Søren Johansen

## LIKELIHOOD BASED

INFERENCE FOR COINTEGRATION
OF NON-STATIONARY

## TIME SERIES

Institute of Mathematical Statistics University of Copenhagen

## Søren Johansen

# LIKELIHOOD BASED INFERENCE FOR COINTEGRATION OF NON-STATIONARY TIME SERIES 

Preprint 1993 No. 5

# LIKELIHOOD BASED INFERENCE FOR COINTEGRATION OF NON-STATIONARY TIME SERIES. 

by
Søren Johansen ${ }^{1 .}$
October 1993

Institute of Mathematical Statistics
University of Copenhagen
Universitetsparken 5
2100 Copenhagen $\emptyset$
Denmark

## Abstract

This paper presents a survey of the statistical analysis of the cointegration model for vector autoregressive processes. The focus is on likelihood based inference, but for comparison the regression approach is briefly discussed. It is not the intention to give a complete survey of all results obtained in cointegration, but rather to present in an informal way the basic problems and some results, in the hope that those who catch an interest in the problem area, will be able to find the relevant references for a deeper study.

Key words and phrases: Cointegration, time series, autoregressive processes, likelihood inference, econometrics.

[^0]
## 1. Introduction

This paper deals with likelihood inference for non-stationary time series and as application and motivation we consider some simple economic problems and show how the analysis of the statistical model helps in gaining insight and understanding of economic phenomena. This section contains the basic definitions of integrated variables, cointegration, and common trends and discuss by examples the formulation of models and processes in terms of common trends and the error correction model.

The notion of cointegration has become one of the more important concepts in time series econometrics since the papers by Granger (1983) and Engle and Granger (1987).

The basic idea is very simple: Let $\left\{\mathrm{X}_{\mathrm{t}}, \mathrm{t}=0,1, \ldots\right\}$ be a p -dimensional stochastic process, we then give the definitions.

Definition 1. If $X_{t}$ is non-stationary but $\Delta X_{t}=X_{t}-X_{t-1}$ is stationary we call $X_{t}$ integrated (of order 1).

Definition 2. If $X_{t}$ is integrated of order 1 but some linear combination, $\mu^{\prime} X_{t}, \mu \neq 0$, is stationary then $X_{t}$ is called cointegrated and $\mu$ is the cointegrating vector and $\mu^{\prime} X=0$ the cointegrating relation.

Just to get an idea of the concept consider the following simple example of a three dimensional stochastic process:

Example 1 We define the 3-dimensional process by

$$
\begin{aligned}
& \mathrm{X}_{1 \mathrm{t}}=\sum_{\mathrm{i}=1}^{\mathrm{t}} \epsilon_{1 \mathrm{i}}+\epsilon_{2 \mathrm{t}} \\
& \mathrm{X}_{2 \mathrm{t}}=\frac{1}{2} \sum_{\mathrm{i}=1}^{\mathrm{t}} \epsilon_{1 \mathrm{i}}+\epsilon_{3 \mathrm{t}}, \\
& \mathrm{X}_{3 \mathrm{t}}=\epsilon_{4 \mathrm{t}} \quad \mathrm{t}=1, \ldots, \mathrm{~T} .
\end{aligned}
$$

Here all $\epsilon$ 's are independent identically distributed Gaussian variables with mean zero and variance $\sigma^{2}$. It is seen that $X_{t}$ is non-stationary and that $\Delta X_{t}$ is stationary. Thus $X_{t}$ is an $I(1)$ process. It is also seen that $X_{1 t}-2 X_{2 t}$ is stationary, such that $X_{t}$ is cointegrated with $(1,-2,0)$ as a cointegrating vector. Note that $(0,0,1)$ is a cointegrating vector too, such that the process has two cointegrating relations and one common (stochastic) trend which is a random walk, $\Sigma_{1}^{\mathrm{t}} \epsilon_{1 \mathrm{i}}$. Hence we can include stationary processes in the analysis at the expense of adding an extra cointegrating vector. The processes we consider are composed of a stationary process plus a random walk. If we want to define a linear combination of $X_{t}$ as a common trend, then this is not uniquely defined. We can take $\mathrm{X}_{1 \mathrm{t}}$ or $\mathrm{X}_{2 \mathrm{t}}$ or any combination of these except of course $(1,-2)$.

The idea behind cointegration is that sometimes the non-stationarity of a multidimensional process is caused by common stochastic trends, which can be eliminated by taking suitable linear combinations of the process, thereby making the linear combination stationary.

In economics and other applications of statistics the autoregressive processes have long been applied to describe stationary phenomena and the idea of explaining the process by its past values has been very useful for prediction. If, however, we want to find relations between simultaneous values of the variables in order to understand the interactions of the economy one would get a lot more information by relating the value of a variable to the value of other variables at the same time point rather than relating it to its own past. In the above example Granger's idea is to relate a variable like $\mathrm{X}_{1 \mathrm{t}}$ to $2 \mathrm{X}_{2 \mathrm{t}}$ rather than to $\mathrm{X}_{1 \mathrm{t}-1}$ to obtain stationarity.

A totally different line of development starts with the so called error correction model. These ideas can be traced back to Phillips (1954) who used ideas from engineering to formulate continuous time models and to Sargan (1964) who used the ideas to formulate models for discrete time data. The simplest example of such a
model which still illustrates the main idea is given by
Example 2 We define the processes by the autoregressive model

$$
\begin{array}{ll}
\Delta \mathrm{X}_{1 \mathrm{t}}=-\alpha_{1}\left(\mathrm{X}_{1 \mathrm{t}-1}-2 \mathrm{X}_{2 \mathrm{t}-1}\right)+ & \epsilon_{1 \mathrm{t}} \\
\Delta \mathrm{X}_{2 \mathrm{t}}= & \epsilon_{2 \mathrm{t}}, \mathrm{t}=1, \ldots, \mathrm{~T}
\end{array}
$$

This model expresses the changes in $X_{t}$ at time $t$ as reacting through the adjustment coefficient $\alpha_{1}$ to a disequilibrium error $\mathrm{X}_{1}-2 \mathrm{X}_{2}$ at time $\mathrm{t}-1$. It is not difficult to see that for $0<\alpha_{1}<2$ the model defines $X_{t}$ as non-stationary, but that $\Delta \mathrm{X}_{\mathrm{t}}$ is stationary, and also $\mathrm{X}_{1 \mathrm{t}}-2 \mathrm{X}_{2 \mathrm{t}}$ is stationary.

In fact we can solve the equations for $\mathrm{X}_{\mathrm{t}}$ as a function of the initial values $\mathrm{X}_{0}$ and the disturbances $\epsilon_{1}, \ldots, \epsilon_{\mathrm{T}}$ and find, since $\mathrm{X}_{1 \mathrm{t}}-2 \mathrm{X}_{2 \mathrm{t}}$ is an autoregressive process of order 1 the representation

$$
\mathrm{X}_{1 \mathrm{t}}-2 \mathrm{X}_{2 \mathrm{t}}=\sum_{\mathrm{i}=0}^{\mathrm{t}-1}\left(1-\alpha_{1}\right)^{\mathrm{i}}\left(\epsilon_{1}-2 \epsilon_{2}\right)_{\mathrm{t}-\mathrm{i}}+\left(1-\alpha_{1}\right)^{\mathrm{t}}\left(\mathrm{X}_{10}-2 \mathrm{X}_{20}\right)
$$

so that

$$
\begin{aligned}
& \mathrm{X}_{1 \mathrm{t}}=2{\underset{\mathrm{i}=1}{\mathrm{t}} \epsilon_{2 \mathrm{i}}+\sum_{\mathrm{i}=0}^{\mathrm{t}-1}\left(1-\alpha_{1}\right)^{\mathrm{i}}\left(\epsilon_{1}-2 \epsilon_{2}\right)_{\mathrm{t}-\mathrm{i}}+\left(1-\alpha_{1}\right)^{\mathrm{t}}\left(\mathrm{X}_{10}-2 \mathrm{X}_{20}\right)+2 \mathrm{X}_{20}}_{\mathrm{X}_{2 \mathrm{t}}=\sum_{\mathrm{i}=1}^{\mathrm{t}} \epsilon_{2 \mathrm{i}}+\mathrm{X}_{20} .} .
\end{aligned}
$$

Thus $X_{t}$ is a non-stationary $I(1)$ variable, such that $X_{1 t}-2 X_{2 t}$ is stationary if the initial value $\mathrm{X}_{10}-2 \mathrm{X}_{20}$ is given its invariant distribution, hence $\mathrm{X}_{\mathrm{t}}$ is cointegrated with the cointegrating vector $(1,-2)$.

The conclusion of this is that the simple error correction model can generate processes that are non-stationary but cointegrated.

The first example shows how the presence of common trends in the moving average representation of $X_{t}$ can generate cointegration. The second example shows that suitable restrictions on the parameters of the autoregressive process will produce cointegration. A general result about the relations between the two approaches were
proved by Granger, see Engle and Granger (1987), and is given in the next section.
These definitions raise a number of interesting mathematical, statistical and probabilistic questions, as well as a number of questions concerning the interpretation of cointegration in the various applications.

## Mathematical:

- What kind of non-stationary processes are $\mathrm{I}(1)$ ?
- Which models generate cointegrated processes?


## Interpretational:

- How does one formulate interesting economic hypotheses in terms of cointegrating relations?
- What is the interpretation of the cointegration relations and how can error correction models be usefully applied?


## Statistical:

- How does one determine the number of cointegrating relations and common trends?
- How does one estimate the cointegrating relations and the common trends?
- How does one test hypotheses concerning the cointegrating rank?
- How does one test economic hypotheses on the cointegrating relations?


## Probabilistic:

- What is the (asymptotic) distribution theory for test statistics and estimators?

We discuss some of these questions in the following and illustrate with an application to an economic problem in sections 3 and 6 .
2.

Granger's representation theorem
This section contains a mathematical discussion of properties of autoregressive processes with respect to the question of integration and cointegration. The results are given in Theorem 2 and illustrated by some examples.

Only autoregressive processes will be considered, since they form a convenient framework for the statistical analysis. These processes are easy to estimate and their properties are well understood.

Consider therefore the general vector autoregressive model for the p -dimensional process $\mathrm{X}_{\mathrm{t}}$ defined by the equations

$$
\begin{equation*}
X_{t}=\Pi_{1} X_{t-1}+\ldots+\Pi_{k} X_{t-k}+\epsilon_{t}, t=1, \ldots, T \tag{2.1}
\end{equation*}
$$

where $\epsilon_{\mathrm{t}}, \mathrm{t}=1, \ldots, \mathrm{~T}$ are independent Gaussian variables in p dimensions with mean zero and variance matrix $\Omega$. The initial values $\mathrm{X}_{0}, \ldots, \mathrm{X}_{-\mathrm{k}+1}$ are fixed.

Example 3 As a very simple example consider the univariate autoregressive process

$$
\mathrm{X}_{\mathrm{t}}=\rho \mathrm{X}_{\mathrm{t}-1}+\epsilon_{\mathrm{t}}, \mathrm{t}=1, \ldots, \mathrm{~T}
$$

It is well known that the solution is

$$
\mathrm{X}_{\mathrm{t}}=\sum_{\mathrm{i}=0}^{\mathrm{t}-1} \rho^{\mathrm{i}} \epsilon_{\mathrm{t}-\mathrm{i}}+\rho^{\mathrm{t}} \mathrm{X}_{0}
$$

which shows that if $|\rho|<1$ we can choose $\mathrm{X}_{0}=\sum_{\mathrm{i}=0}^{\infty} \rho^{\mathrm{i}} \epsilon_{-\mathrm{i}}$ or $\rho^{\mathrm{t}} \mathrm{X}_{0}=\sum_{\mathrm{i}=\mathrm{t}}^{\infty} \rho^{\mathrm{i}} \epsilon_{\mathrm{t}-\mathrm{i}}$, such that $\mathrm{X}_{\mathrm{t}}=\sum_{\mathrm{i}=0}^{\infty} \rho^{\mathrm{i}} \epsilon_{\mathrm{t}-\mathrm{i}}$ becomes stationary.

Thus a condition on $\rho$ is needed to produce a stationary process. In more dimensions the situation is a lot more complex. Consider the following simple example

Example $4 \quad$ The process is generated by

$$
\begin{aligned}
& \Delta \mathrm{X}_{1 \mathrm{t}}=\epsilon_{1 \mathrm{t}} \\
& \Delta \mathrm{X}_{2 \mathrm{t}}=\mathrm{X}_{1 \mathrm{t}-1}+\epsilon_{2 \mathrm{t}}, \mathrm{t}=1, \ldots, \mathrm{~T}
\end{aligned}
$$

It is seen that

$$
\mathrm{X}_{1 \mathrm{t}}=\mathrm{X}_{10}+\sum_{\mathrm{i}=1}^{\mathrm{t}} \epsilon_{1 \mathrm{i}}
$$

and hence

$$
\begin{aligned}
& \mathrm{X}_{2 \mathrm{t}}=\sum_{\mathrm{i}=1}^{\mathrm{t}} \mathrm{X}_{1 \mathrm{i}-1}+\sum_{\mathrm{i}=1}^{\mathrm{t}} \epsilon_{2 \mathrm{i}}+\mathrm{X}_{20} \\
& =\sum_{\mathrm{i}=1 \mathrm{j}=1}^{\mathrm{t}=1} \epsilon_{1 \mathrm{j}}+\sum_{\mathrm{i}=1}^{\mathrm{t}} \epsilon_{2 \mathrm{i}}+\mathrm{t} \mathrm{X}_{10}+\mathrm{X}_{20}
\end{aligned}
$$

Thus $X_{2 t}$ and hence $X_{t}$ are not $I(1)$ processes since even $\Delta X_{t}$ is non-stationary.
This example shows that even simple autoregressive porcesses with 1 lag can generate a process which meeds 2 differences to become stationary. Thus we need a theorem that gives precise conditions for an autoregressive rpprocess to be an $\mathrm{I}(1)$ process.

We want to formulate this result for a multivariate process. As usual the properties of the matrix polynomial

$$
\Pi(z)=I-\Pi_{1} z-\ldots-\Pi_{k^{z}} z^{k}
$$

determine the properties of the process, and the first well known result see Anderson (1971) is given here for completeness. We let $|\Pi(z)|$ denote the determinant of $\Pi(z)$.

Theorem 1. If $X_{t}$ is given by (2.1) and if $|\Pi(z)|=0$ implies that $|z|>1$, then $X_{t}$ can be given an initial distribution such that it becomes stationary. In this case $X_{t}$ has the representation

$$
X_{t}=\sum_{i=0}^{\infty} C_{i} \epsilon_{t-i^{\prime}}
$$

where the coefficients are given by $C(z)=\sum_{i=0}^{\infty} C_{i} z^{i}=\Pi(z)^{-1},|z|<1+\delta$ for some $\delta>$ 0.

This result shows that if $|\Pi(\mathrm{z})|$ has all roots outside the unit disk then the process generated by (2.1) is stationary. Thus we have to allow other roots of $|\Pi(z)|$ for $X_{t}$ to be non-stationary. In order to formulate a general result we expand the polynomial $|\Pi(z)|$ around $z=1$ :

$$
\Pi(z)=-\Pi+(1-z) \Psi+(1-z)^{2} \Pi_{2}(z)
$$

The matrix $\Pi=-\Pi(1)=-I+\sum_{i=1}^{k} \Pi_{i}$ has the interpretation as the "total impact" matrix and $\Psi=-\mathrm{d} \Pi(\mathrm{z}) /\left.\mathrm{dz}\right|_{\mathrm{z}=1}=\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{i} \Pi_{\mathrm{i}}$ is the "mean lag" matrix. Finally for any ( $\mathrm{p} \times \mathrm{r}$ ) ( $\mathrm{r}<\mathrm{p}$ ) matrix $\alpha$ of full rank we let $\alpha_{\perp}$ denote a ( $\mathrm{p} \times(\mathrm{p}-\mathrm{r})$ ) matrix of full rank such that $\alpha^{\prime} \alpha_{\perp}=0$. We can then formulate Granger's representation theorem

Theorem 2. If $X_{t}$ satisfies (2.1), and if

$$
\begin{align*}
& |\Pi(z)|=0 \text { implies that }|z|>1 \text { or } z=1  \tag{2.2}\\
& \Pi=\alpha \beta^{\prime} \quad(\alpha, \beta p \times r \text { of full rank } r<p)  \tag{2.3}\\
& \alpha_{\perp}^{\prime} \Psi \beta_{\perp} \quad \text { full rank, } \tag{2.4}
\end{align*}
$$

then $\Delta X_{t}$ and $\beta^{\prime} X_{t}$ can be given initial distributions such that $X_{t}$ is $I(1)$, and cointegrated with cointegrating relations $\beta$, i.e. $X_{t}$ is non-stationary, $\Delta X_{t}$ is stationary and $\beta^{\prime} X_{t}$ is stationary.

Further the process $X_{t}$ can be given the representation:

$$
\begin{equation*}
X_{t}=\beta_{\perp}\left(\alpha_{\perp}^{\prime} \Psi \beta_{\perp}\right)^{-1} \alpha_{\perp}^{\prime} \sum_{i=1}^{t} \epsilon_{i}+C_{1}(L) \epsilon_{t}+P_{\beta_{\perp}} X_{0^{\prime}} \tag{2.5}
\end{equation*}
$$

and satisfies the reduced form error correction equation

$$
\begin{equation*}
\Delta X_{t}=\alpha \beta^{\prime} X_{t-1}+\sum_{i=1}^{k-1} \Gamma_{i} \Delta X_{t-i}+\epsilon_{t^{\prime}} t=1, \ldots, T \tag{2.6}
\end{equation*}
$$

Example 2 continued
We find in this case the matrix $\Pi$ and $\Psi$ to be

$$
\Pi=\left[\begin{array}{cc}
-\alpha_{1} & 2 \alpha_{1} \\
0 & 0
\end{array}\right]=\left[\begin{array}{c}
-\alpha_{1} \\
0
\end{array}\right](1,-2), \Psi=\left[\begin{array}{cc}
1-\alpha_{1} & 2 \alpha_{1} \\
0 & 1
\end{array}\right]
$$

so that

$$
\Pi(z)=\left[\begin{array}{cc}
\alpha_{1} & -2 \alpha_{1} \\
0 & 0
\end{array}\right]+(1-z)\left[\begin{array}{cc}
1-\alpha_{1} & 2 \alpha_{1} \\
0 & 1
\end{array}\right]
$$

with determinant

$$
|\Pi(z)|=(1-z)\left(1+z \alpha_{1}-z\right)
$$

and roots $\mathrm{z}=1$ and $\mathrm{z}=\left(1-\alpha_{1}\right)^{-1}, \alpha_{1} \neq 1$. Thus condistion (2.2) is satisfied for $0<$ $\alpha_{1}<2$. We see that $\Pi$ has reduced rank and $\alpha^{\prime}=\left(-\alpha_{1}, 0\right)$ and $\beta^{\prime}=(1,-2)$. It is seen that $\alpha_{\perp}^{\prime}=(0,1)$ and $\beta_{\perp}^{\prime}=(2,1)$ so that $\alpha_{\perp}^{\prime} \Psi \beta_{\perp}=1$, so that condition (2.4) is satisfied and the process is $\mathrm{I}(1)$.

## Example 4 continued

For this example we find the coefficients

$$
\Pi_{1}=\left[\begin{array}{ll}
1 & 0 \\
1 & 1
\end{array}\right], \Pi=\left[\begin{array}{ll}
0 & 0 \\
1 & 0
\end{array}\right] \Psi=\left[\begin{array}{ll}
1 & 0 \\
1 & 1
\end{array}\right]
$$

which shows that

$$
\Pi(z)=\left[\begin{array}{cc}
1-z & 0 \\
-z & 1-z
\end{array}\right],|\Pi(z)|=(1-z)^{2} .
$$

Thus $z=1$ is a double root, condition (2.2) is satisfied and

$$
\alpha=\left[\begin{array}{l}
0 \\
1
\end{array}\right], \beta=\left[\begin{array}{l}
1 \\
0
\end{array}\right], \alpha_{\perp}=\left[\begin{array}{l}
1 \\
0
\end{array}\right], \beta_{\perp}=\left[\begin{array}{l}
0 \\
1
\end{array}\right]
$$

with the result that

$$
\alpha_{\perp}^{\prime} \Psi \beta_{\perp}=0
$$

Hence for this example condition (2.4) breaks down with the result that the process $\mathrm{X}_{\mathrm{t}}$ is not $\mathrm{I}(1)$.

For a univariate series condition (2.3) says that the sum of the coefficients to the lagged levels add up to 1 , implying that the characteristic polynomial has a unit root, and condition (2.4) then says that the derivative of the polynomial is different from zero at $\mathrm{z}=1$ so that the process has only one unit root. This condition is needed
to make sure that the process has to differenced only once to become stationary.
The parametric restrictions (2.2), (2.3) and (2.4) determine exactly when the autoregressive equations define an $\mathrm{I}(1)$ process that allows for cointegrating relations, and (2.5) gives the representation in terms of common trends or as a random walk plus a stationary process. The equations can be written in error correction form (2.6) and Granger's representation then shows that the common trends formulation and the error correction formulation are equivalent. Condition (2.4) guarantees that the number of roots of $|\Pi(z)|=0$ at $z=1$ equals the rank deficiency of $\Pi$, i.e. $p-r$. The proof of Theorem 2 is given in Johansen (1991).

We can now define a parametric statistical model, the error correction model, given by the vector autoregressive model (2.1) with the restriction (2.3), or rewritten as (2.6), which we want to use for describing the statistical variation of the data. The parameters are $\left(\alpha, \beta, \Gamma_{1}, \ldots, \Gamma_{\mathrm{k}-1}, \Omega\right)$ which vary freely. We thus express in parametric form the hypothesis of cointegration, and hypotheses of interest can be formulated as parametric restrictions on the cointegrating relations. What remains is, in connection with applications, to see which hypotheses could be of interest and then to analyze the model in order to find estimators and test statistics, and describe their (asymptotic) distributions.

## 3. Purchasing power parity, an illustrative example

The law of one price states that if the same quantity of the same commodity is purchased in two different countries at prices $P_{1}$ and $P_{2}$ respectively then

$$
\begin{equation*}
\mathrm{P}_{1}=\mathrm{P}_{2} \mathrm{E}_{12} \tag{3.1}
\end{equation*}
$$

where $\mathrm{E}_{12}$ is the exchange rate between the two currencies. It is by no means clear that the same relation will be found if two price indices are compared with the exchange rate. This is due to the definition of the price index, which could vary between countries, and the different pattern of consumption in the different countries.

Still if this relation is not satisfied approximately there will be pressures on the economy either for changing the price levels or for changing the exchange rate.

It is therefore of interest to see if such relations hold or to look for the so called purchasing power parity (PPP).

The data we analyze is kindly supplied by Ken Wallis and analysed in Johansen and Juselius (1992). It consists of quarterly observations from 1972.1 to 1987.3 for the UK wholesale price index $\left(\mathrm{p}_{1}\right)$ compared to a trade weighted foreign price index $\left(\mathrm{p}_{2}\right)$ and the UK effective exchange rate $\left(\mathrm{e}_{12}\right)$. These variables are measured in logs. Also included in the analysis are the three months treasury bill rate in UK ( $\mathrm{i}_{1}$ ) and the three months Eurodollar interest rate $\left(\mathrm{i}_{2}\right)$. The reason for including the interest variables is that one would expect that the interest rates are related to the exchange rates through

$$
\begin{equation*}
\mathrm{i}_{1 \mathrm{t}}-\mathrm{i}_{2 \mathrm{t}}=\Delta \mathrm{e}_{12 . \mathrm{t}+1} \tag{3.2}
\end{equation*}
$$

if there are no restrictions in the movement of capital between countries.
Inspection of the plots of the time series shows immediately that we have processes that are not stationary. We fit an autoregressive model with 2 lags and allow for seasonal dummies and a constant term in model (2.1). In order to find a reasonable description of the data we also included the world oil price and treated it as given for the present analysis. This gives added complications in the analysis, but we shall not go into these in this presentation but refer to Johansen and Juselius (1992) for a full discussion of the application. The residuals show no systematic deviation from independent Gaussian variables and we continue with the analysis of model (2.1).

The economic relation (3.1) expressed in logs is the PPP relation

$$
\begin{equation*}
\mathrm{p}_{1}-\mathrm{p}_{2}-\mathrm{e}_{12}=0 \tag{3.3}
\end{equation*}
$$

This equation is clearly not satisfied by the data. Granger's formulation is that in general a linear combination of variables will be non-stationary, just like individual variables, but one would like to find the most "stable" or stationary ones and identify
them as the interesting economic relations like (3.3), and the error correction idea is that changes in prices and exchange rates are influenced by disequilibrium errors like $p_{1 t}-p_{2 t}-e_{12 t}$ through adjustment coefficients.

In this formulation the PPP relation is identified with the vector ( $1,-1,-1,0,0$ ) and one can test the hypothesis that it is a cointegrating vector, or the slightly weaker hypothesis that there exist cointegrating relations between the variables $\mathrm{p}_{1}, \mathrm{p}_{2}$, and $\mathrm{e}_{12}$, that is, there exists a vector of the form ( $\mathrm{a}, \mathrm{b}, \mathrm{c}, 0,0$ ) which is a cointegration vector. One can also ask the even simpler question if ( $\mathrm{p}_{1}, \mathrm{p}_{2}, \mathrm{e}_{12}, \mathrm{i}_{1}, \mathrm{i}_{2}$ ) exhibit cointegration in which case one would like to know how many cointegrating relations there are. Our beliefs are that $\mathrm{r}=2$ as given by (3.2) and (3.3). Once the number of cointegration vectors is determined, a natural hypothesis is that relevant economic relations should only depend on the ratio of the prices, or in other words the coefficients to $\mathrm{p}_{1}$ and $\mathrm{p}_{2}$ should add to zero. The next section contains a discussion of a number of hypotheses that can be formulated on cointegrating relations.

## 4. Formulation of the reduced form error correction model and various hypotheses on the cointegrating relations

If model (2.1) describes an $I(1)$ process having cointegration we should restrict the parameters as given by (2.2), (2.3) and (2.4). The restriction in (2.2), that the roots are outside the unit disk, is very difficult to handle analytically. Fortunately it rarely turns out that the roots are inside the unit disk, and if they are, it is more important to know where they are than to force them to the boundary of the unit disk. Hence we disregard that part of condition (2.2) in the statistical calculations, but check that it is satisfied by the estimates. Condition (2.4) is easily satisfied, since matrices with full rank are dense in the space of all matrices, thus the estimator derived without the restriction that $\alpha_{\perp}^{\prime} \Psi \beta_{\perp}$ has full rank will automatically have full rank with probability one.

Definition 3 The reduced form error correction model is described by the equations

$$
\begin{equation*}
\Delta X_{t}=\alpha \beta^{\prime} X_{t-1}+\Sigma_{1}^{k-1} \Gamma_{i} \Delta X_{t-1}+\epsilon_{t^{\prime}} \tag{4.1}
\end{equation*}
$$

where $\epsilon_{1}, \ldots, \epsilon_{T}$ are independent Gaussian $N_{p}(0, \Omega)$ and the parameters $\left(\alpha, \beta, \Gamma_{1}, \ldots, \Gamma_{k-1}, \Omega\right)$ are freely varying.

The condition that $\Pi=\alpha \beta^{\prime}$ is sometimes referred to as that of imposing ( $\mathrm{p}-\mathrm{r}$ ) unit roots, but we shall think of it as a parametric representation of the existence of (at most) r cointegrating relations. Thus the model is a submodel of the general autoregressive model defined by the reduced rank hypothesis on the coefficient matrix $\Pi$ of the levels.

The above allows one to formulate a nested sequence of hypotheses

$$
\mathrm{H}_{0} \subset \ldots \subset \mathrm{H}_{\mathrm{r}} \subset \ldots \subset \mathrm{H}_{\mathrm{p}}
$$

and the test of $\mathrm{H}_{\mathrm{r}}$ in $\mathrm{H}_{\mathrm{p}}$ is then the test that there are (at most) r cointegrating relations. Thus $H_{0}$ is just a vector autoregressive model for $X_{t}$ in differences and $H_{p}$ the unrestricted autoregressive model for $\mathrm{X}_{\mathrm{t}}$ in levels, and the models in between $\mathrm{H}_{1}, \ldots, \mathrm{H}_{\mathrm{p}-1}$ give the possibility to exploit the information in the reduced rank matrix П. A standard way of analyzing non-stationary processes is to difference them sufficiently to obtain stationarity and then analyze the differences by an autoregressive model. Note that this model is just $\mathrm{H}_{0}$, the adequacy of which can be tested if we start with the general model $H_{p}$

Note that in model $\mathrm{H}_{\mathrm{r}}$ the parameters $\alpha$ and $\beta$ are not identified since $\Pi=$ $\alpha \beta^{\prime}=\alpha \xi^{-1}\left(\beta \xi^{\prime}\right)^{\prime}$ for any rxr matrix $\xi$ of full rank, but that one can estimate the space spanned by $\alpha$ and $\beta$ respectively.

Thus cointegration analysis is formulated as the problem of making inference on the cointegration space, $\operatorname{sp}(\beta)$, and the adjustment $\operatorname{space}, \operatorname{sp}(\alpha)$, and hypotheses in
the following will be expressed as restrictions on these.
Once the cointegrating rank has been determined we can test hypotheses about the coefficients $\alpha$ and $\beta$, and we next give examples of such hypotheses.

The hypothesis that only real prices enter the cointegrating relations, can be expressed as the hypothesis that the coefficients to $\mathrm{p}_{1}$ and $\mathrm{p}_{2}$ sum to zero, or $(1,1,0,0,0) \beta=0$. This is the same restriction on all cointegrating relations

$$
\begin{equation*}
\beta=\mathrm{H} \varphi \tag{4.2}
\end{equation*}
$$

where $\mathrm{H}=(1,1,0,0,0)_{\perp}^{\prime}$ is known and $\varphi(1 \times r)$ is unknown. This hypothesis on $\beta$ does not depend on $\beta$ being identified uniquely, since it is the same set of restriction on all the relations. If $\beta$ satisfies (4.2) then so does $\beta \xi$ for any matrix $\xi(\mathrm{rxr})$. Hence (4.2) is a testable hypothesis on the cointegrating space.

The hypothesis that some cointegration vectors like ( $1,-1,-1,0,0$ ) and ( $0,0,0,1,-1$ ) are known can be formulated as

$$
\begin{equation*}
\beta=(\mathrm{b}, \varphi), \tag{4.3}
\end{equation*}
$$

where $\mathrm{b}\left(\mathrm{p} \times \mathrm{r}_{1}\right)$ is known and $\varphi\left(\mathrm{p} \times \mathrm{r}_{2}\right)$ is unknown, $\mathrm{r}_{1}+\mathrm{r}_{2}=\mathrm{r}$. In particular it means that the test that an individual variable is stationary can be expressed in the form (4.3) for b equal to a unit vector.

A more general hypothesis can, for $\mathrm{r}=2$ say, be formulated as

$$
\begin{equation*}
\beta=\left(\mathrm{H}_{1} \varphi_{1}, \mathrm{H}_{2} \varphi_{2}\right) \tag{4.4}
\end{equation*}
$$

where $H_{i}\left(\mathrm{p}_{\mathrm{s}} \mathrm{s}_{\mathrm{i}}\right)$ are known and $\varphi_{\mathrm{i}}\left(\mathrm{s}_{\mathrm{i}} \times \mathrm{r}_{\mathrm{i}}\right)$ are unknown, and $\mathrm{r}_{1}+\mathrm{r}_{2}=\mathrm{r}$, see Johansen and Juselius (1993).

Example $5 \quad$ An example of (4.4) is given by the hypothesis that $\mathrm{p}_{1}, \mathrm{p}_{2}$ and $e_{12}$ cointegrate and that the interest rates cointegrate, see section 3. In this case we are looking for two relations of the form ( $\mathrm{a}, \mathrm{b}, \mathrm{c}, 0,0$ ) and ( $0,0,0, \mathrm{~d}, \mathrm{e}$ ), which clearly form a set of uniquely identified equations. The hypothesis has the form (4.4) with

$$
H_{1}=\left[\begin{array}{lll}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{array}\right], H_{2}=\left[\begin{array}{ll}
0 & 0 \\
0 & 0 \\
0 & 0 \\
1 & 0 \\
0 & 1
\end{array}\right] .
$$

Thus we are, in the econometric language, testing for the overidentifying restriction that there is a cointegrating relation between the variables that has two zeros as coefficients to the interest rates. These hypotheses are hypotheses on the cointegrating space and thus do not depend on how the cointegrating relations are identified.

Often one want to estimate structural equations, rather than just the cointegrating space, which is more difficult to interpret. Thus we have a number, 2 say, of meaningful economic relations that satisfy linear restrictions like exclusion restrictions. One should then check the rank condition and if it is satisfied (in a generic) sense one can then go on to formulate the hypothesis as (4.4), with $H_{i}=R_{i_{\perp}}$, i $=1,2$. The problem of identification in this formulaton is discussed by Johansen and Juselius (1993).

The economic insight is used in formulating the problem of interest, and therefore in the choice of variables, as well as in the discussion of which economic relations we expect to find. The statistical model is then used as a description of the non-stationary statistical variation of the data. The cointegration relations are used as a tool for discussing the existence of long-run economic relations and the various hypotheses are then tested in view of the statistical variation of the data. The interpretation of the cointegrating relations also require a thorough understanding of the underlying economic problems.

## 5. Estimation of cointegrating relations and calculation of test statistics

This section contains a brief description of the OLS solution to the estimation problem and then discuss how the estimation problem of the various hypotheses from section 4
can be solved by analysing the Gaussian likelihood function.
A time honored procedure for finding linear relations between two variables $Y_{t}$ and $X_{t}$ is to regress $Y_{t}$ on $X_{t}$ and then to discuss the properties of the estimator, $\hat{\beta}_{\text {ols }}$, under various assumptions on the processes. This was of course the first to be used by Engle and Granger (1987) in their fundamental paper. The problem with the analysis is that since the regressor $\mathrm{X}_{\mathrm{t}}$ in general is a non-stationary process the usual simple asymptotic normality does not hold for the estimator.

Stock (1987) proved the, at first sight, rather surprising result that one gets a superconsistent estimator in the sense that

$$
\mathrm{T}^{1-\delta}\left(\hat{\beta}_{\mathrm{ols}}-\beta \stackrel{\mathrm{P}}{\rightarrow} \stackrel{0}{\rightarrow} \delta>0\right.
$$

under the assumption that the regressor is an $I(1)$ process, and that $\left(Y_{t}, X_{t}\right)$ cointegrate with cointegrating vector $(1,-\beta)$, i.e. $\mathrm{Y}_{\mathrm{t}}-\beta \mathrm{X}_{\mathrm{t}}$ is stationary.

Behind this result is the following very simple idea: In the simple regression model

$$
\mathrm{Y}_{\mathrm{t}}=\beta \mathrm{X}_{\mathrm{t}}+\epsilon_{\mathrm{t}}
$$

where $\epsilon_{\mathrm{t}}$ are independent Gaussian variables with mean zero and variance $\sigma^{2}$, and the X 's are deterministic one finds that $\hat{\beta}_{\text {ols }}$ is Gaussian with mean $\beta$ and variance $\sigma^{2} / \Sigma_{1}^{\mathrm{T}} \mathrm{X}_{\mathrm{t}}^{2}$. If the X 's are bounded away from zero and infinity then the sum will increase like T and usual asymptotics holds in the sense that

$$
\mathrm{T}^{\frac{1}{2}}(\hat{\beta}-\beta)
$$

is asymptotically Gaussian. For $I(1)$ processes, however, it holds that $X_{t}$ behaves asymptotically like a random walk, see (2.5), and then

$$
\begin{equation*}
\mathrm{T}^{-2} \Sigma_{1}^{\mathrm{T}} \mathrm{X}_{\mathrm{t}}^{2} \xrightarrow{\mathrm{w}} \int_{0}^{1} \mathrm{~W}(\mathrm{u})^{2} \mathrm{du} \tag{5.1}
\end{equation*}
$$

where $\mathrm{W}(\mathrm{u})$ is a Brownian motion given by the limit of $\mathrm{T}^{-\frac{1}{2}} \mathrm{X}_{[\mathrm{Tu}]}$. The faster rate ( $\mathrm{T}^{2}$ ) implies that one gets superconsistency of the regression estimator and the random limit implies that the limiting distribution is not Gaussian, but a rather complicated
mixed Gaussian distribution involving Brownian motion and nuisance parameters. Inference for the remaining parameters $\vartheta=\left(\Gamma_{1}, \ldots, \Gamma_{\mathrm{k}-1}, \alpha, \Omega\right)$ is relatively simple since superconsistency of the estimate for $\beta$ implies that inference on $\vartheta$ can be conducted as if $\beta$ were known, in which case model (4.1) only involves the stationary observables $\beta^{\prime} \mathrm{X}_{\mathrm{t}-1}$ and the differences of $\mathrm{X}_{\mathrm{t}}$.

This type of result has created a very large literature, see Stock and Watson (1988) for the estimation of the cointegrating rank and the cointegrating relations, Chan and Wei (1988) for inference in unstable processes and the work of Phillips and his coworkers, on how to do regression with integrated regressors Phillips $(1987,1991)$, Phillips and Durlauf (1986), Phillips and Ouliaris (1990), Park and Phillips $(1988,1989)$. It has lead to a new class of limit distributions, which are combinations of mixed Gaussian and the so call unit root distributions. This type of problem has also been taken up be Jeganathan (1992).

The negative aspects of the findings of these authors is that the limiting distribution for the regression estimator is very complicated and this make inference and hypothesis testing difficult. There are ways of eliminating the nuisance parameters by modifying the regression method, see Park (1992) and Phillips and Hansen (1990). Another way of modifyig ordinary least squares is to analyse the Gaussian likelihood function and use that as a tool for generating estimators of the various hypotheses investigated in section 4 . One would expect that if any estimator would have a simple limit distribution is would have to be the maximum likelihood estimator. Similarly one would expect that the likelihood ratio test statistic has a simple (asymptotic) distribution. It is therefore that section 4 contains a fairly detailed discussion of the model and the hypotheses, and we now turn to likelihood based inference for the cointegration model.

Model (4.1) gives rise to a reduced rank regression and the solution is available as an eigenvalue problem. It was solved by Anderson (1951) in the regression
context and runs as follows:
First we eliminate the parameters $\Gamma_{1}, \ldots, \Gamma_{\mathrm{k}-1}$ by regressing $\Delta \mathrm{X}_{\mathrm{t}}$ and $\mathrm{X}_{\mathrm{t}-1}$ on $\Delta \mathrm{X}_{\mathrm{t}-1}, \ldots, \Delta \mathrm{X}_{\mathrm{t}-\mathrm{k}+1}$. The residuals are $\mathrm{R}_{0 \mathrm{t}}$ and $\mathrm{R}_{1 \mathrm{t}}$ respectively. Next form the sum of squares and products

$$
\mathrm{S}_{\mathrm{ij}}=\mathrm{T}^{-1} \sum_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{R}_{\mathrm{it}} \mathrm{R}_{\mathrm{jt}}^{\prime}, \quad \mathrm{i}, \mathrm{j}=0,1
$$

and solve the eigenvalue problem

$$
\begin{equation*}
\left|\lambda S_{11}-S_{10} S_{00}^{-1} S_{01}\right|=0 \tag{5.2}
\end{equation*}
$$

for eigenvalues $\hat{\lambda}_{1}>\ldots>\hat{\lambda}_{p}$ and eigenvectors $\hat{V}=\left(\hat{\mathrm{v}}_{1}, \ldots, \hat{\mathrm{v}}_{\mathrm{p}}\right)$, that is,

$$
\lambda_{i} S_{11} v_{i}=S_{10} S_{00}^{-1} S_{01} v_{i}, i=1, \ldots, p
$$

and $\hat{V}^{\prime} \mathrm{S}_{11} \hat{\mathrm{~V}}=\mathrm{I}$.
A maximum likelihood estimator for $\beta$ is given by

$$
\begin{equation*}
\hat{\beta}=\left(\hat{\mathrm{v}}_{1}, \ldots, \hat{\mathrm{v}}_{\mathrm{r}}\right) \tag{5.3}
\end{equation*}
$$

An estimator for $\alpha$ is then

$$
\hat{\alpha}=S_{01} \hat{\beta},
$$

and the maximized likelihood function is given by

$$
\begin{equation*}
\mathrm{L}_{\max }^{-2 / \mathrm{T}}=\left|\mathrm{S}_{00}\right| \stackrel{\mathrm{H}}{\mathrm{i}=1}_{\mathrm{r}}\left(1-\hat{\lambda}_{\mathrm{i}}\right) \tag{5.4}
\end{equation*}
$$

see Johansen (1988) and Johansen and Juselius (1990) for details and applications. One can interpret $\hat{\lambda}_{\mathrm{i}}$ as the squared canonical correlation between $\Delta \mathrm{X}_{\mathrm{t}}$ and $\mathrm{X}_{\mathrm{t}-1}$ conditional on $\Delta \mathrm{X}_{\mathrm{t}-1}, \ldots, \Delta \mathrm{X}_{\mathrm{t}-\mathrm{k}+1}$. Thus the estimate of the "most stable" relations between the levels are those that correlate most with the stationary process $\Delta \mathrm{X}_{\mathrm{t}}$ corrected for its lags.

Since only $\operatorname{sp}(\beta)$ is identifiable without further restrictions, one really estimates the cointegration space as the space spanned by the first r eigenvectors. This is seen by the fact that if $\hat{\beta}$ is given by (5.3) then $\hat{\beta} \xi$ is also maximizing the likelihood function for any choice of $\xi(\mathrm{r} \times r)$ of full rank.

This solution provides the answer to the estimation of the models $H_{r}, r=$
$0, \ldots, p$ and by comparing likelihoods one can test $H_{r}$ in $H_{p}$, i.e test for $r$ cointegration relations, by the statistic

$$
\begin{equation*}
-2 \ln \mathrm{Q}(\mathrm{r} \mid \mathrm{p})=-\mathrm{T} \sum_{\mathrm{i}=\mathrm{r}+1}^{\mathrm{p}} \ln \left(1-\hat{\lambda}_{\mathrm{i}}\right) \tag{5.5}
\end{equation*}
$$

The above estimator (5.3) is an estimator of all cointegrating relations and it is sometimes convenient to normalize (or identify) the vectors by choosing a specific coordinate system in which to express the variables in order to facilitate the interpretation and in order to be able to give an estimate of the variability of the estimators.

If c is any pxr matrix, such that $\beta^{\prime} \mathrm{c}$ has full rank, one can normalize $\beta$ as

$$
\beta_{\mathrm{c}}=\beta\left(\mathrm{c}^{\prime} \beta\right)^{-1}
$$

which satisfies $c^{\prime} \beta_{c}=I$ provided that $\left|c^{\prime} \beta\right| \neq 0$. A particular example is given by $c^{\prime}$ $=(\mathrm{I}, 0)$ and $\beta^{\prime}=\left(\beta_{1}, \beta_{2}\right)$ in which case $\beta^{\prime} \mathrm{c}=\beta_{1}$ and $\beta_{\mathrm{c}}^{\prime}=\left(\mathrm{I}, \beta_{1}^{-1} \beta_{2}\right)$ which corresponds to solving the cointegrating relations for the first r variables, if the coefficient matrix of these $\left(\beta_{1}\right)$ has full rank.

The maximum likelihood estimator of $\beta_{c}$ is then

$$
\hat{\beta}_{c}=\hat{\beta}\left(c^{\prime} \hat{\beta}\right)^{-1}
$$

The hypotheses (4.2) and (4.3) are easily analyzed the same way. If (4.2) holds then $\alpha \beta^{\prime} \mathrm{X}_{\mathrm{t}}=\alpha \varphi^{\prime} \mathrm{H}^{\prime} \mathrm{X}_{\mathrm{t}}$, which shows that the cointegration relations are found in $\mathrm{sp}(\mathrm{H})$ by solving the eigenvalue problem

$$
\begin{equation*}
\left|\lambda H^{\prime} \mathrm{S}_{11} \mathrm{H}-\mathrm{H}^{\prime} \mathrm{S}_{10} \mathrm{~S}_{00}^{-1} \mathrm{~S}_{01} \mathrm{H}\right|=0 \tag{5.6}
\end{equation*}
$$

Under hypothesis (4.3) there are some known cointegration relations and $\alpha \beta^{\prime} \mathrm{X}_{\mathrm{t}}=$ $\alpha_{1} b^{\prime} X_{t}+\alpha_{2} \varphi^{\prime} X_{t}$, which shows that the coefficients to the observable $b^{\prime} X_{t-1}$ can be eliminated together with the parameters $\left(\Gamma_{1}, \ldots, \Gamma_{k-1}\right)$ such that the eigenvalue problem, that has to be solved, is

$$
\begin{equation*}
\left|\lambda S_{11 . b}-S_{10 . b} S_{00 . b}^{-1} S_{01 . b}\right|=0 \tag{5.7}
\end{equation*}
$$

where

$$
S_{i j . b}=S_{i j}-S_{i 1} b\left(b^{\prime} S_{11} b\right)^{-1} b^{\prime} S_{1 j}
$$

The maximal value of the likelihood function is given by expressions similar to (5.4) and the test of hypotheses (4.2) and (4.3) then consists of comparing the r largest eigenvalues under the various restrictions.

The hypothesis (4.4) is slightly more complicated, but can be solved by a switching algorithm, where each step involves an eigenvalue problem, see Johansen and Juselius (1993). The problem of estimating uniquely identified relations admits a simple solution in the likelihood framework.

Thus it is seen that a number of interesting hypotheses can be solved provided one has an eigenvalue routine and that one can perform the basic operations on a covariance matrix, namely that of marginalization (transformation) and conditioning. We have programs written in RATS which perform these analyses, and programs exist in SAS, AREMOS and GAUSS.

## 6. The empirical example continued

For the example of section 3 we find by solving (5.2) the results in Table 1
TABLE 1
Eigenvectors and eigenvalues for the UK data in section 3.

| $\hat{\lambda}$ | .401 | .285 | .254 | .102 | .083 |
| :--- | :---: | :---: | :---: | :---: | :---: |
| $\mathrm{p}_{1}$ | -16.64 | -1.68 | 4.71 | 9.94 | -9.93 |
| $\mathrm{p}_{2}$ | 15.12 | 1.92 | -5.99 | -23.84 | 14.42 |
| $\mathrm{e}_{12}$ | 15.51 | 5.65 | 5.24 | 11.15 | 4.77 |
| $\mathrm{i}_{1}$ | 56.14 | -59.17 | 12.93 | -4.06 | -22.61 |
| $\mathrm{i}_{2}$ | 31.45 | 55.27 | -13.34 | 29.67 | -7.57 |

Note that the first eigenvector has coefficients for $\mathrm{p}_{1}, \mathrm{p}_{2}$, and $\mathrm{e}_{12}$ which looks like the PPP relation, corresponding to the relation (3.3), whereas the two next have equal coefficient with opposite sign for the interest rates corresponding to the interest rate differential (3.2). The challenge in the analysis lies in the interpretation of the eigenvectors corresponding to the largest eigenvalues. We often find that they have an immediate interpretation, but one should remember, that what one estimates is the cointegration space spanned by the eigenvectors. Thus one should sometimes "rotate" or take linear combinations of the vectors to find out what they mean, or still better re-estimate them under (over) identifying restrictions as expressed in (4.4).

We discuss the asymptotic distribution of tests statistics and estimators in the next section, but continue here with a brief description of the findings in the data.

The first question is how many cointegration vectors that are consistent with the data. Table 2 contains the test statistics and fractiles for the hypotheses $\mathrm{H}_{0}, \ldots, \mathrm{H}_{5}$, see (5.5).

TABLE 2
Likelihood ratio test statistics for testing the number of cointegration relations

| $\hat{\lambda}$ | r | $-\mathrm{T} \Sigma_{\mathrm{r}+1}^{5} \ln \left(1-\hat{\lambda}_{\mathrm{i}}\right)$ | $95 \%$ fractile |
| :---: | :---: | :---: | :---: |
| .083 | 4 | 5.19 | 8.08 |
| .102 | 3 | 11.66 | 17.84 |
| .254 | 2 | 29.26 | 31.26 |
| .285 | 1 | 49.42 | 48.41 |
| .401 | 0 | 80.77 | 69.98 |

It is seen from Table 2 that the hypothesis $H_{0}$ which gives a test statistic of 80.77 corresponds to an extreme observation in the asymptotic distribution of the test statistic, but that the test statistic for $\mathrm{H}_{1}$ and $\mathrm{H}_{2}$ linger around the $95 \%$ fractile. Note that $\hat{\lambda}_{2}$ and $\hat{\lambda}_{3}$ are almost identical.

In the further investigation of the data we took $\mathrm{r}=2$, and performed various tests of the type mentioned above. In particular we found that the hypothesis that both cointegration vectors have the form ( $a,-\mathrm{a},-\mathrm{a}, \mathrm{b}, \mathrm{c}$ ) was easily accepted by the data. This is of the type (4.2) of a common restriction on both cointegration relations. When we investigated if the vector $(1,-1,-1,0,0)$ was a cointegration vector it was strongly rejected by the data. Thus our main relation (3.3), the PPP relation, does not hold, but the stationarity can be achieved by including the interest rates in the relation. It turns out, however, that the vector $(0,0,0,1,-1)$ is a cointegrating relation, such that the interest rate differential is stationary, in accordance with equation (3.2). Finally one can ask if $\mathrm{p}_{1}, \mathrm{p}_{2}$ and $\mathrm{e}_{12}$ are cointegrated. This hypothesis is of the form (4.4) and was found not to be supported by the data.

Our understanding of the two-dimensional cointegration space is then the following: One of the vectors is just ( $0,0,0,1,-1$ ) indicating that the interest rate differential is stationary. If we choose as the other vector the PPP relation $(1,-1,-1,0,0$,$) then it is inconsistent with the data, but we have to include the interest$ rates and can choose a vector of the form ( $1,-1,-1, a, b)$, which describes a modified PPP-relation. Note that in the system consisting of the relations ( $1,-1,-1, \mathrm{a}, \mathrm{b}$ ) and $(0,0,0,1,-1)$ the first is not identified, but if we put either a or $b$ to zero we get a uniquely identified system of cointegrating relations. This clearly does not conclude the economic analysis of the data, but the findings are useful for formulating an empirical economic model that describes this small set of variables. Thus the methods presented here are meant only as a tool for investigating long-run relations in the economy.

It is also of interest to discuss the estimates of the adjustment coefficients $\alpha$, since they show how the various variables change with past disequilibrium errors, like the (modified) PPP relation. We found in the present data that the changes in foreign prices were not influenced by the disequilibrium errors, that is, the $\alpha$-coefficient was zero. This has the consequence that the whole analysis can be performed conditionally on the foreign prices, or differently expressed: The foreign prices were weakly exogenous for the cointegrating relations, see Johansen (1992).

## 7. Asymptotic theory

Section 7 contains a brief description of the asymptotic theory of test statistics and estimators, as well as a discussion of how the results can be applied to conduct inference about the cointegrating rank and the cointegrating vectors.

The reason that inference for non-stationary processes is interesting and that so many people work on it now, is that it is non-standard, in the sense that estimators are not asymptotically Gaussian and test statistics are not asymptotically $\chi^{2}$. This was systematically explored by Dickey and Fuller, see Fuller (1976), in testing for unit roots in univariate processes.

Consider the simple model of an autoregressive process of order 1

$$
\mathrm{Y}_{\mathrm{t}}=\rho \mathrm{Y}_{\mathrm{t}-1}+\epsilon_{\mathrm{t}}
$$

where $\epsilon_{\mathrm{t}}$ are independent Gaussian variables with mean zero and variance $\sigma^{2}$. The hypothesis of interest is

$$
\rho=1,
$$

and it is seen that it means that $\mathrm{Y}_{\mathrm{t}}$ is a random walk, i.e a non-stationary process. They found among other results that under the null of $\rho=1$, it holds that

$$
\mathrm{T}(\hat{\rho}-1)=\mathrm{T}^{-1} \Sigma_{1}^{\mathrm{T}} \epsilon_{\mathrm{t}} \mathrm{Y}_{\mathrm{t}-1} / \mathrm{T}^{-2} \Sigma_{1}^{\mathrm{T}} \mathrm{Y}_{\mathrm{t}-1}^{2} \xrightarrow{\mathrm{w}} \underset{0}{1} \mathrm{WdW} / \int_{0}^{1} \mathrm{~W}^{2}(\mathrm{u}) \mathrm{du}
$$

where $\mathrm{W}(\mathrm{t})$ is a Brownian motion on $[0,1]$ and $\int_{0}^{1} \mathrm{WdW}=\frac{1}{2}\left(\mathrm{~W}(1)^{2}-1\right)$. The implication
is that the likelihood ratio test statistic is asymptotically distributed as

$$
\left[\int_{0}^{1} \mathrm{WdW}\right]^{2} / \int_{0}^{1} \mathrm{~W}^{2}(\mathrm{u}) \mathrm{du} .
$$

This distribution is often called the "unit root" distribution and its multivariate version plays an important role for the asymptotic inference for cointegration. We give the main results obtained for likelihood inference, and refer to Johansen $(1988,1991)$ and Ahn and Reinsel (1990) for the technical details.

Theorem 3. Under the null of $r$ cointegrating relations it holds that

$$
-2 \ln Q(r)=-T \Sigma_{r+1}^{p} \ln \left(1-\hat{\lambda}_{i}\right) \xrightarrow{w} \operatorname{tr}\left\{\int_{0}^{1} W d W^{\prime}\left[\int_{0}^{1} W(u) W(u)^{\prime} d u\right]^{-1} \int_{0}^{1} W d W^{\prime}\right\} .
$$

The process W is a ( $\mathrm{p}-\mathrm{r}$ )-dimensional Brownian motion with covariance matrix equal to I. Thus the limit distribution only depends on the number of common trends of the problem. This distribution has then been tabulated by simulation, see Johansen (1988), Johansen and Juselius (1990), Reinsel and Ahn (1990) and Osterwald-Lenum (1992). It is seen that the distribution is a multivariate generalization of the unit root distribution. This is not surprising, since one can think of testing $\rho=1$ in the above model as a test for no cointegration, i.e. of $\mathrm{r}=0$, when $\mathrm{p}=1$, and $\mathrm{k}=1$.

Although the limit distribution given in Theorem 3 only depends on the degrees of freedom or the dimension of the Brownian motion, it turns out that if a constant term or a linear term is allowed in the model then the limit distribution changes. This leads to a number of complications as described in Johansen (1993).

It is quite satisfactory, however, that the other tests described in section 4 on $\alpha$ and $\beta$ all have asymptotic $\chi^{2}$ distributions. Thus the only non-standard test is the test for cointegration rank. The reason for this is that the asymptotic distribution of the estimator of $\beta$ is a mixed Gaussian distribution:

Theorem 4. The asymptotic distribution of $\hat{\beta}_{c}$ is given by

$$
\begin{equation*}
T\left(\hat{\beta}_{c}-\beta_{c}\right)^{\mathrm{w}} \rightarrow\left(I-\beta_{c} c^{\prime}\right) \beta_{\perp}\left[\int_{0}^{1} W_{1}(u) W_{1}(u)^{\prime} d u \Gamma^{1} \int_{0}^{1} W_{1} d W_{2},\right. \tag{7.1}
\end{equation*}
$$

where $W_{1}$ and $W_{2}$ are independent Brownian motions of dimension $p-r$ and $r$ respectively. The asymptotic conditional variance matrix is

$$
\begin{equation*}
\left(I-\beta_{c} c^{\prime}\right) \beta_{\perp}\left[\int_{0}^{1} W_{1}(u) W_{1}(u)^{\prime} d u\right]^{-1} \beta_{\perp}\left(I-c \beta_{c}^{\prime}\right) \otimes\left(c^{\prime} \Pi^{\prime} \Omega^{-1} \Pi c\right)^{-1} \tag{7.2}
\end{equation*}
$$

which is consistently estimated by

$$
\begin{equation*}
T\left(I-\hat{\beta}_{c} c^{\prime}\right) S_{11}^{-1}\left(I-c \hat{\beta}_{c}^{\prime}\right) \otimes\left(c^{\prime} \hat{\Pi}^{\prime} \hat{\Omega}^{-1} \hat{\Pi}_{c}\right)^{-1} \tag{7.3}
\end{equation*}
$$

Thus for given value of $W_{1}$ the limit distribution of $\hat{\beta}$ is just a Gaussian distribution with mean zero and variance given by (7.1). It is this result that implies, by a simple conditioning argument, that the likelihood ratio test statistics for hypotheses about restrictions on $\beta$ is asymptotically distributed as $\chi^{2}$, which again makes inference about $\beta$ very simple if likelihood based methods are used.

Another way of reading the results (7.1), (7.2) and (7.3) is that since $c^{\prime}\left(\hat{\beta}_{c}-\beta_{c}\right)=0$ we need only consider the coefficients $c_{\perp}^{\prime}\left(\hat{\beta}_{c}-\beta_{c}\right)$. It now follows from (7.3) that we can act as if these are asymptotically Gaussian with a "variance" matrix given by

$$
T c_{\perp}^{\prime}\left(\mathrm{I}-\hat{\beta}_{\mathrm{c}} \mathrm{c}^{\prime}\right) \mathrm{S}_{11}^{-1}\left(\mathrm{I}-\hat{c}_{\mathrm{c}}^{\prime}\right) \mathrm{c}_{\perp} \otimes\left(\mathrm{c}^{\prime} \hat{\Pi}^{\prime} \hat{\Omega}^{-1} \hat{\Pi} \mathrm{c}\right)^{-1}
$$

Despite the complicated formulation the result is surprisingly simple since it only states that if $\beta$ is estimated as identified parameters the asymptotic variance of $\beta$ is given by the inverse information matrix which is the Hessian used in the numerical maximization of a function. This result is exactly the same as the result that holds for inference in stationary processes. The only difference is the interpretation of (7.3), which for a stationary process would be an estimate of the asymptotic variance, but for $\mathrm{I}(1)$ processes is a consistent estimator of the asymptotical conditional variance. The basic property, however, is the same in both cases, namely that it is the approximate
scale parameter to use for normalizing the deviation $\hat{\beta}-\beta$.
Table 1 contains a lot of information. We use $\hat{\lambda}_{\mathrm{r}+1}, \ldots, \hat{\lambda}_{\mathrm{p}}$ to test the model, and $\hat{\mathrm{v}}_{1}, \ldots, \hat{\mathrm{v}}_{\mathrm{r}}$ to estimate $\beta$. The remaining vectors $\hat{\mathrm{v}}=\left(\hat{\mathrm{v}}_{\mathrm{r}+1}, \ldots, \hat{\mathrm{v}}_{\mathrm{p}}\right)$ together with $\mathrm{D}=$ $\operatorname{diag}\left(\hat{\lambda}_{1}, \ldots, \hat{\lambda}_{\mathrm{r}}\right)$ contain information on the "variance" of the estimate of $\hat{\beta}$. This can be exemplified in the form of a Wald test. One can prove that under the hypothesis $\mathrm{K}^{\prime} \beta$ $=0$, which is of the form (4.2) for $\mathrm{H}=\mathrm{K}_{\perp}$, it holds that

$$
\left.\operatorname{Ttr}\left\{\mathrm{K}^{\prime} \hat{\beta}\left(\hat{\mathrm{D}}^{-1}-\mathrm{I}\right)^{-1} \hat{\beta}^{\prime} \mathrm{K}\right)\left(\mathrm{K}^{\prime} \hat{\hat{v}}^{\prime} \hat{\mathrm{v}}^{\prime} \mathrm{K}\right)^{-1}\right\}
$$

is asymptotically $\chi^{2}$ distributed. Thus in particular if $\mathrm{r}=1$, and $\hat{\beta}=\left(\hat{\beta}_{1}, \ldots, \hat{\beta}_{\mathrm{p}}\right)$, say, and one wants to see if the coefficient $\beta_{1}$ is significant, one can calculate the quantity

$$
\mathrm{T}^{\frac{1}{2}} \hat{\beta}_{1}\left\{\left(\hat{\lambda}_{1}^{-1}-1\right)\left(\Sigma_{2}^{\mathrm{p}^{\hat{\mathrm{v}}}}{ }_{1 \mathrm{j}}^{2}\right)\right\}^{-\frac{1}{2}}
$$

and compare it to the fractiles in a normalized Gaussian distribution, see Johansen (1991).

One can now discuss why inference about $\beta$ becomes difficult when based on the simple regression estimator. This is because the limiting distribution of $\hat{\beta}_{\text {ols }}$ is expressed as an integral as in Theorem 4, but with dependent $\mathrm{W}_{1}$ and $\mathrm{W}_{2}$. This again implies that for given $\mathrm{W}_{1}$ the limit distribution of the estimator does not have mean zero, which implies that the test statistics based upon the regression estimator will have some "non-central" distribution with nuisance parameters.

Inference for the remaining parameters $\vartheta=\left(\alpha, \Gamma_{1}, \ldots, \Gamma_{\mathrm{k}-1}, \Omega\right)$ is different. This is explained by Phillips (1991) and the idea is roughly the following. The second derivative of the likelihood function with respect to $\beta$ tends to infinity as $\mathrm{T}^{2}$, see (5.1), whereas the second derivative with respect to $\vartheta$ and the mixed derivatives tend to infinity like T . This means that $\hat{\beta}-\beta$ has to be normalized by T and $\hat{\vartheta}-\vartheta$ by $\mathrm{T}^{\frac{1}{2}}$. This on the other hand requires a normalization of the mixed derivatives by $\mathrm{T}^{3 / 2}$ and makes them disappear in the limit. Thus in the limit the information matrix, which is usde to normalize $(\hat{\beta}-\beta, \hat{\vartheta}-\vartheta)$, is block diagonal with one block for $\beta$ and on block for the remaining parameters $\vartheta$. Thus inference concerning $\beta$ can be conducted as if $\vartheta$
were known and vice versa, see Johansen (1991).

## 8. Conclusion

We have shown that the notion of cointegration as a way of describing long-run economic relations can be formulated in the autoregressive framework as the hypothesis of reduced rank of a certain matrix. This allows explicit maximum likelihood estimation of the cointegration relations, both unrestricted as well as under certain types of linear restrictions that seem to correspond to interesting economic hypotheses. We have found the asymptotic distribution of the likelihood ratio test for the cointegration rank, and tabulated it by simulation, and shown that restrictions on $\beta$ can be tested using the $\chi^{2}$ distribution. This shows that the property of non-stationarity of the processes instead of being a nuisance to be eliminated by differencing can be used as a strong tool to investigate long-run dependencies between variables.

## 9. References

Ahn, S.K. and G.C. Reinsel (1990): " Estimation for partially non-stationary multivariate autoregressive models" Journal of the American Statistical Association, 85, 813-823.

Anderson, T.W. (1951): "Estimating linear restrictions on regression coefficients for multivariate normal distributions," Annals of Mathematical Statistics, 22, 327-351.

Anderson, T. W. (1971): "The statistical analysis of time series," New York: Wiley.
Chan, N.H. and C.Z.Wei (1988): "Limiting distributions of least squares estimates of unstable autoregressive processes," Annals of Mathematical Statistics, 16, 367-410.

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

Fuller W. (1976): "Introduction to Statistical Time Series," New York: Wiley.
Granger, C.W.J. (1983): " Cointegrated Variables and Error Correction Models," UCSD Discussion paper 83-13a.

Jeganathan, P. (1993): "Some Aspects of Asympotic Theory with Applications to Time Series Models," Econometric Theory (to appear)

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," Journal of Economic Dynamics and Control, 12, 231-254.

Johansen, S. (1989): "Likelihood based inference on cointegration," Lecture notes, Centro Inter-universitario di Econometria, Bologna

Johansen, S. (1991): "Estimation and Hypothesis Testing of cointegration vectors in Gaussian Vector Autoregressive Models", Econometrica, 59, 1551-1580.

Johansen, S. (1992): "Cointegration in partial systems and the efficiency of single equation analysis," Journal of Econometrics, 52, 389-402.

Johansen, S. (1993): "The role of the constant and linear terms in cointegration analysis of nonstationary variables," to appear in Econometric Reviews.

Johansen, S. and K. Juselius (1990): "Maximum Likelihood Estimation and Inference on Cointegration - with Applications to the demand for money," Oxford Bulletin of Economics and Statistics, 52, 169-210.

Johansen, S. and K. Juselius (1992): "Structural hypotheses in a multivariate cointegration analysis of the PPP and UIP for UK," Journal of Econometrics, 53, 211-244..

Johansen, S. and K. Juselius (1993): "Identification of the long-run and the short run structure. An application to the ISLM model. Journal of Econometrics (to appear)

Osterwald-Lenum, M.(1992): "A note with fractiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistic: Four cases," Oxford Bulletin of Economics and Statistics 54, 461-471.

Phillips, A. W. (1954): " Stabilisation policy in a closed economy," Economic Journal, 64, 290-323..

Phillips, P.C.B. (1987): "Time series with a unit root," Econometrica, 55, 277-301.
Phillips, P.C.B. (1991): "Optimal inference in cointegrated systems," Econometrica 59, 283-306.

Phillips, P.C.B., and S.N. Durlauf (1986): "Multiple Time Series Regression with Integrated Processes," Review of Economic Studies, 53, 473-495.

Phillips, P.C.B., and B.E. Hansen (1990): "Statistcal inference on instrumental variables regression with I(1) processes," Review of Economic Studies, 57, 99-124.

Phillips, P.C.B., and S. Ouliaris (1990): "Asymptotic Properties of Residual based Tests for Cointegration," Econometrica, 58, 165-193.

Park, J.Y. (1992): "Canonical cointegrating regressions," Econometrica, 60, 119-143.
Park, J.Y. and P.C.B. Phillips (1988): "Statistical Inference in Regressions with Integrated Processes : Part 1," Econometric Theory, 4, 468-498.

Park, J.Y. and P.C.B. Phillips (1989): "Statistical Inference in Regressions with Integrated Processes : Part 2," Econometric Theory, 5, 95-131.

Reinsel G.C. and S.K. Ahn (1990): " Vector autoregressive models with unit roots and reduced rank structure, estimation, likelihood ratio test, and forecasting," Journal of Time Series, 13,283-295.

Sargan, J.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 for National Economic Planning, Butterworths, London.

Stock, J.H. (1987): "Asymptotic Properties of Least Squares Estimates of Cointegration Vectors," Econometrica, 55, 1035-1056.

Stock, J.H., and M.W. Watson (1988): " Testing for Common Trends," Journal of the American Statistical Association 83, 1097-1107.

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, DK-2100 COPENHAGEN $\varnothing$, DENMARK. TELEPHONE + 4535320899.

No. 1 Johansen, Søren: The Role of the Constant Term in Cointegration Analysis of Nonstationary Variables.

No. 2 Paruolo, Paolo: Asymptotic Inference on the Moving Average Impact Matrix in Cointegrated I(1) VAR Systems.

No. 3 Johansen, Søren and Juselius, Katarina: Identification of the Long-Run and the Short-Run Structure. An Application to the ISLM Model.

No. 4 Johansen, Søren: Identifying Restrictions of Linear Equations.

## Preprints 1993

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, DK-2100 COPENHAGEN $\varnothing$, DENMARK. TELEPHONE +45 35320899 .

No. 1 Hansen, Henrik and Johansen, Søren: Recursive Estimation in Cointegration VAR-Models.
No. 2 Stockmarr, A. and Jacobsen, M.: Gaussian Diffusions and Autoregressive Processes: Weak Convergence and Statistical Inference.
No. 3 Nishio, Atsushi: Testing for a Unit Root against Local Alternatives
No. 4 Tjur, Tue: StatUnit - An Alternative to Statistical Packages?
No. 5 Johansen, Søren: Likelihood Based Inference for Cointegration of Non-Stationary Time Series.


[^0]:    ${ }^{1}$ Cramer Lecture presented at the 2nd World Conference of the Bernoulli Society for Mathematical Statistics and Probability. Uppsala, Sweden 13-18 August 1990.

