUS*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.612676	0.160374	-34.99740	0.0000
LYUK	1.208592	0.014419	83.81657	0.0000
R-squared 0.978413, S.E. of regression 0.052291, DW-stat 0.140469				

Note that the regression features a very high R^2 and the UK disposable income is very significant in explaining US consumption. We have a case of a spurious regression, which witnesses the relevance of non-stationarity in economic time-series. To elaborate on this point consider the two following simple univariate time-series models for LYUS and LYUK.

TABLE 4: Univariate Time-series models for US consumption and UK disposable income

	Coefficient	Std. Error	t-Statistic	Prob.
Dependent variable LCUS				
C	0.039	0.008	4.91	0.0000
LCUS(-1)	0.996	0.001	964.9	0.0000
R-squared 0.999835, S.E. of regr 0.004537, DW stat 1.397403.				
Dependent variable LCUS				
C	0.050	0.049	1.00	0.3185
LYUK(-1)	0.996	0.004	222	0.0000
R-squared 0.999835, S.E. of regr 0.004537, DW stat 1.397403.				

despite the simplicity of the two time-series models for LYUS and LYUK, we note that they can both be approximated by random walk models:

$$
\begin{aligned}
LCUS_t &= a_0 + LCUS_{t-1} + \epsilon_{1t} \\
LYUK_t &= b_0 + LYUK_{t-1} + \epsilon_{2t} \\
\epsilon_{1t} &\sim n.i.d.\left(0, \sigma^2_{\epsilon_1}\right) \\
\epsilon_{2t} &\sim n.i.d.\left(0, \sigma^2_{\epsilon_2}\right)
\end{aligned}
$$

As we already know, recursive substitution yields:

$$
\begin{aligned}
LCUS_t &= LCUS_0 + a_0 t + \sum_{i=0}^{t-1} \epsilon_{1t-i} \\
LYUK_t &= LYUK_0 + b_0 t + \sum_{i=0}^{t-1} \epsilon_{2t-i}
\end{aligned}
$$

When the following model is estimated

$$
LCUS_t = \widehat{\alpha} + \widehat{\beta} LYUK_t + \widehat{u}_t,
$$

the coefficient $\widehat{\beta}$ is significant as both series have a deterministic trend. However, in order to have a non-spurious relation we would need that the regression removes also the stochastic trend from the dependent variables, leaving stationary residuals. If this does not happen, then the correlation we observe can be labelled as spurious. We report in Figure 2.6 the residuals from the OLS regression of LCUS on LYUK,

visual impression confirms the intuition that the regression has delivered a spurious relation, having not removed the stochastic trend form the non-stationary dependent variable. The reported DW statistic of 0.14 gives a more formal background to the visual impression. In fact the Durbin-Watson statistic, originally designed to test for the presence of first order autocorrelation in the residuals, can be re-calibrated to test for stationarity. We have

$$
DW = \frac{\sum_{i=2}^{T} \left(\widehat{u}_t - \widehat{u}_{t-1}\right)^2}{\sum_{i=2}^{T} \widehat{u}_t} \simeq 2\left(1 - \widehat{\rho}\right)
$$

where $\widehat{\rho}$ is the OLS coefficient from the regression of $\widehat{u}_t$ on $\widehat{u}_{tt-1}$. The test was originally tabulated to test the hypothesis $H_0 : \rho = 0$, but critical values for the null of non-stationarity $H_0 : \rho = 1$, have been provided by Sargan-Bhargava([51]). According to such critical values the null of non-stationarity cannot be rejected by an observed value of 0.14 for the DW statistic.

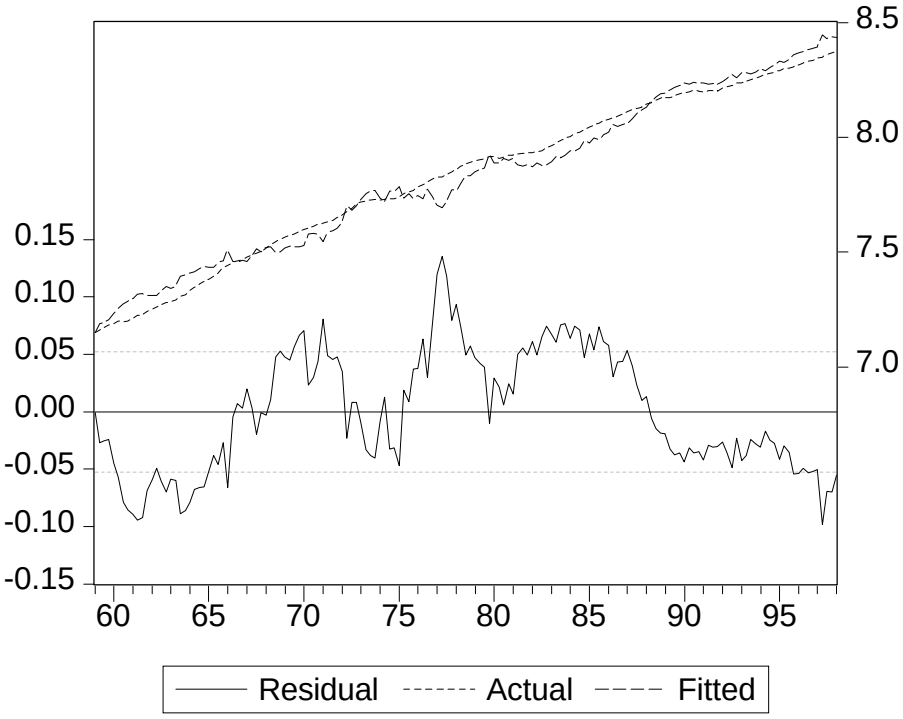

FIG. 2.6. A spurious regression

In conclusion we note that non-stationarity of time-series is problematic in that it might generate spurious regression and it does not allow the use of standard large-sample theory for valid estimation and inference in the linear model. Before considering the solutions to the problem we shall in the section clarify it further by re-illustrting it form a different perspective.

2.5.1 *Non-stationarity and the likelihood function.*

Consider a vector $\mathbf{x}_t$ containing observations an time series variables at time t. A sample of T time series observations on all the variables can be represented as follows:

$$\mathbf{X}_T^1 = \begin{bmatrix} \mathbf{x}_1 \\ . \\ . \\ . \\ \mathbf{x}_T \end{bmatrix}.$$

In general, estimation is performed by considering the joint sample density function, known also as the likelilihood function, which we can express as $\mathrm{D}\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right)$. The likelihood function is defined on the parameters space Θ, given the observation of the observed sample $\mathbf{X}_T^1$ and of a set of initial conditions $\mathbf{X}_0$. Such initial conditions can be interpreted as the pre-sample observations on the relevant variables (which are usually not available). In case of independent observations the likelihood function can be written as the product of the density functions for each observation. However this is not the relevant case for time-series, as time-series observations are in general sequentially correlated. In the case of time-series the sample density is then constructed using the concept of sequential conditioning. The likelihood function, conditioned with respect to initial conditions, can always be written as the product of a marginal density and a conditional density as follows:

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) D\left(\mathbf{X}_T^2 \mid \mathbf{X}_1, \boldsymbol{\theta}\right).$$

Obviously we also have

$$D\left(\mathbf{X}_T^2 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_2 \mid \mathbf{X}_1, \boldsymbol{\theta}\right) D\left(\mathbf{X}_T^3 \mid \mathbf{X}_2, \boldsymbol{\theta}\right)$$

and, by recursive substitution, we eventually obtain :

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = \prod_{t=1}^{T} D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right).$$

Having obtained $D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right)$ we can in theory derive $D\left(\mathbf{X}_T^1, \boldsymbol{\theta}\right)$ by integrating with respect to X_0 the density conditional on pre-sample observations. In practice this could be not tractable analitically as $D\left(X_0\right)$ is not known. The hypothesis of stationarity becomes crucial at this stage, as stationarity restricts the memory of time series and limits to the first observations in the sample the effects of pre-sample observations. This is the reason why, in the case of stationary processes, initial conditions can be simply ignored. Clearly the larger the sample, the better, as the weight of the information lost becomes smaller. Moreover note also that, even by omitting initial conditions we have:

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) \prod_{t=2}^{T} D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right).$$

Therefore the likelihood function is separated in the product on $T-1$ conditional distribution and one unconditional distribution. In the case of non-stationarity the unconditional distribution is not defined. On the other hand, in the case of stationarity the DGP is completely described by the conditional density function $D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right)$.

2.5.1.1 *An illustration: the first order autoregressive process* To give more empirical content to our case, let us consider again the case of the univariate first order autoregressive process.

$$x_t \mid \mathbf{X}_{t-1} \sim N\left(\lambda x_{t-1}, \sigma^2\right) \tag{2.7}$$

$$D\left(\mathbf{X}_T^1 \mid \lambda, \sigma^2\right) = D\left(x_1 \mid \lambda, \sigma^2\right) \prod_{t=2}^{T} D\left(x_t \mid \mathbf{X}_{t-1}, \lambda, \sigma^2\right). \tag{2.8}$$

From (2.8) it is clear that the likelihood function involves $T-1$ conditional densities and one unconditional densities. The conditional densities are given by (2.7) , the unconditional density can be derived only in the case of stationarity. In fact given :

$$\begin{aligned} x_t &= \lambda x_{t-1} + u_t \\ u_t &\sim N.I.D\left(0, \sigma^2\right), \end{aligned}$$

we can obtain by recursive substitution:

$$x_t = u_t + \lambda u_{t-1} + ... + \lambda^{n-1} u_1 + \lambda^n x_0.$$

And only if $|\lambda| < 1$, the effect of the initial condition disappear and we can write the unconditional density of x_t as:

$$D\left(x_t \mid \lambda, \sigma^2\right) = N\left(0, \frac{\sigma^2}{1-\lambda^2}\right).$$

There under stationarity we can write down the exact likelihood function as:

$$D\left(\mathbf{X}_T^1 \mid \lambda, \sigma^2\right) = (2\pi)^{-\frac{T}{2}} \sigma^{-T} \left(1-\lambda^2\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(\left(1-\lambda^2\right) x_1^2 + \sum_{t=2}^{T} \left(x_t - \lambda x_{t-1}\right)^2\right)\right] \tag{2.9}$$

and estimates of the parameters of interest are derived by maximizing this function. Note that $\widehat{\lambda}$ cannot be derived by analytical methods using the exact likelihood function,but it requires conditioning the likelihood and operating a grid search. Note also that the idea of using in large sample the approximate likelihood function by dropping the first observation works only under the hypothesis of stationarity in a large samples. When the first observation is dropped and the approximate likelihood fnction is considered, it can be shown analytically that the ML estimate of λ coincides with the OLS estimate.

2.6 Univariate decompositions of time-series

The general solution proposed to the problem introduced in the previous section is the search for a stationary representation of non-stationary time-series. This has been done both in an univariate and in a multivariate framework. As an introduction we shall briefly discuss methodologies used in a uni-variate framework to move swiftly to decompositions in a multivariate framework, which are at the heart of our discussion of modern macroeconometrics.

Beveridge-Nelson (1981) provide an elegant way of decomposing a non-stationary time-series into a permanent component and a temporary, cyclical,component by applying ARIMA methods. For any non-stationary time-series x_t integrated of the first order the Wold decomposition theorem could be applied to its first difference, to deliver the following representation:

$$\begin{aligned} \Delta x_t &= \mu + C(L)\,\epsilon_t \\ \epsilon_t &\sim n.i.d.\left(0, \sigma_\epsilon^2\right) \end{aligned}$$

where $C(L)$ is a polynomial of order q in the lag operator. Consider now the polynomial $D(L)$ defined as follows:

$$D(L) = C(L) - C(1) \tag{2.10}$$

given that $C(1)$ is a constant, also $D(L)$ will be of order q. It can immediately be seen that

$$D(1) = 0$$

therefore 1 is a root of $D(L)$, and we can write

$$D(L) = C^*(L)(1 - L) \tag{2.11}$$

where $C^*(L)$ is a polynomial of order $q - 1$.
By equating (2.10) to (2.11) we have:

$$C(L) = C^*(L)(1 - L) + C(1)$$

and

$$\Delta x_t = \mu + C^*(L)\,\Delta\epsilon_t + C(1)\,\epsilon_t \tag{2.12}$$

by integrating (2.12) we finally have:

$$\begin{aligned} x_t &= C^*(L)\,\epsilon_t + \mu t + C(1)\,z_t \\ &= C_t + TR_t \end{aligned}$$

where z_t is a process for which we have $\Delta z_t = \epsilon_t$. C_t is the cyclical component and TR_t is the trend component made of a deterministic trend and a stochastic trend. Note that the trend component can be represented as follows:

$$TR_t = TR_{t-1} + \mu + C(1)\,\epsilon_t.$$

2.6.1 *Beveridge-Nelson decomposition of an IMA(1,1) process*

Consider the process:

$$\Delta x_t = \epsilon_t + \theta\epsilon_{t-1}, \quad 0 < \theta < 1.$$

In this case we have:

$$C(L) = 1 + \theta L$$

$$C(1) = 1 + \theta$$

$$\begin{aligned} C^*(L) &= \frac{C(L) - C(1)}{1 - L} \\ &= -\theta. \end{aligned}$$

The BN decomposition gives the following result:

$$\begin{aligned} x_t &= C_t + TR_t \\ &= -\theta\epsilon_t + (1+\theta)\,z_t. \end{aligned}$$

2.6.2 *Beveridge-Nelson decomposition of an ARIMA(1,1) process*

Consider the process:

$$\Delta x_t = \rho\Delta x_{t-1} + \epsilon_t + \theta\epsilon_{t-1}$$

In this case we have

$$
\begin{aligned}
C(L) &= \frac{1+\theta L}{1-\rho L} \\
C(1) &= \frac{1+\theta}{1-\rho} \\
C^*(L) &= \frac{C(L)-C(1)}{1-L} \\
&= -\frac{\theta+\rho}{(1-\rho)(1-\rho L)}
\end{aligned}
$$

and the BN decomposition gives the following result:

$$
\begin{aligned}
x_t &= C_t + TR_t \\
&= -\frac{\theta+\rho}{(1-\rho)(1-\rho L)}\epsilon_t + \frac{1+\theta}{1-\rho} z_t
\end{aligned}
$$

2.6.3 *Deriving the Beveridge-Nelson decomposition in practice*

The practical derivation of a BN decomposition for any ARIMA process is easily derived by applying a methodology suggested by Cuddington and Winters([6]). For any I(1) process, we have seen that the stochastic trend can be represented as follows:

$$
TR_t = TR_{t-1} + \mu + C(1)\,\epsilon_t \tag{2.13}
$$

The decomposition can then be applied by the following steps:

- identify the appropriate ARIMA model and estimate ϵ_t and all the parameters in μ and $C(1)$ and
- given an initial values for TR_0 use (2.13) to generate the permanent component of the time-series
- generate the cyclical component as the difference between the observed value in each period and the permanent component

The above procedure will give the permanent component up to constant, if the precision of this procedure isnot satisfactory, one can use further conditions to identify more precisely the decomposition. For example one can impose the condition that the sample mean of the cyclical component is zero to pin down the constant in the permanent component.

To illustrate how the procedure works in practice we have simulated an ARIMA(1,1,1) in E-Views for a sample of 200 observations, by running the following programme:

```
smpl 1 2
```

```
genr x=0
smpl 1 200
genr u=nrnd
smpl 3 200
series x= x(-1) +0.6*x(-1)-0.6*x(-2) +u+0.5*u(-1)
```

From the previous section we know the exact BN decomposition of our x_t :

$$\begin{aligned} x_t &= C_t + TR_t \\ &= -\frac{1.1}{(1-0.6)(1-0.6L)}\epsilon_t + \frac{1.5}{0.4}z_t \\ TR_t &= TR_{t-1} + \frac{1.5}{0.4}\epsilon_t \end{aligned}$$

we can therefore generate the permanent component of X and the transitory component as follows:

```
smpl 1 2
genr p=0
smpl 3 200
series TR= TR(-1)+(1.5/0.4)*u
genr CYCLE=X-TR
```

The series X, TR and CYCLE are reported in Figure 2.7.

This is exactly the procedure that we follow in practice except that we estimate parameters rather than impute them from the known DGP.

2.6.4 *Assessing the Beveridge-Nelson decomposition*

The properties of the permanent and temporary component of an integrated time-series delivered by the BN decomposition are worth some comments. The innovations in the permanent and the transitory components are perfectly negatively correlated, moreover the trend component is more volatile than the actual time series as the negative correlation between the permanent and the transitory component acts to smooth the original time-series. These results are easily seen for the simplest case we have already discussed. For example in the case of the IMA(1,1) process the correlation between the innovations in the permanent and transitory component is -$E_t(1.5\epsilon_t 0.5\epsilon_t) = 1$, the variance of the innovation in trend component is $(1.5)^2\sigma_\epsilon^2 > \sigma_\epsilon^2$. Note that in general the variance of innovation might have economic interpretation and economic theory might suggests different pattern of correlations between innovations from a perfect negative correlation. As we shall see in one of the next chapters, an interesting pattern could be the absence of correlation between the innovation in the cycle and trend component of an integrated time-series. In general, different restrictions on the correlation between the trend and cycle components lead to the identification of different stochastic trends for integrated time-series. As a consequence the

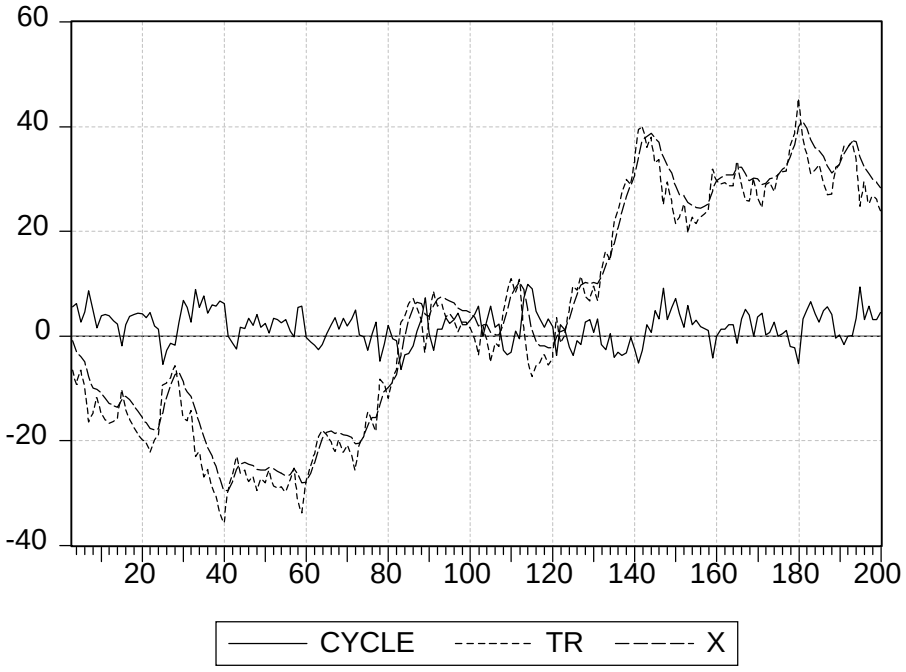

FIG. 2.7. A Beveridge-Nelson decomposition of an ARIMA(1,1,1) process

Beveridge-Nelson decomposition is not unique. In general uni-variate decompositions are not unique. To see this point more explicitly we can compare the BN trend with the trend extracted using an alternative technique which has been recently very successful in time-series analysis: The Hodrick-Prescott filter.

Hodrick and Prescott proposed their method to analyze postwar U.S. business cycles in a working paper circulated in the early 1980s and published in 1997([27]). The Hodrick-Prescott (HP) filter computes the permanent component TR_t of a series x_t by minimizing the variance of x_t around TR_t, subject to a penalty that constrains the second difference of TR_t . That is, the HP filter is derived by minimizing the following expression:

$$\sum_{t=1}^{T} (x_t - TR_t)^2 + \lambda \sum_{t=2}^{T-1} \left[(TR_{t+1} - TR_t)^2 - (TR_t - TR_{t-1})^2 \right].$$

The penalty parameter λ controls the smoothness of the series, by controlling the ratio of the variance of the cyclical component to the variance of the series. The larger the λ, the smoother the TR_t approaches a linear trend as λ goes

to infinite. In practical applications λ is set to 100 for annual data, 1600 for quarterly data and 14400 for monthly data.

In the following Figure we report the BN trend and the HP trend (with $\lambda = 100$) for the data generated in the previous section.

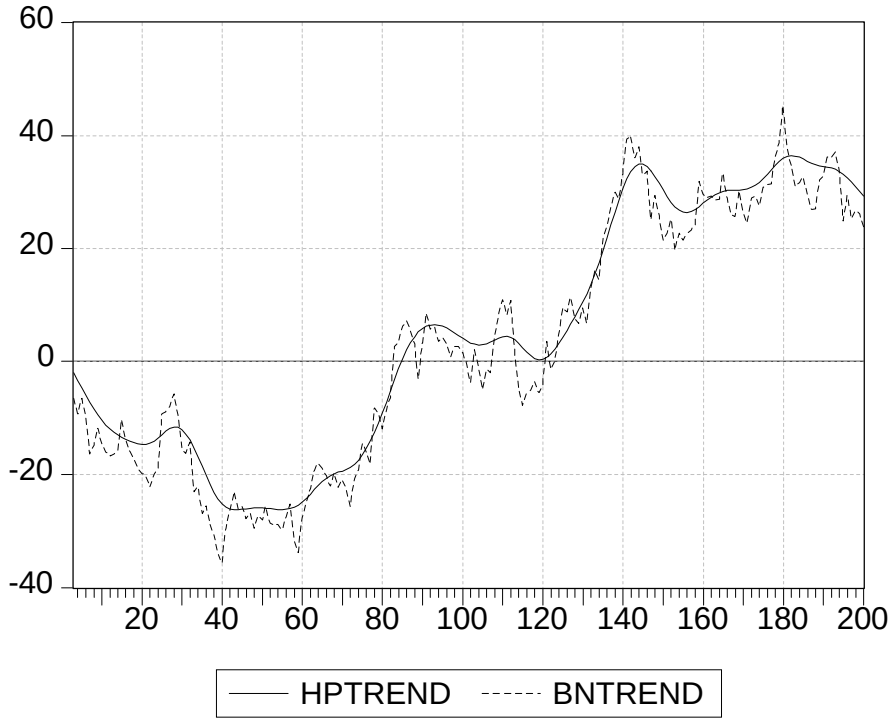

FIG. 2.8. Trend components: Hodrick-Prescott versus Beveridge-Nelson

Note that the BN trend is more volatile than the HP trend. It is possible to increase the volatility of the HP trend by reducing the parameter λ, however the HP filter can reach at most the volatility of the actual time series which, as we already know, is smaller than the volatility of the BN trend.

The HP filter has the advantage of removing the same trend from all time series; this might be desirable as some theoretical models, as for example real business cycle models, indicate that macroeconomic variables share the same stochastic trend. However, it has been shown by Harvey and Jaeger([22])that the use of such filter can lead to the identification of spurious cyclical behaviour. In fact the two authors above predicate a different approach to modelling time-series, known as structural time series modelling, which, we do not consider in our analysis as it is not related to macroeconomic models, but certainly merits

some attention.[2]

The comparison between the HP and the BN trend reinforces the argument of non-uniqueness of univariate decomposition made before, moreover we are left with the problem of how to use the filtered series in applied macroeconometrics and how to relate them to theoretical models. The empirical counterparts of theoretical macroeconomic models are multivariate time-series. Theoretical models often predict that different time-series share the same stochastic trend. The natural question at this point is if it is possible that the problem of non-stationarity in time-series could be resolved by considering multivariate models. In this context, stationarity is obtained by considering combination of non-stationary time series sharing the same stochastic trends. If such results could be achieved, it would be in principle possible to justify the identification of trends by relating them to macroeconomic theory. We shall consider this possibility in the next sections.

2.7 Multivariate decompositions and dynamic models

Let us reconsider our spurious regression for US consumption in the context of a dynamic model. We do so by augmenting the static regression to consider consumption and income lagged up to one year, i.e. we consider four lags of each variables. Results shown over Table 4, witness that the spurious regression result disappears: i.e contemporaneous and lagged US disposable income is significant in explaining US consumption, while contemporaneous and lagged UK disposable income is not.

[2]We refer the interested reader to the work by Andrew Harvey and Augustin Maravall.

TABLE 4: A dynamic model for US consumption

	Dependent variable $LCUS_t$, regression by OLS, 1960:1-1998:1			
	Model with US income		*Model with UK income*	
	Coefficient	S.E.	Coefficient	S.E
c	0.367	0.106	0.333	0.150
$LCUS_{t-1}$	0.987	0.087	1.197	0.083
$LCUS_{t-2}$	-0.006	0.120	-0.156	0.131
$LCUS_{t-3}$	0.012	0.121	0.142	0.130
$LCUS_{t-4}$	-0.172	0.085	-0.196	0.082
LYUS	0.258	0.037		
$LYUS_{t-1}$	-0.126	0.049		
$LYUS_{t-2}$	-0.068	0.050		
$LYUS_{t-3}$	0.021	0.049		
$LYUS_{t-4}$	0.034	0.042		
LYUK			0.009	0.020
$LYUK_{t-1}$			0.018	0.028
$LYUK_{t-2}$			-0.034	0.028
$LYUK_{t-3}$			-0.0163	0.028
$LYUK_{t-4}$			0.0015	0.0229
Trend	0.00039	0.0001	0.00023	0.0001
R^2	0.99		0.99	
S.E.	0.0037		0.0042	
F-test on income	F(5,155)=10.324		F(5,155)=1.239	

.

This is an interesting result which leads to think that, in the case also the problems related to non-stationarity could be solved, dynamic multivariate time-series models are the right foundation for macroeconometrics.

2.7.1 *Cointegration and Error Correction Models*

To explain why the spurious results disappear when dynamic models are estimated let us consider a simplified version of the dynamic specification estimated for consumption:

$$c_t = a_0 + a_1 c_{t-1} + a_2 y_t + a_3 y_{t-1} + u_t \tag{2.14}$$

This specification has some interesting dynamic properties which are worth discussing. First note that the short-run elasticity of consumption with respect to income is different from the long-run elasticity. In fact the short-run elasticity is a_2 while the long-run elasticity is $\frac{a_2+a_3}{1-a_1}$. The long-run elasticity is found by setting all variables in the dynamic model (2.14)to their steady state value $c_{t+i} = \overline{c}$, $y_{t+i} = \overline{c}$. To see immediately this point consider the following re-parameterisation of (2.14):

$$\Delta c_t = a_0 + a_2 \Delta y_t - \alpha \left(c_{t-1} - \beta_1 y_{t-1} \right) + u_t \tag{2.15}$$

$$\alpha = (1 - a_1), \ \beta_1 = \frac{a_2 + a_3}{1 - a_1} \tag{2.16}$$

The estimated dynamic model includes both first differences and levels. The presence of the level variables generates a long-run solution, derived by setting all first differences either to zero (steady-state with no deterministic trend) or to a constant(steady-state). Note now the role of the terms in level: we can interpret $\beta_1 y_{t-1}$ as the long-run equilibrium level c^* for the log of real consumption c. When $\alpha < 0$ consumption increases at time t whenever $c_{t-1} < c^*_{t-1}$, and decreases whenever $c_{t-1} > c^*_{t-1}$.The system equilibrates in presence of disequilibrium (i.e. a discrepancy between c and c^*) such error correction features guarantees that in the long-run the consumption will converge to its equilibrium value. For this reason the specification (2.15) ,with $\alpha < 0$, is termed Error Correction Model. Note that, in the case of an ECM representation, the difference between c and c^* is a stationary series. This defines co-integration. We say that two non stationary variables integrated of order q are cointegrated of order p if there exist a linear combination of them which is integrated of order $p - q$. The case $p = 1$, $q = 1$, is interesting in that co-integration implies an ECM representation, which allows to re-write a model in levels, which involves non-stationary time-series, as a model which involves only stationary variables. Such variables are stationary either because they are first diferences of non-stationary variables or because they are stationary linear combination of non-stationary variables (cointegrating vectors).

The inclusion of both differences and levels in the estimated relationship is the key factor to the solution of the problems related to non-stationarity of the level of variables included in the specification. This solution to the non-stationarity problem has also the feature of revealing immediately to the economist the long-run properties of the estimated model. To see this point practically we can use E-Views to simulate the following bivariate model:

$$\begin{aligned} \Delta c_t &= 0.25 \Delta y_t - 0.2 \left(c_{t-1} - y_{t-1} \right) + 0.003 u_{1t} \\ \Delta y_t &= 0.02 + 0.009 u_{2t} \end{aligned} \tag{2.17}$$

where u_{1t} and u_{2t} are independently distributed standard normal, the parameters are calibrated to reflect the long-run properties of the consumption function reported in Table 4. The volatility of the innovations are again calibrated to estimated processes on real data for the US economy; income is more volatile than consumption.

To show the properties of the model, we first generate samples for the two innovation process, then we generate artificial data for consumption and income by constructing the above model and solving it dynamically. We do so for a

sample of 100 observations, the simulated series are plotted in Figures 2.9 and 2.10.

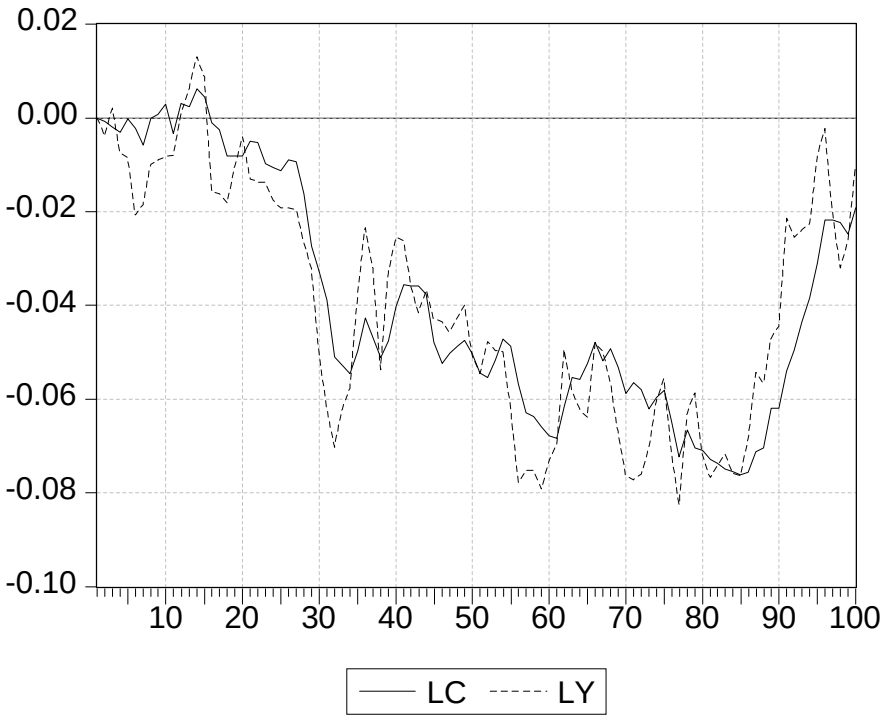

FIG. 2.9. Two cointegrated series

Note that the the levels of LC and LY, share a stochastic trend, which disappears from the series (LC-LY). The parameter α in the ECM spefication determines the speed of adjustment in presence of disequilibrium. To illustrate the role of this parameter we report the two series (LC-LY) generated by taking the same innovations for the sample 1 200. The innovations are drawn for normal independent for all observations, except for observation 101 where the residuals in the income process are augmented by 0.036. We have then a shock four standard deviation away from the mean of the distribution, we can then visually inspect the behaviour of the simulated series in the presence of an outlier. The process (2.17) is used to generate the first time-series of disequilibria, while the second time-series is generated using keeping all parameters unchanged with the exception of α, which is trebled to 0.6 from 0.2. The resulting observations for disequilibria are reported in the Figure 2.11.

The disequilibria from the second simulation run are less persitent to witness

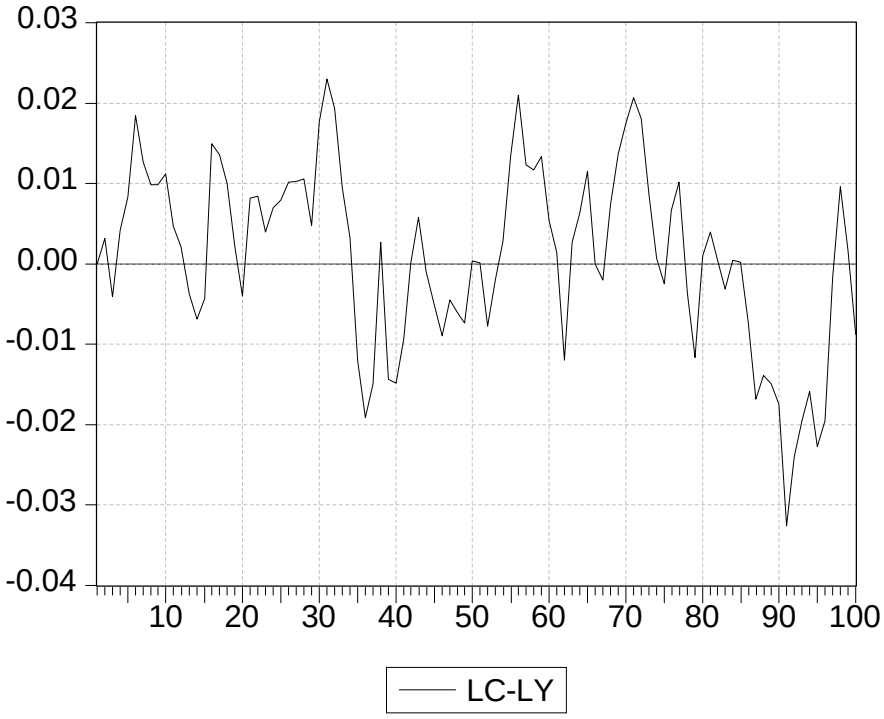

FIG. 2.10. Disequilibrium

that the second system feature a fastest speed of adjustment in presence of disequilibrium. All the simulated series are contained in the E-Views workfile ECM.WF1, with which the reader can experiment to convince herself of the properties of Error Correction Models.

As an application of further interest let us reconsider the static regression in the light of our discussion of dynamic models.

Given the following DGP:

$$\begin{aligned} y_t &= a_1 y_{t-1} + a_2 x_t + a_3 x_{t-1} + u_{1t} \\ x_t &= b_1 x_{t-1} + u_{2t} \\ \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} &\sim N.I.D.\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} \right) \end{aligned} \tag{2.18}$$

a static model is estimated by OLS:

$$y_t = \gamma x_t + \varepsilon_t$$

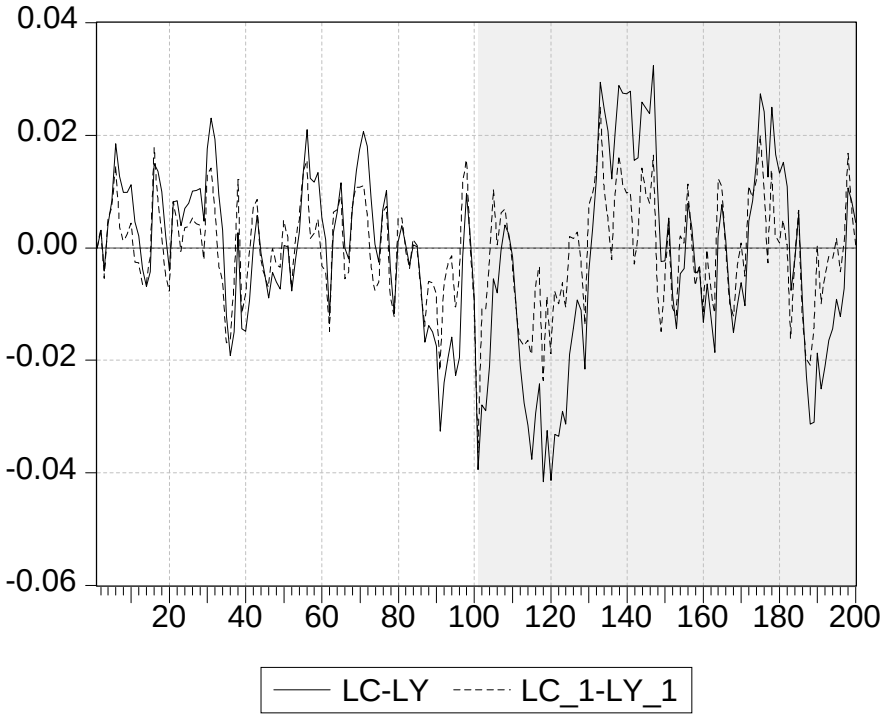

FIG. 2.11. Speed of adjustments and disequilibrium

$$\widehat{\gamma} = \frac{\sum x_t y_t}{\sum x_t^2}$$

Assess the results of running the static model by taking $p \lim \hat{\gamma}$:

$$p \lim \widehat{\gamma} = p \lim \left[a_1 \frac{\sum x_t y_{t-1}/T}{\sum x_t^2/T} + a_2 + a_3 \frac{\sum x_t x_{t-1}/T}{\sum x_t^2/T} + \frac{\sum x_t u_t/T}{\sum x_t^2/T} \right]$$

under the hypothesis that (2.18) is stationary ($b_1 < 1$) we can substitute for x_t in terms of x_{t-1} and u_{2t} and apply Slutsky's and Cramer's theorem to derive the following result:

$$p \lim \widehat{\gamma} = \frac{a_2 + a_3 b_1}{1 - a_1 b_1}$$

$$a_2 \leq p \lim \widehat{\gamma} \leq \frac{a_2 + a_3}{1 - a_1}$$

Note that as b_1 approaches 0 the elasticy of y with respect to x delivered by the static regression goes asymptotically to the true short-run elasticity, while as b_1 approaches to 1 such elasticity converges to the long-run elasticity. Technically speaking we cannot show what happens when b_1 is one because this violates the stationarity conditions which we have used to derive the asymptotic behaviour of the OLS estimator. However, confirming the above intuition, Stock([55]) has shown that the OLS estimator of the parameters determining the long-run relationship non-stationary cointegrated series is super-consistent. In fact it converges towards the true value at speed $\left(\frac{1}{T}\right)$,higher than the speed of much $\left(\frac{1}{\sqrt{T}}\right)$,with which the OLS estimator converges to its true value in regression between stationary time series. This result has given some background to a two-step research strategy, according to which the cointegrating relations is estimated first in static model and the used to estimate a dynamic ECM model, involving only stationary variables. This strategy is less efficient than the simultaneous estimation of short-run and long-run dynamics. In fact the static regression delivers super-consitent estimates of the cointegrating parameter despite being mis-specified, because the omitted variables are the stationary variables determining the short-run dynamics, which, in large-samples, should not affect the estimation of cointegrating parameters. It has been shown through Monte-Carlo simulation that the dimension of the samples required to appeal to the super-consitency theorem are much higher than the dimension of the samples usually available for time-series modelling (see, for example, [2],[3], [4]). Moreover the empirical counterpart of macroeconomic models are usually dynamic multivariate time-series models. Therefore, there must be a price to be paid in considering static uni-variate models as a basis for empirical work. We shall devote more attention to this issue in the next section.

2.7.2 *Cointegration in a multivariate framework*

So far we have stressed the importance of the magnitude of the adjustment parameter α as the relevant discriminant to decide on cointegration, but we have not yet provided a statistical framework to test for such an hypothesis.We also mentioned the importance of dimensionality of the system to be considered in empirical work. In this section we shall elaborate on these points and illustrate the Johansen's([30], [34]) approach to cointegration in a multivariate framework.

So far we have considered cointegration in a bi-variate context. Things differ when we consider a multivariate context. In fact , in general between n non-stationary series we can have up to $n-1$ cointegrating vector and the single equation dynamic modelling can cause serious troubles when there are multiple cointegrating vectors. To illustrate the problem let us consider the case of an investigator who uses cointegration techniques to investigate money demand. Following the standard economic background to empirical investigations of money demand (see, for example, Hendry and Ericsson,[26]) the chosen data set includes money, m, a price index, p, real income, y, the own interest rate on money, R^m, and the opportunity cost of holding money, R^b. All variables are in logarithms,

with the exception of interest rates. The investigator specifies a dynamic single-equation model for real money, towards the identification of a money demand equation, which takes the following shape:

$$\begin{aligned}(m-p)_t = a_0 + a_1 (m-p)_{t-1} + a_2 y_{t-1} + a_3 y_{t-2} + \\ + a_4 R^m_{t-1} + a_5 R^m_{t-2} + a_6 R^b_{t-1} + a_7 R^b_{t-2} + u_t\end{aligned} \tag{2.19}$$

This statistical model fits the data well. As it is found that $a_1 < 1$, the investigation leads to the identification of a long-run equilibrium money demand, which results clearly form the ECM re-parametrization of the dynamic model (2.19) :

$$\begin{aligned}\Delta (m-p)_t &= a_0 - a_3 \Delta y_{t-1} - a_5 \Delta R^m_{t-1} - a_6 \Delta R^b_{t-1} + \\ &\quad (a_1 - 1)\left[(m-p)_{t-1} - (m-p)^*_{t-1}\right] + u_t \\ (m-p)^*_{t-1} &= \frac{a_2 + a_3}{1-a_1} y_{t-1} + \frac{a_4 + a_5}{1-a_1} R^m_{t-1} + \frac{a_6 + a_7}{1-a_1} R^b_{t-1}\end{aligned} \tag{2.20}$$

However, the good fit of the statistical model might be combined with an incorrect identification of the long-run solution. Think, for example, of the case in which the non-stationary vector containing the five variables of interests admits two cointegration relationships:$(m-p-y)$ and $\left(R^m - \beta_{22} R^b\right)$. Where the first one is generated by the stationarity of the velocity of circulation of money and the second one by the behaviour of the banking sector, which sets the interest rate on money as a mark-down on the opportunity cost of holding money. In the short-run money reacts to disequilibria with respect to both long-run solutions, hence money demand is correctly parametrised as follows:

$$\begin{aligned}\Delta (m-p)_t = \pi_0 + \pi_1 \Delta y_{t-1} + \pi_2 \Delta R^m_{t-1} + \pi_3 \Delta R^b_{t-1} \\ - \alpha_1 (m_{t-1} - p_{t-1} - y_{t-1}) + \alpha_2 \left(R^m_{t-1} - \beta_{22} R^b_{t-1}\right) + u_t\end{aligned} \tag{2.21}$$

Note that the statistical specification of (2.15) and (2.21) is identical, in fact the residuals u_t are the same, however identification is very different. In fact when (2.21) represents the correct specification, (2.15) identifies as long-run elasticities what in fact are mixtures of cointegrating parameters and parameters determining the speed of adjustment with respect to disequilibria in the true model. Single-equation approach leads to believe that the long-run elasticity of money demand with respect to the opportunity cost of holding money is $\frac{a_6+a_7}{1-a_1}$,while in fact such estimated coefficient is a convolution of the parameter α_2, determining the speed with which money demand reacts to a disalignment of interest rates with respect to their equilibrium value, and the parameter c, determining the mark-down of the own interest rate on money with respect to the

interest rate on the opportunity cost of holding money. This identification has serious consequences in the intepretation of estimated parameters. In fact when the above problem occurs a structural instability in the short-term adjustment parameter β_{22} would mis-lead the researcher into the belief that long-run money demand is unstable.

The solution of this identification problem requires a framework to allow the researcher to find the number of cointegrating vectors among a set of variables and to identify them. The procedure proposed by Johansen([30], [32]) within the framework of the Vector Autoregressive Model achieves both results.

2.7.3 *The Johansen Procedure*

To illustrate the procedure proposed by Johansen consider the multivariate generalisation of the single-equation dynamic model discussed so far, i.e. a Vector Autoregressive Model (VAR) for the vector of, possibly non-stationary, m-variables $\mathbf{y}$:

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + ... + \mathbf{A}_n\mathbf{y}_{t-n} + \mathbf{u}_t \tag{2.22}$$

by proceeding in the same way we did for the simple single-equation dynamic model, we can reparameterise the VAR in levels as a model involving levels and first-differences of variables.

Start by subtracting $\mathbf{y}_{t-1}$ from both sides of the VAR to obtain:

$$\Delta\mathbf{y}_t = (\mathbf{A}_1 - \mathbf{I})\,\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + ... + \mathbf{A}_n\mathbf{y}_{t-n} + \mathbf{u}_t \tag{2.23}$$

now subtract $(\mathbf{A}_1 - \mathbf{I})\,\mathbf{y}_{t-2}$ from both sides to obtain:

$$\Delta\mathbf{y}_t = (\mathbf{A}_1 - \mathbf{I})\,\Delta\mathbf{y}_{t-1} + (\mathbf{A}_1 + \mathbf{A}_2 - \mathbf{I})\,\mathbf{y}_{t-2} + ... + \mathbf{A}_n\mathbf{y}_{t-n} + \mathbf{u}_t \tag{2.24}$$

By iterating this procedure until n-1, we end up with the following specification:

$$\Delta\mathbf{y}_t = \Pi_1\Delta\mathbf{y}_{t-1} + \Pi_1\Delta\mathbf{y}_{t-2} + ... + \Pi\mathbf{y}_{t-n} + \mathbf{u}_t \tag{2.25}$$

$$= \sum_{i=1}^{n-1}\Pi_i\Delta\mathbf{y}_{t-i} + \Pi\mathbf{y}_{t-n} + \mathbf{u}_t \tag{2.26}$$

where :

$$\Pi_i = -\left(I - \sum_{j=1}^{i}\mathbf{A}_j\right)$$

$$\Pi = -\left(I - \sum_{i=1}^{n}\mathbf{A}_i\right)$$

Clearly the long-run properties of the system are described by the properties of the matrix Π. There are three cases of interest:

- rank (Π) = 0. The system is non-stationary, with no cointegration between the variables considered. This is the only case in which non-stationarity is correctly removed just by taking first difference of the variables considered
- rank (Π) = m, $full$. The system is stationary.
- rank (Π) = $k < m$. The system is non stationary but there are k cointegrating relationships among the considered variables. In this case we have $\Pi = \alpha\beta'$, where α is an $(m \times k)$ matrix of weights and β is an $(m \times k)$ matrix of parameters determining the cointegrating relationships.

Therefore, the rank of Π is crucial in determining the number of cointegrating vectors. The Johansen procedure is based on the fact that the rank of a matrix is equal to the number of its characteristic roots that differ from zero. Here is the intuition on how the tests can be constructed. Having obtained estimates for the parameters in the Π matrix, we associate to them estimates for the m characteristic roots and we order them as follows $\lambda_1 > \lambda_2 > ... > \lambda_m$. If the variables are not cointegrated, then the rank of Π is zero and all the characteristic roots will be zero. In this case each of the expression $\ln(1 - \lambda_i)$ will be zero. If instead the rank of Π is one, and $0 < \lambda_1 < 1$, then $\ln(1 - \lambda_1)$ will be negative and $\ln(1 - \lambda_2)$ $=\ln(1 - \lambda_3) = ... = \ln(1 - \lambda_m) = 0$. Johansen derives a test on the number of characteristic roots that are different from zero by considering the two following statistics:

$$\lambda_{trace}(k) = -T \sum_{i=k+1}^{m} \ln\left(1 - \widehat{\lambda}_i\right)$$

$$\lambda_{\max}(k, k+1) = -T \ln\left(1 - \widehat{\lambda}_{k+1}\right)$$

where T is the number of observations used to estimate the VAR. The first statistic test the null of at most k cointegrating vectors against a generic alternative. The test should be run in sequence starting from the null of at most zero cointegrating vectors up to the case of at most m cointegrating vectors. The second statistic tests the null of at most k cointegrating vectors against the alternative of at most $k + 1$ coitegrating vectors. Both statistics will be small under the null hypothesis. Critical values are tabulated by Johansen and they depend on the number of non-stationary component under the null and on the specification of the deterministic component of the VAR. Johansen has shown in the past ([33]) some preference for the trace test on the argument that the maximum eigenvalue test does not give rise to a coherent testing strategy.

To illustrate briefly the intuition behind the procedure, consider the VAR representation of our simple dynamic model (2.18) , introduced in one of the previous sections, for the two variables x and y:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{2.27}$$

(2.27) can be reparameterised as follows in terms of the VECM representation:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} a_{11} - 1 & a_{12} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{2.28}$$

from which it is clear that

$$\Pi = \begin{pmatrix} a_{11} - 1 & a_{12} \\ 0 & 0 \end{pmatrix}, \alpha = \begin{pmatrix} a_{11} - 1 \\ 0 \end{pmatrix}, \beta' = \begin{pmatrix} 1 & -\frac{a_{12}}{1-a_{11}} \end{pmatrix}$$

To expand on this intuition let us reconsider our example on money demand from the previous section.

The baseline VAR could be specified as follows:

$$\begin{bmatrix} (m-p)_t \\ y_t \\ R_t^m \\ R_t^b \end{bmatrix} = A_0 + A_1 \begin{bmatrix} (m-p)_{t-1} \\ y_{t-1} \\ R_{t-1}^m \\ R_{t-1}^b \end{bmatrix} + A_2 \begin{bmatrix} (m-p)_{t-2} \\ y_{t-2} \\ R_{t-2}^m \\ R_{t-2}^b \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix}$$

Which could then be reparameterised in VECM form:

$$\begin{bmatrix} \Delta (m-p)_t \\ \Delta y_t \\ \Delta R_t^m \\ \Delta R_t^b \end{bmatrix} = \Pi_0 + \Pi \begin{bmatrix} (m-p)_{t-1} \\ y_{t-1} \\ R_{t-1}^m \\ R_{t-1}^b \end{bmatrix} + \Pi_1 \begin{bmatrix} \Delta (m-p)_{t-1} \\ \Delta y_{t-1} \\ \Delta R_{t-1}^m \\ \Delta R_{t-1}^b \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix}$$

Given that we know that there are two cointegrating vectors, we have:

$$\begin{aligned} \Pi &= \alpha\beta' \\ rank\ \Pi &= 2 \\ \beta' &= \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -\beta_{22} \end{bmatrix} \end{aligned}$$

As we have analysed only one equation in our previous discussion of the system, the only constraints we have on the specification for α are $\alpha_{11} < 0, \alpha_{12} > 0$. A possible specification for α would then be:

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \\ 0 & \alpha_{32} \\ 0 & 0 \end{bmatrix}$$

with the above specification for the loadings, money demand adjusts both in presence of misalignements of velocity with respect to the equilibrium velocity and of misalignments of interest rates with respect to their equilibrium spread. In particular money demand increases when velocity is "too high" and the opportunity cost of holding money is "too low". In case of disequilibrium in interest rates it is the interest on money which adjusts, while the dynamics of interest rates on the alternative of money in agents'portfolio does not react to disequilibria.

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \\ 0 & \alpha_{32} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -\beta_{22} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & -\alpha_{11} & \alpha_{12} & -\alpha_{12}\beta_{22} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_{32} & -\alpha_{32}\beta_{22} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

2.7.4 *Identification of multiple cointegrating vectors*

The Johansen procedure allows to identify the number of cointegrating vectors. However, in the case of existence of multiple cointegrating vectors, an interesting identification problem arises. In fact, α and β, are only determined up to the space spanned by them and, for any non-singular matrix ξ conformable by product, we have:

$$\Pi = \alpha\beta' = \alpha\xi^{-1}\xi\beta'$$

In other words β and $\beta'\xi$ are two observationally equivalent basis of the cointegrating space. The obvious implication is that, before solving such identification problem, no meaningful economic interpretation of coefficients in cointegrating vectors can be proposed. The solution to such problem is achieved by imposing a sufficient number of restrictions on parameters such that the matrix satisfying such constraints in the cointegrating space is unique. Such criterion is derived in Johansen(1992) and discussed in the work of Johansen-Juselius, Giannini([15]) and Hamilton([20]). Given the matrix of cointegrating vectors $\boldsymbol{\beta}$ we can formulate linear constraints on the different cointegrating vectors using the R_i matrices of dimensions $r_i \times n$. Let us consider the columns of $\boldsymbol{\beta}$, i.e. the parameters in each cointegrating vectors, ignoring the normalisation constraint to1 of one variable in each cointegrating vector. Any structure of linear contraints can be represented as follows:

$$\mathbf{R}_i\boldsymbol{\beta}_i = 0$$

$R_i\,(r_i \times n)$, $\boldsymbol{\beta}_i(n \times 1)$,rank $R_i = r_i$.

The same constraints can be expressed in explicit forms as follows:

$$\boldsymbol{\beta}_i = \mathbf{S}_i \theta_i$$

$S_i\,(n \times (n - r_i))$, $\boldsymbol{\beta}_i(n \times 1), \theta_i\,((n - r_i) \times 1)$, rank $S_i = n - r_i, \mathbf{R}_i \mathbf{S}_i = 0$.

A necessary and sufficient condition for identification of parameters in the $i - th$ cointegrating vectors is the following:

$$rank\,(\mathbf{R}_i \boldsymbol{\beta}) = r - 1 \tag{2.29}$$

when (2.29) is satisfied it is not possible to replicate the cointegrating vector $i - th$ by taking linear combinations of the parameters in the other cointegrating vectors. In this case the matrix obtained by applying to the cointegrating space the restrictions of the $i - th$ cointegrating vectors has rank $r - 1$.

A necessary condition for identification is immediately derived in that $\mathbf{R}_i \boldsymbol{\beta}$ must have enough rows to satisfy condition (2.29) , therefore a necessary condition for identification is that each cointegrating vectors has at least $r - 1$ restrictions.

A sufficient condition for identification is provided by Johansen by considering the implicit and explicit form of expressing constraints:

Theorem 2.1 *The i-th cointegrating vector is identified by the constraints* $\mathbf{S}_1, \mathbf{S}_2, ...\mathbf{S}_r$ *if for each k=1,...,r-1 and for each set of indices* $1 < j_1 < ... < j_k < r$, *not containing i, we have that : rank* $[R_i S_{j_1}, ...R_i S_{j_k}] > k$

Given identification of the system we can distinguish the case of just-identification and over-identification. In case of over-identification, the over-identifying restrictions are testable.

2.7.4.1 *An illustrative example* Let us reconsider our example on money demand.Considering the following vectorial representation of the series $\left(m - p\ y\ R^m\ R^b \right)'$, and leaving aside normalizations, the matrix $\boldsymbol{\beta}$ can be represented as follows:

$$\begin{pmatrix} \beta_{11} & 0 \\ -\beta_{11} & 0 \\ 0 & \beta_{32} \\ 0 & -\beta_{42} \end{pmatrix}$$

given the following general representation of the matrix $\boldsymbol{\beta}$:

$$\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \\ \beta_{41} & \beta_{42} \end{pmatrix}$$

our constraints imply the following specification for the matrices R_i ans S_i :

$$R_1 = \begin{pmatrix} 1\ 1\ 0\ 0 \\ 0\ 0\ 1\ 0 \\ 0\ 0\ 0\ 1 \end{pmatrix}, S_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$$

$$R_2 = \begin{pmatrix} 1\ 0\ 0\ 0 \\ 0\ 1\ 0\ 0 \end{pmatrix}, S_2 = \begin{pmatrix} 0\ 0 \\ 0\ 0 \\ 1\ 0 \\ 0\ 1 \end{pmatrix}$$

The necessary conditions for identification are obviously satisfied, while the sufficient conditions for identification requires: $\text{rank}(R_1 S_2) \geq 1, \text{rank}(R_2 S_1) \geq 1$. They are also satisfied, in fact:

$$R_1 S_2 = \begin{pmatrix} 0\ 0 \\ 1\ 0 \\ 0\ 1 \end{pmatrix}, R_2 S_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

2.7.5 *Hypothesis testing with multiple cointegrating vectors*

The Johansen procedure allows for testing the validity of restricted forms of cointegrating vectors. More precisely, the validity of restrictions in additions(over-identifying restrictions) to those necessary to identify the long-run equilibria could be tested. The intuition behind the construction of all tests is that when there are r cointegrating vectors only these r linear combination of variables are stationary, therefore the test statistics involve comparing the number of cointegrating vectors under the null and the alternative hypothesis. Following this intuition, we understand immediately why only over-identifying restrictions can be tested, in fact just-identified model feature the same long-run matrix Π, and therefore the same eigenvalues of Π. Consider the case of testing restrictions on a set r of identified cointegrating vectors stacked in the matrix $\boldsymbol{\beta}$. The test statistic involves comparing the number of cointegrating vectors under the null and the alternative hypothesis. Let $\widehat{\lambda}_1, \widehat{\lambda}_2, ..., \widehat{\lambda}_r$, the ordered eigenvalues of the Π matrix in the unrestricted model, and $\widehat{\lambda}_1^*, \widehat{\lambda}_2^*, ..., \widehat{\lambda}_r^*$the ordered eigenvalues of Π matrix in the restricted model, restricitions on $\boldsymbol{\beta}$ are testable by forming the following test statistic:

$$T \sum_{i=1}^{r} \left[\ln\left(1 - \widehat{\lambda}_i^*\right) - \ln\left(1 - \widehat{\lambda}_i\right) \right] \tag{2.30}$$

Johansen ([32]) shows that the statistic (2.30) takes a χ^2 distribution with degrees of freedom equal to the number of over-identifying restrictions. Note that

small values of $\widehat{\lambda}_i^*$ with respect to $\widehat{\lambda}_i$ imply a reduction of rank of Π when the restrictions are imposed and hence the rejection of the null hypothesis. This testing procedure can be extended to tests on restrictions on the matrix of weigths $\boldsymbol{\alpha}$ or on the deterministic components (constant and trends) of the cointegrating vectors.

2.7.6 *Cointegration and Common Stochastic Trends*

Having discussed the VECM representation for a vector of m non-stationary variables admitting k cointegrating relationships, it is interesting to compare it with the multivariate extension of the Beveridge-Nelson decomposition. Consider the simple case of an I(1) vector $\mathbf{y}_t$ featuring first order dynamics and no deterministic components:

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{y}_{t-1} + \mathbf{u}_t \tag{2.31}$$

where α is the $(m \times k)$ matrix of loadings and $\boldsymbol{\beta}$ is the $(m \times k)$ matrix of parameters in the cointegrating relationships. When As $\mathbf{y}_t$ is I(1), we can apply the Wold decomposition theorem to $\Delta \mathbf{y}_t$ to obtain the following representation:

$$\Delta \mathbf{y}_t = \mathbf{C}(L)\,\mathbf{u}_t$$

From which, by applying the algebra illustrated in our discussion of the univariate Beveridge-Nelson decomposition, we can derive the following stochastic trends representation:

$$\mathbf{y}_t = \mathbf{C}^*(L)\,\mathbf{u}_t + \mathbf{C}(1)\,\mathbf{z}_t$$

where $\mathbf{z}_t$ is a process for which we have $\Delta \mathbf{z}_t = \mathbf{u}_t$. The existence of cointegration imposes some restrictions on the C matrices, in fact the stochastic trends must cancel out when the k stationary linear combinations of the variables in $\mathbf{y}_t$ are considered in other words we must have:

$$\boldsymbol{\beta}' \mathbf{C}(1) = 0$$

By investigating further the relation between the VECM and the stochastic trends representations we can give a more precise parameterisation of the matrix $\mathbf{C}(1)$.

Note first that equation (2.21) can be re-written as :

$$\mathbf{y}_t = \left(I_m + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\mathbf{y}_{t-1} + \mathbf{u}_t \tag{2.32}$$

Premultiplying this system by $\boldsymbol{\beta}'$ yields:

$$\begin{aligned}\boldsymbol{\beta}'\mathbf{y}_t &= \boldsymbol{\beta}'\left(I_m + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\mathbf{y}_{t-1} + \boldsymbol{\beta}'\mathbf{u}_t \\ &= \left(I_k + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\beta}'\mathbf{u}_t\end{aligned}$$

Solving this model recursively, we obtain the MA representation for the k cointegrating relationships:

$$\boldsymbol{\beta}'\mathbf{y}_t = \sum_{i=0}^{\infty}\left(I_k + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)^i \boldsymbol{\beta}'\mathbf{u}_{t-i} \tag{2.33}$$

By substituting (2.33) in (2.21) we have the MA representation for $\Delta\mathbf{y}_t$:

$$\Delta\mathbf{y}_t = \sum_{i=1}^{\infty}\boldsymbol{\alpha}\left(I_k + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)^{i-1}\boldsymbol{\beta}'\mathbf{u}_{t-i} + \mathbf{u}_t$$

from which we have:

$$\mathbf{C}(1) = I_n - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}' \tag{2.34}$$

Now note the beatiful[3] relation

$$I_n = \beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp' + \alpha\left(\beta'\alpha\right)^{-1}\beta' \tag{2.35}$$

where $\beta_\perp, \alpha_\perp$ are $((m \times (m-k))$ matrices of rank $m-k$ such that $\alpha_\perp'\alpha = 0, \beta_\perp'\beta = 0$.

By using (2.35) in (2.34) , we have

$$\mathbf{C}(1) = \beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp'$$

and

$$\mathbf{y}_t = \mathbf{C}^*(L)\,\mathbf{u}_t + \beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}\left(\alpha_\perp'\mathbf{z}_t\right)$$

which shows that a system of m variables with k cointegrating relationships features $(m-k)$ linearly independent common trends (**TR**). The common trends are given by $(\alpha_\perp'\mathbf{z}_t)$, while the coefficients on these trends are $\beta_\perp\left(\alpha_\perp'\beta_\perp\right)^{-1}$. Note also that stochastic trends depend on a set of initial conditions and on cumulated disturbances in fact

[3]See Johansen([34]) ,page 40

$$\mathbf{TR}_t = \mathbf{TR}_{t-1} + C(1)\,\mathbf{u}_t$$

Our brief discussion should have made clear that the VECM model and the MA model are complementary. As a consequence the identification problem relevant for the vector of parameters in the cointegrating vectors β is also relevant for the vector of parameters determining the stochastic trends $\alpha_\perp$. However, there is one aspect in which the two concepts are different. In theory, identified cointegrating relationships on a given set of variables should be robust to augmentation of the information set by adding new variables, which should have a zero coefficient in the cointegrating vectors of the VECM representation of the larger information set. This is not true for the stochastic trends. Consider the case of augmenting an information set consiting of m variables admitting k cointegrating vectors to $m+n$ variables, the number of cointegrating vectors is constant while the number of stochastic trends increases by n, moreover an unanticipated shock in a small system need not to be unanticipated in a larger system. Note that we have added in theory to our statement, this is because, in practice, given the size of available samples application of the procedure to analyze cointegration in a larger set of variables might lead to identify different cointegrating relationships from those obtained on a smaller set of variables.

2.7.7 *VECM and common trends representations*

The joint behaviour of consumption and income under the Permanent Income Hypothesis (PIH) is a good empirical example to illustrate VECM and common trends representations. Let $y_t, y_t^p$and c_t denote respectively the logarithms of aggregate disposable income, permanent income and consumption. Under PIH the joint distribution of consumption and income can be characterised as follows:

$$\begin{aligned} y_t &= y_t^p + v_t \\ y_t^p &= \mu_y + y_{t-1}^p + u_t \\ c_t &= y_t^p \end{aligned}$$

permanent income is the stochastic trend of income, which is made of the permanent component and of a transitory component, v_t and u_t are the shocks to the transitory and the permanent component of income, it is natural to think of them as orthogonal shocks normally and independently distributed. Consumption and income are cointegrated, in fact they share the single unobservable common stochastic trend in this system.

By eliminating the unobservable stochastic trend from the system, we have a bi-variate structural representation:

$$\begin{aligned} y_t &= c_t + v_t \\ c_t &= \mu_y + c_{t-1} + u_t \end{aligned} \tag{2.36}$$

We obtain the VAR(1) representation by substituting for c_t in the first equation from the second equation of (2.36)

$$\begin{pmatrix} y_t \\ c_t \end{pmatrix} = \begin{pmatrix} \mu_y \\ \mu_y \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} w_t \\ u_t \end{pmatrix}$$
$$w_t = u_t + v_t$$

From which we immediately obtain the VECM representation

$$\begin{pmatrix} \Delta y_t \\ \Delta c_t \end{pmatrix} = \begin{pmatrix} \mu_y \\ \mu_y \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} w_t \\ u_t \end{pmatrix}$$

where

$$\begin{aligned} \Pi &= \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} -1 & 1 \end{pmatrix} \\ &= \alpha \beta' \end{aligned}$$

The derivation of the common trends representation is derived by considering that, as $y_t - c_t = v_t$, the MA representation for consumption and income growth is then

$$\begin{pmatrix} \Delta y_t \\ \Delta c_t \end{pmatrix} = \begin{pmatrix} \mu_y \\ \mu_y \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w_t \\ u_t \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} w_{t-1} \\ u_{t-1} \end{pmatrix}$$

from which we have

$$\begin{pmatrix} y_t \\ c_t \end{pmatrix} = \begin{pmatrix} \mu_y \\ \mu_y \end{pmatrix} t + \mathbf{C}^* (L) \begin{pmatrix} w_t \\ u_t \end{pmatrix} + C(1)\mathbf{z}_t$$

and

$$\begin{aligned} C(1) &= \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp \\ \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \left[\begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right]^{-1} \begin{pmatrix} 0 & 1 \end{pmatrix} \end{aligned}$$

Given that in this application $\left(\alpha'_\perp \beta_\perp\right)^{-1} = 1$, it follows that consumption and income have a single common stochastic trend. Such trend can be represented as $\alpha'_\perp \left(\begin{pmatrix} \mu_y \\ \mu_y \end{pmatrix} t + \begin{pmatrix} \sum_{i=1}^{t} w_t \\ \sum_{i=1}^{t} u_t \end{pmatrix} \right)$, and only shocks to the permanent component of income enter the trend.

2.8 Multivariate cointegration: an application to US data

To illustrate empirically how cointegration analysis is performed let us consider monthly data from the US economy for the variables considered in basic macroeconomic models: the log of the real M2 $(m-p)$,annual, seasonally adjusted, CPI inflation (π) , the log of monthly real GDP (y) , the nominal own return on M2 (R^m) , the nominal opportunity cost of holding money as measured by the interest rate on three-month Treasury Bills $\left(R^b\right)$. All series except R^m are those used in Leeper-Sims-Zha([38]) , R^m has been retrieved by the St.Louis FED Website at *http://www.stls.frb.org/fred/*. They are available in the file LSZUSA.XLS. We shall perform cointegration analysis using the package PC-FIML by Doornik and Hendry ([10]), alternative menu-driven packages are available in RATS (see [41], [21]) , E-Views does not allow to perform all the steps of the analysis in that specification and testing of the long-run restrictions are not (yet) available.

2.8.1 *Specification of the VAR*

The first step of the empirical analysis is the specification of the VAR. The specification of the VAR requires the consideration of two issues pertaining respectively to the set of variables included in the VAR and to the lag length of the VAR. These are important issues in that mis-specification of the VAR leads to misleading inference. In general the set of variables to be included in the VAR is determined by the economic problem at hand, however this criterion does not rule out the possibility of mis-specification. Consider the case of the set of variables chosen for our example, they include all the variables used in a simple IS-LM model of a closed economy, but nothing guarantees that the US economy is correctly described by such model. Suppose that the central bank targets expected inflation by using short term interest rates as an instrument. The model is mis-specified if it omits any leading indicator for inflation monitored by Central Bank. An obvious candidate is the commodity price index but there might be more, such as long-term interest rates or other asset prices. In absence of an obvious baseline model, the behaviour of residuals is taken as an indicator of mis-specification. In a correctly specified model residuals should be random normal variables with zero mean and constant variance-covariance matrix, departure of fitted residuals from those hypotheses could be taken as an indicator of mis-specification. However, even when all the relevant variables have been included, the model could be still mis-specified because of omitted relevant dynamics. The selection of the order of the VAR is an important step in the specification. Sims([52]) suggests a statistics to test the validity of restrictions imposed on a general model:

$$(T-k)\left[\log|\Sigma_r| - \log|\Sigma_{unr}|\right]$$

where T is the sample size, k is the number of parameters estimated in each equation of the VAR, $|\Sigma_r|$ is the determinant of the variance-covariance ma-

trix of the residuals of the fitted restricted model, $|\Sigma_{unr}|$ is the determinant of the variance-covariance matrix of the residuals of the fitted general unrestricted model. The statistic has a χ^2 distribution with degrees of freedom equal to the number of restrictions in the system. The term $(T-k)$ includes a small sample correction, in fact as T becomes larger the correction for small sample $(T-k)/T$ converges towards unity. Obviously, the selection of variables and the selection of the lag length are not independent processes, in fact a longer lag length might be the consequence of omission of one relevant variable from the VAR. In practice we shall start from a baseline VAR including the set of variables suggested by the theory and a sufficiently long lag, check the behaviour of residuals. When well-behaved residuals are obtained, we proceed to reduction of the lag length by testing the validity of the implied restrictions.

Our general baseline model is a VAR estimated over the sample 1960:1-1979:6, including fifteen lags of each of the five variables, a constant and a trend, so we have:

$$\begin{pmatrix} y_t \\ \pi_t \\ (m-p)_t \\ R_t^m \\ R_t^b \end{pmatrix} = \mathbf{a}_o + \mathbf{a}_1 t + \sum_{i=1}^{14} A_i L^i \begin{pmatrix} y_t \\ \pi_t \\ (m-p)_t \\ R_t^m \\ R_t^b \end{pmatrix} + \mathbf{u}_t. \tag{2.37}$$

We have chosen to end our estimation in 1979 because, from the second part of 1979 to 1982, the Fed has changed is operating procedure moving from an interest rate targeting regime to a reserves targeting regime. As a consequences paramaters in the Fed's reaction function must have changed. It is very important to estimate cointegrating models using data from a single regime. In fact, structural instability might be very dangerous when cointegrated models are used. The intuition is very simple, in the presence of parameters instability a cointegrated model is very likely to push the system towards the "wrong" long-term equilibria with very serious consequences for forecasting and policy simulation. Checking residual behavious is very important also with this respect, as pathologic behaviour of residuals is a clear symptom of parameters' instability.

The estimation of our base-line model delivers the set of residuals reported in Figure 2.12.

The residual are normalized, hence residual with absolute value higher than 1.96 occur with a probability of one percent under the null of normality. We note many outliers, in fact when a formal test of normality of residuals is performed the null is strongly rejected[4]. This is worrying in that non-normality might signal mis-specification but also in that departure from normality might induce misleading inference in the application of the Johansen procedure. Interestingly, most outliers occur on occasion of the oil price crises. So, prob-

[4]We shall discuss tests for normality at a later stage of the book

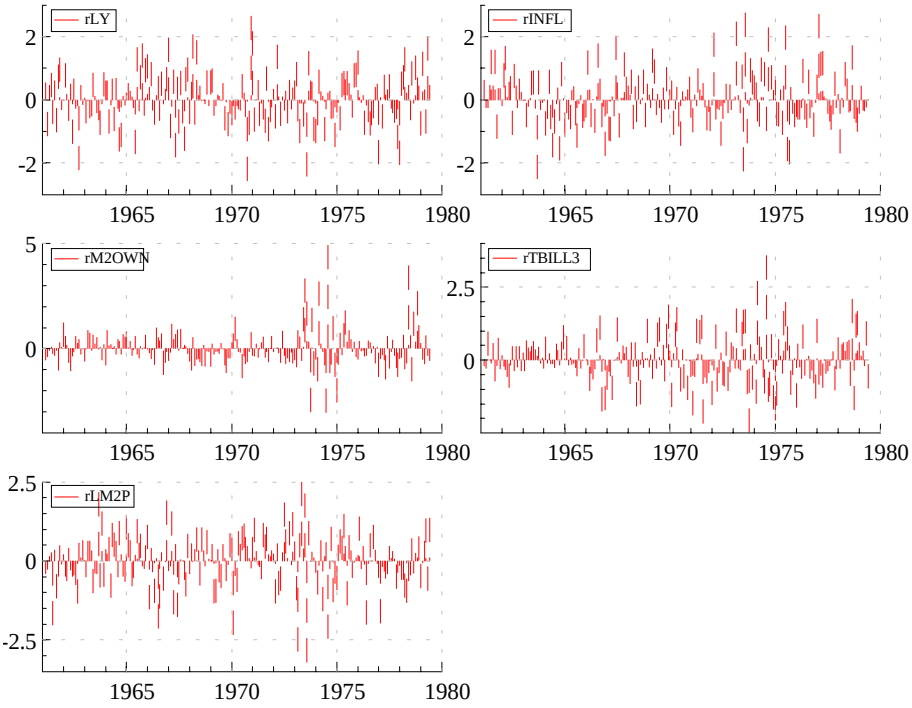

FIG. 2.12. VAR residuals with outliers

ably a commodity price index is the relevant omitted variables causing non-normality in the residuals. However, the inclusion of a commodity price index as a further endogenous variable in our system would simply shift the outlier problem from our equation for interest rates to the commodity price index. In fact no variable included in this system has an high explanatory potential for a commodity price index. We have then included in the system contemporaneous and lagged (up to the sixth lag) commodity price inflation. We consider commodity price inflation as a stationary exogenous variable, this choice shall be discussed later on. We have also included dummies for exceptional periods during the oil price crises as exogenous variables in our system. In general, dumMMYY is a variable taking value 1 in the MM month of the year YY and zero anywhere else. We include the following dummy variables: dum7306,dum7307, dum7308, dum7310, dum7311, dum7312, dum7402, dum7403, dum7407, dum7408,dum7409, dum7501, dum7505, dum7806, dum7808, dum7811, dum7904. Note that the inclusion of dummies and exogenous variables in the specification modifies the deterministic nucleus of our model and appropriate critical values for the tests should be re-computed(see??,??). We do not

take this step and report the critical values automatically indicated by version 9 of Pc-Fiml in Table 7.

The inclusion of the dummies in the system delivers a new set of residuals, reported in Figure 2.13, which are virtually free from outliers and do not show any departure from the hypothesis of normality.

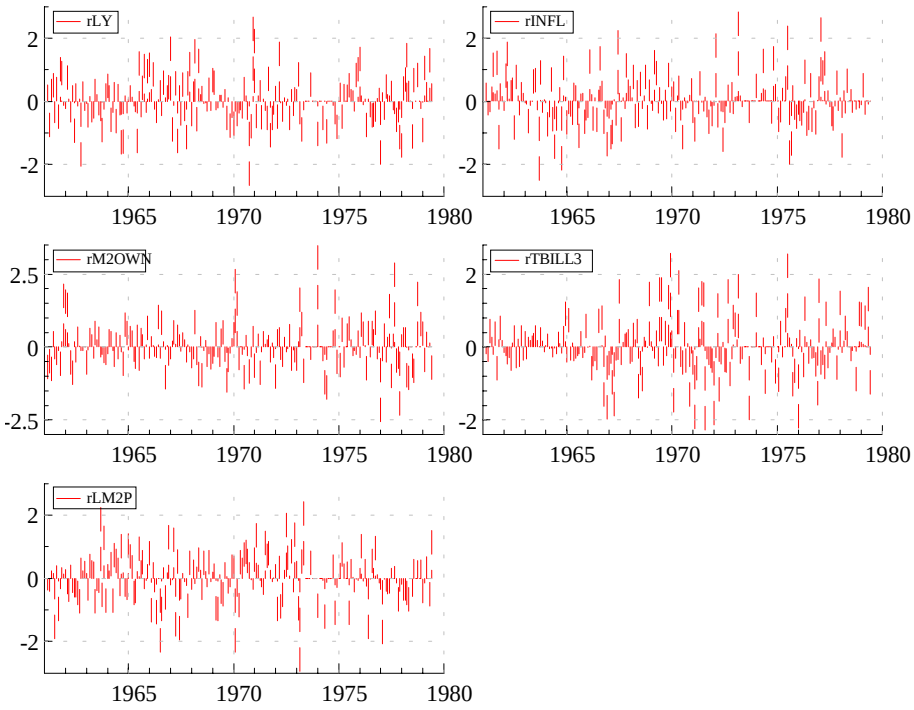

FIG. 2.13. VAR residuals without outliers

We proceed then to assess the possibility of simplification of the system. The progressive simplification strategy, based on the likelihood ratio tests discussed above, leads us to a specification with 6 lags.

2.8.2 *Selection of the deterministic components in the VECM specification*

The choice of the determistic components in the VAR is not trivial, given that it affects the distribution of the relevant statistics to perform cointegration analysis. Given the following general VECM model:

$$\Delta \mathbf{y}_t = \mu_0 + \mu_1 t + \Pi_1 \Delta \mathbf{y}_{t-1} + \Pi_1 \Delta \mathbf{y}_{t-2} + ... + \Pi \mathbf{y}_{t-n} + \mathbf{u}_t$$

five possible specifications for the deterministic components have been considered in the literature:

(*i*) $\mu_0 = 0, \mu_1 = 0$ this would determine a zero mean in the I(0) components and a non-zero mean in the I(1) components

(*ii*) $\mu_0 = \alpha\beta_1, \mu_1 = 0$ this would restrict the constant to belong to the cointegrating space inducing a non-zero mean both in the I(0) and the I(1) components

(*iii*) $\mu_0 = unrestricted, \mu_1 = 0$ this would generate a zero mean in the I(0) components and a linear trend in the I(1) component

(*iv*) $\mu_0 = unrestricted, \mu_1 = \alpha\beta_1$ this would generate a linear trend both in the I(0) and the I(1) components

(*v*) $\mu_0 = unrestricted, \mu_1 = unrestricted$, this would generate a linear trend in the I(0) components and a quadratic trend in the I(1) components

Different critical values have been tabulated for each specification ([**?**]) and are now automatically available in all packages peforming the Johansen procedure. Note that the inclusion of intervention dummies also modifies the deterministic components of the VAR and this requires in principle in ad-hoc tabulation of the relevant critical values([37]).

In our application we choose specification (*iv*) as some of our series show trends in levels and, as already stated, we ignore the modification of the relevant distribution generated by the inclusion of dummies.

2.8.3 *Test for the rank of* Π

Having specified the VAR and chosen the specification of th deterministic component , we can estimate the Π matrix and start our analysis of the long-run properties of the system.We apply the Johansen procedure to identify the rank of the matrix Π in the following re-parameterisation of our model:

$$\begin{bmatrix} \Delta y_t \\ \Delta \pi_t \\ \Delta R_t^m \\ \Delta R_t^b \\ \Delta\,(m-p)_t \end{bmatrix} = \begin{bmatrix} d_{0,11} \\ d_{0,21} \\ d_{0,31} \\ d_{0,41} \end{bmatrix} + \sum_{i=0}^{6} F_i \Delta_{12} LPCM_{t-i} + \mathbf{g}'\mathbf{DUM}_t +$$

$$\sum_{i=1}^{5} D_i \begin{bmatrix} \Delta y_{t-i} \\ \Delta \pi_{t-i} \\ \Delta R_{t-i}^m \\ \Delta R_{t-i}^b \\ \Delta\,(m-p)_{t-i} \end{bmatrix} + \Pi \begin{bmatrix} y_{t-6} \\ \pi_{t-6} \\ R_{t-6}^m \\ R_{t-6}^b \\ (m-p)_{t-6} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \\ u_{5t} \end{bmatrix}$$

The results of the Johansen procedure are reported in Table 6:

TABLE 6: Analysis of the Π matrix in the estimated VAR model

Eigenvalue	$H_0 : rank = p$	Eigenval	C.Eigenval	95%	Trace	C.Trace	95%
0.185	$p = 0$	45.74	39.61	37.5	134.1	116.1	87.3
0.148	$p \leq 1$	35.93	31.11	31.5	88.36	76.53	63
0.133	$p \leq 2$	32.14	27.84	25.5	52.44	45.41	42.4
0.083	$p \leq 3$	19.56	16.94	19	20.3	17.58	25.3
0.0033	$p \leq 4$	0.74	0.64	12.3	0.74	0.64	12.3

where Eigenvalis the max eigenvalue test, C.Eigenvalue is the max eigenvalue test corrected for small sample, i.e. using $T - k$ instead of T, Trace is the trace test, C.Trace is the trace test corrected for small sample amd 95% are the critical values tabulated for our specification of the deterministic components. Table 7 poses an interesting problem to the applied reseracher in that the trace statistics and the maximum eigenvalue statistic deliver different results, with more relevant differences in the case of the adoption of small sample correction for the statistics. We opt for rejecting the null of at most one cointegating vector and do not reject the null of at most two. Of course, such choice is debatable.

Note that, before any identifying restrictions are introduced, most available cointegrating packages do deliver some point estimates of the α and β matrices as follows:

TABLE 7: Cointegrating vectors: the Johansen interpretation.

		Standardised β'			
y	π	$m-p$	R^m	R^b	*Trend*
1	0.078	-0.40	3.55	-3.96	-0.18
1.08	1	-0.61	-1.20	-1.08	-0.15
		Standardised α			
y	-0.02	-0.013			
π	0.02	-0.005			
$m-p$	0.047	-0.018			
R^m	-0.00008	-0.002			
R^b	0.03	0.01			

These estimates are obtained by imposing a default identification which delivers cointegrating vectors orthogonal to each other ([36]). In some context, for example a demand and supply system, this assumption might be the economic case of interest. However, this is not the case in general and in our specific example. In the next section we shall evaluate the potential of different identification of economic interest by checking the validity of over-identifying restrictions.

2.8.4 *Specification and testing of the long-run restrictions*

We consider two different proposals. In the first one we identify a traditional money demand and a relation beween the own interest rate on money and the opportunity cost of holding money. As an alternative, we identify an interest rate reaction funcion in which the nominal interest rate responds to inflation, output and a linear trend, alongwith a relation between interest rates and inflation. We have selected these two specifications because, as we shall see, they form the basis for two alternative targeting strategies: inflation targeting via the control of money growth and inflation targeting via the control of interest rates.

We can parameterise the restrictions implied the first identification scheme as follows:

$$\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \\ \alpha_{41} & \alpha_{42} \\ \alpha_{51} & \alpha_{52} \\ \alpha_{61} & \alpha_{62} \end{pmatrix}$$

$$\boldsymbol{\beta}' = \begin{pmatrix} \beta_{11} & 0 & 1 & \beta_{41} & \beta_{51} & \beta_{61} \\ 0 & \beta_{22} & 0 & 1 & \beta_{52} & 0 \end{pmatrix}$$

The results, reported in Table 9, show that the two over-identifying restrictions are not rejected. The first cointegrating vector is consistent with a money demand function as far as the semi-elasticities with respect to interest rates are concerned, the elasticity with respect to income is somewhat high, although it is compensated by a deterministic trend with the opposite sign. However, looking at the weights on the cointegrating vectors we note that real money reacts very little to disequilibrium in the first cointegrating relationship. In fact the only strongly significant weight is the one describing the reaction of real income to disequilibrium in the second cointegrating relationship.

TABLE 9: A scheme of overidentified cointegrating vectors

Standardised β'					
y	π	$m-p$	R^m	R^b	*Trend*
-2.20	0	1	-7.29	7.51	0.38
(0.17)			(2.16)	(0.96)	(0.06)
0	1.08	0	-3.16	1	0
	(0.22)		(0.59)		

Standardised α		
y	0.064	-0.17
	(0.015)	(0.036)
π	-0.0016	-0.034
	(0.009)	(0.021)
$m-p$	-0.019	-0.014
	(0.009)	(0.023)
R^m	-0.0006	0.002
	(0.0001)	(0.003)
R^b	-0.023	0.03
	(0.008)	(0.02)

LR-test, rank=2:$\chi^2(2) = 1.03$ [0.59]

We then consider the second alternative and parameterise restrictions as follows:

$$\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \\ \alpha_{41} & \alpha_{42} \\ \alpha_{51} & \alpha_{52} \\ \alpha_{61} & \alpha_{62} \end{pmatrix}$$

$$\boldsymbol{\beta}' = \begin{pmatrix} \beta_{11} & -1 & 0 & 0 & 1 & \beta_{51} \\ 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

The results, reported in Table 9, show the plausibility of the interpretation of the first cointegrating relation as a reaction function for the monetary policy maker. Policy rates react to inflation, with a coefficient which can be restricted to one, and to deviation of output form a deterministic trend (a non-stationary variable in our specification). The estimated weights strongly support the identification of this relationship as an equilibrium for the policy rates. The second cointegrating vector does not differ significantly from the one obtained within the first identification scheme.

TABLE 9: A scheme of overidentified cointegrating vectors

Standardised β'					
y	π	$m-p$	R^m	R^b	$Trend$
-0.22	-1	0	0	1	0.08
(0.03)				(0.68)	(0.009)
0	-0.96	0	2.75	1	0
	(0.17)		(0.47)		

Standardised α		
y	0.13213	-0.002
	(0.069)	(0.039)
π	0.047	-0.017
	(0.038)	(0.022)
$m-p$	-0.09	-0.09
	(0.042)	(0.034)
R^m	0.007	0.003
	(0.005)	(0.0029)
R^b	-0.13	-0.062
	(0.036)	(0.021)

LR-test, rank=2: $\chi^2(4) = 6.1$ [0.19]

We conclude our analysis of these two alternative identification schemes by stressing that statistical criteria do not lead to an unequivocal identification, then the choice between the two alternatives is very likely to rely on economic criteria.

2.9 Multivariate decompositions: some considerations

The purpose of our illustration of the Johansen procedure in the previous section was to show that the identification of cointegrating vectors requires involves a multi-step process. The outcome of many of these steps is not so clear-cut and therefore the final product might be differ across researchers. The presence of structural breaks paired with the specification issues and size of available samples have an important impact on the empirical application of the procedure. Alternative methods to the this procedure have been proposed in the literature, see, for example, Horvath and Watson([28]), Phillips ([45]) , Reimers([47]) and Saikkonen([49]) . However, it is important to remember that the specification of a dynamic model in levels has proved sufficient to remove the spurious regression problem and that the VECM representation of a VAR model in level is just a reparameterisation, before the rank reduction restrictions are imposed. Sims,Stock and Watson([53]) , argue that a VAR model in levels in the presence of cointegration is over-parameterised and therefore leads to inefficient but consistent estimates of the parameters of interest. The loss of efficiency has to be weighted against the risk of inconsistency of estimates which occurs when the "wrong" cointegrating restrictions are imposed. Imposing the "wrong" cointegrating parameters will make the system converge to the "wrong" long-run equilibria but it will also bias the short-run dynamics as the system is pulled in the wrong direction. For this reason the recent research has taken a defined line

and VAR in levels rather than co-integrated VARs are used when the issue of economic interest is not related to the short-run rather than to the long-run. A standard example is the analysis of the monetary transmission mechanism. Of course there is more in Sims,Stock and Watson([53]) than these considerations. In fact they show that standard distribution can be applied when doing inference in a VAR model which involves variables admitting stationary linear combinations, reverting to non-standard distributions is necessary only when the subsets of variables on which inference is performed do not admit any stationary linear combination.

As a matter of fact cointegrated VARs are mainly data-driven specifications. The macro-model for the relevant DGP is not fully specified, as it is clearly the case with the example discussed in this chapter where we started off our investigation by a model centered on money and we end up specifying a long-run structure where the quantity of money, being fully demand determined, plays no role in the monetary transmission mechanism. It is not easy to interpret the results from a simultaneous model, when we have (loose) theories generating only a subset of the equations. Moreover, there are difficulties with an approach aimed discriminating between theories on the basis of the outcome of test statistics, based on a number of joint hypothesis, some of which are clearly independent from the theories tested. There is also an issue with the critical values for the testing procedures in the Johansen framework. First, they depend crucially on the specification of the deterministic nucleus of the VAR , so the inclusion of dummies for outliers introduces modifications in the relevant critical values. A solution to this problem is available, see Johansen-Nielsen[37]. Second, recent work by Johansen[35], has shown that it is important to implement small sample corrections for the asymptotic critical values, when applicable. Taking these two aspects together, it is likely that a re-assessment of all the empirical evidence proposed in the nineties without implementing the appropriate corrections is necessary. So what do we make of all the sentences issued on theories using the wrong critical values?

Note also that cointegration analysis based on a multi-step framework: specification of the VAR and its deterministic component, identification of the number of cointegrating vectors, identification of the parameters in cointegrating vectors, tests on the speed of adjustment with respect to disequilibria. The results of the final test depend on the outcome of the previous stages in the empirical analysis, but the outcome of each step is not so easily and uniquely established empirically.

Of course there is something to be said for a methodology aimed at exploiting cointegration to deliver a stationary representation of a non-stationary vector autoregressive process in which short-run and long-run dynamics are naturally separated and sound statistical inference can be applied. However, the practical implementation of such methodology requires theresearcher to deal with specification and identification problems which are not easily, and above all not uniquely, solved.

REFERENCES

Amisano G., Giannini C.(1997) Topics in Structural VAR Econometrics, second edition, Springer Verlag, New York

Banerjee A., Dolado J.J., Hendry D.F. and Smith G.W. (1986) Exploring Equilibrium Relationships in Econometrics through Static Models: some Monte Carlo Evidence, Oxford Bulletin of Economics and Statistics, August

Banerjee A., Dolado J.J., Galbraith J.W. and Hendry D.F.(1993) Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data, Oxford University Press, August

Banerjee A., Hendry D.F.(1992) "Testing Integration and Cointegration", Oxford Bulletin of Economics and Statistics, 54, Special Issue

Beveridge S. and C.-Nelson(1981) A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to the Measurement of Business Cycle Journal of Monetary Economics, 151-174

Cuddington J.T. and L.A.Winters(l987) "The Beveridge-Nelson decomposition of economic time-series. A quick computational method" Journal of Monetary Economics,19, 125-127.

Bhargava A.(1986) "On the Theory of Testing for Unit Roots in Observed Time Series", Review of Economic Studies,53 pp. 369-384

DeJong D.N. and C.H. Whiteman (1991) Reconsidering "Trends and random walks in macroeconomic time-series" , Journal of Monetary Economics, 28, 221-254

Dickey D.A.,Fuller W.A.(1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", Econometrica,49,pp.1057-1072

Doornik J. and Hendry D.F.(1994)"PcFiml 8.0. Interactive Econometric Modelling of Dynamic Systems", International Thomson Publishing, London

Enders W.(1995) Applied Econometric Time Series, Wiley Series in Probability and Mathematical Statistics, New York

Engle R.F.,Granger C.W.J.(1987) "Cointegration and Error Correction : Representation, Estimation and Testing" Econometrica,55,pp.251-276

Engle R.F.,Granger C.W.J.,Hallman J.J.(1987) "Combining Short and Long Run Forecasts: an Application of Seasonal Cointegration to Monthly Electricity Sales Forecasting",Unpublished paper, University of California San Diego

Fuller W.A. (1976) Introduction to Statistical Time Series, New York,J.Wiley ans Sons

Giannini C. (1992) Topics in Structural VAR Econometrics, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag

Giannini C., Lanzarotti S. and M. Seghelini (1994) "A Traditional Interpretation of Macroeconomi Fluctuations: the Case of Italy.", European Journal of Political Economy

Granger C.W.J.(1986) "Developments in the Study of Cointegrated Economic Variables",Oxford Bulletin of Economic and Statistics,48,pp.213-228

Gonzalo J. (1994) "Five Alternative Methods of Estimating Long-run Equilibrium Relationships", Journal of Econometrics, 60, 203-233

Granger C.W.J.,Newbold P.(1974) "Spurious Regressions in Econometrics", Journal of Econometrics,2, pp.111-120

Hamilton J. (1994) Time-Series Analysis, Princeton, Princeton University Press,

Hansen H. and K.Juselius(1995) "CATS in RATS. Cointegration Analysis of Time Series", Estima, Evanston, Illinois

Harvey,A.C. and Jaeger A.(1993) 'Detrending, Stylized Facts and the Business Cycles' Journal of Applied Econometrics

Hatanaka M.(1996) Time-Series-based-econometrics, Oxford university Press, Oxford

Hendry D.F.(1987) "Econometric Modelling with Cointegrated Variables", Oxford Bulletin of Economic and
Statistics,48,pp.201-212

Hendry D.F. (1995) Dynamic Econometrics Oxford University Press

Hendry D.F. and N.Ericsson(1991) Modeling the demand for narrow money in the United Kingdom and the United States, European Economic Review, 35, 833-886

Hodrick, R.J. and E.C. Prescott (1997) Postwar U.S. Business Cycles: An Empirical Investigation, Journal of Money, Credit, and Banking, 29, 116.

Horvath M.T.K., and M.W. Watson(1995) "Testing for cointegration when some of the cointegrating vectors are known" Econometric Theory, 11, 984-1014

Kiwiatkowski D., P.C.B. Phillips, P.Schmidt and Y. Shin (1992) Testing the Null Hypothesis of Stationarity against the alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root ?, Journal of Econometrics, 54, 159-178

Johansen S. (1988), Statistical Analysis of Cointegration Vectors, "Journal of Economics Dynamics and Control, March.

Johansen S. (1989) Likelihood-based Inference on Cointegration: Theory and Applications, mimeo, University of Copenhagen

Johansen S.(1992) Identifying Restrictions of Linear Equations, University of Copenaghen, Institute of Mathematical Statistics

Johansen S.(1994) The role of the Constant and Linear Terms in Cointegration Analysis of Nonstationary Variables, Econometric Reviews,13, 205-229

Johansen S. (1995) Likelihood Based Inference on Cointegration in the Vector Autoregressive Model, Oxford University Press, Oxford

Johansen, S. (1999), "A Bartlett Correction Factor for Tests on the Cointegrating Relationships", downloadable from http://www.iue.it/Personal/Johansen/Welcome.html

Johansen S., Juselius, K. (1990), Maximum Likelihood Estimation and Inference on Cointegration with Applications to the Demand for Money, "Oxford Bulletin of Economics and Statistics", May.

Johansen S., Nielsen B.G.(1993) Asymptotics for the cointegration rank test tests in the presence of intervantion dummies, Manual for the simulation program DisCo, manual and program are available at the URL http://www.nuff.ox.ac.uk/users/nielsen/d

Leeper E.M., C.A. Sims and T. Zha (1997) "What does monetary policy do?", Brookings Paper on Economic Activity, also available at ftp://ftp.econ.yale.edu/pub/sims/mpolicy

Maddala G.S., and In-Moo Kim(1998) Unit roots, Cointegration and Structural Change, Cambridge University Press, Cambridge

Mellander E., Vredin A. e A.Warne (1993) "Stochastic Trends and Economic Fluctuations in a Small Open Economy" Journal of Applied Econometrics

Mosconi R.(1998) "Malcolm. The theory and practice of cointegration analysis in RATS", Cafoscarina, Venezia

Nelson, C.R. e C.I. Plosser(1982) Trends versus Random Walks in economic time-series. Some evidence and implications, Journal of Monetary Economics, 10, 139-162

Osterwald-Lenum M.(1992) "A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics", Oxford Bulletin of Economics and Statistics, 54, 461-472

Perron P.(1987) "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", Research Paper, Department of Economics, University of Montreal,Quebec, Canada

Phillips P.C.B. (1991) Optimal Inference in Cointegrated Systems, Econometrica, 59, 283-306

Phillips P.C.B. e P. Perron (1988) Testing for a unit root in time-series regression, Biometrika, 75, 335-346.

Reimers H.E. (1992) "Comparison of tests for multivariate cointegration", Statistical Papers, 33, 335-359

Salmon M.(1982) "Error Correction Mechanisms",The Economic Journal,92,pp.615-629.

Saikkonen P.(1993) Estimation of cointegration Vectors with Linear Restrictions, Econometric Theory, 9, 19-35

Sargan D. (1964) Wages and Prices in the United Kingdom: A Study in Econometric Methodology, in P.E. Hart, G.Mills and J.K. Whitaker (eds.), Econometric Analysis of National Economic Planning, London, Butterworth, 25-63

Sargan J.D.,Barghava A.(1983) "Testing Residuals from Least Squares Regression for Being Generated by Gaussian Random Walk",Econometrica,51,pp.153-174

Sims C.(1980) "Macroeconomics and Reality" Econometrica

Sims C., Stock J. e Watson M.(1990) Inference in Linear Time Series Models with Some Unit Roots Econometrica, 58, 1, 113-144

Spanos A.(1986)"Statistical Foundations of Econometric Modelling",Cambridge University Press,Cambridge

Stock J.H.(1987)"Asymptotic Properties of the Least Squares Estimator of Cointegrating Vectors." Econometrica, 55,pp. 1035-1056

Stock J.H.,West K.D.(1987) "Integrated Regressors and the Test of the Permanent Income Hypothesis",Princeton University,Woodrow Wilson School D.P. 127

Zivot, E. e D.W.K. Andrews(1992) Further Evidence on the Great Crash, the Oil-Price Shock and the Unit Root Hypothesis", Journal of Business Economics and Statistics,10, 251.270

Mellander E., A.Vredin and A.Warne(1992) Stochastic trends and economic fluctuations in a small open economy, Journal of Applied Econometrics, 7, 369-394

King R.G., C.I.Plosser, J.H.Stock,M.W.Watson(1991) "Stochastic trends and economic fluctuations" the American Economic Review, 81,4, 820-840

White H. (1984) "Asymptotic theory for econometricians", Academic Press, London