Søren Johansen

Estimating Systems of Trending Variables

Preprint
September
1991
6

Institute of Mathematical Statistics
University of Copenhagen

ISSN 0902-8846

Søren Johansen

ESTIMATING SYSTEMS OF TRENDING VARIABLES

Preprint 1991 No. 6

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

September 1991

University of Copenhagen
Institute of
Mathematical Statistics

Revised September 25 1991

ESTIMATING SYSTEMS OF TRENDING VARIABLES

Søren Johansen[1]

Lecture presented at the ESEM 1991 in Cambridge

---

[1]This research is part of a project supported by the Danish Social Sciences Research Council, acc. 14–5793. The paper was written while the author was visiting Department of Statistics, University of Helsinki.

1.Introduction

This paper will be on the statistical analysis of systems, on common trends and on cointegration but first of all a paper on the construction and analysis of models.

A model is here meant to be a statistical model, that is, a parameterized family of probability measures. It is my hope to demonstrate that by building statistical models that can be considered a framework for the analysis of economic models, one can improve ones understanding of the economic phenomena. The main emphasis in econometrics has for a long time been on methods, where a method is here simply understood as a rule for the calculation of a test statistic or an estimator.

A recent survey by Cambell and Perron (1991) gives an excellent and detailed account of many different techniques and their relations and potential usefulness both for the univariate and multivariate situation. I shall be much less detailed and emphasize the underlying statistical concepts and their relation to the economic concepts, rather than the variety of techniques. It is also apparent that the paper turned out to be a summary of the results that I have obtained on the analysis of the autoregressive model, but I have tried to relate the results to some of the other contributions in the area.

I shall survey some contributions to the topic mentioned in the title with special emphasis on the analysis of the statistical model. The reason for this is that by carefully constructing a statistical model where the economic concepts can be defined precisely and the hypotheses of interest can be tested one would hope that the analysis of the model would lead to relevant methods that have better properties than methods that are suggested on more intuitive grounds. Although some authors emphasize models and others emphasize methods, the two concepts complement one another, in the sense that a prime goal of the analysis of a model is to produce a method, and the properties of a method can only be discussed with some sort of model in the background.

Once the method has been derived by an analysis of a statistical model one can of course use it under all sorts of different circumstances, provided one can prove that it has reasonable properties under these other circumstances, that is, in some other statistical model. Thus one can think of the statistical model as a method generating tool, but my conviction is that it is much more than that, and I hope to demonstrate this in the following.

From my experience with economic data I find that formulating the interrelations between economic variables as a system seems a useful activity. It is not quite clear, however, what one wants to achieve, and this will be discussed in the next section. Certainly the ideal is to have a serious economic theory for the variables actually observed. Since opinions on what is a serious economic theory seem to diverge, it would be good to have some way of checking which of two rival economic theories is better.

Statistics offers such a possibility using the ideas of hypothesis testing. Unfortunately all tests rest on precise assumptions about the underlying statistical model and it is therefore important to have methods that can check if such assumptions are satisfied.

Thus the methodology is to build a statistical model that describes the fluctuations of the data, and express the economic theory in terms of the parameters of the statistical model. Once this has been done an analysis of the model will reveal how the interesting economic parameters can be estimated, and how hypotheses of economic interest can be tested.

2. Building models for systems of trending variables.

2.1. Trending variables, cointegration and common trends.

Many economic variables are trending over time, and when modeling this phenomenon one can focus on two aspects of the economy, either the "stable" economic relations between variables that show less variability than the original variables, that is, those relations between non–trending variables, or the driving economic forces, which create the non–stationarity. The purpose of this section is to give these concepts a precise statistical definition in order to be able to discuss them in detail in a statistical model.

By a trending variable we basically mean a stochastic process which is non–stationary, and which becomes stationary after differencing a suitable number of times. There are however other ways of removing the non–stationarity of a process. Consider for instance a stationary process Y_t and define $X_t = Y_t - Y_1$, $t = 1,2,\ldots$. Then in general X_t is non–stationary and ΔX_t is stationary, but we do not want to describe X_t as trending, since the non–stationarity can be removed by adding Y_1 to the process. This leads to the definition:

DEFINITION 2.1 *A p–dimensional stochastic process* $\{X_t, t = 1,2,\ldots\}$ *is called trending if* $\{X_t + Z, t = 1,2,\ldots\}$ *is non–stationary for all random variables* Z.

The trend can either be deterministic or stochastic, and we define the deterministic trend as $\mu_t = EX_t$, and the stochastic trend as $X_t - EX_t$.

We then focus on the type of non–stationarity that can be removed by differencing, and define

DEFINITION 2.2 *For* $d = 0,1,2,\ldots$ *the trending process* $\{X_t, t = 1,2,\ldots\}$ *is called an I(d) process if* $\Delta^{d-1}(X_t - EX_t)$ *is trending, and* $\Delta^d(X_t - EX_t)$ *is stationary.*

Here $\Delta^{-1}(X_t - EX_t) = \Sigma_1^t(X_i - EX_i)$. The purpose of the definition is of course first of all to describe a random walk as a trending process that is I(1). If we had phrased Definition 2.2 in terms of non–stationarity we would not have excluded the process $X_t = Y_t - Y_1$. In

Engle and Granger (1987) the definition of I(1) excludes processes with a completely deterministic part. The present attempt is another way of solving this problem. We also want to describe the two dimensional process consisting of a random walk and a stationary sequence as an I(1) process thus allowing the component processes to be integrated of different order. The present definition has the property that it is invariant to non singular linear transformations of the process. An I(0) process Y_t is thus a stationary process with the property that its sums are trending, thus if for instance $Y_t = \Delta U_t$ for some stationary process U_t then Y_t is not I(0), since the sum $\Sigma_1^t Y_i = U_t - U_1$ is not trending.

Thus a trending variable is composed of an I(d) variable with mean zero and a possible deterministic trend. Consider the following simple example that illustrates the various possibilities for d = 2:

$$X_t = C_2 \sum_{j=1}^{t} \sum_{i=1}^{j} \epsilon_i + C_1 \sum_{i=1}^{t} \epsilon_i + \tau_0 + \tau_1 t + \tfrac{1}{2}\tau_2 t^2 + Y_t,\ t = 1,2,\ldots \tag{2.1}$$

We assume that ϵ_t are i.i.d, and we let Y_t be some stationary process $Y_t = \sum_{j=0}^{\infty} C_j^* \epsilon_{t-j}$. Clearly X_t is trending, since differencing it twice makes it stationary. Note, however, that if $C_2 = 0$, then we still need to difference it twice to make it stationary. We prefer to call such a process an I(1) process, since the stochastic trend needs only one differencing.

The moving average representation (2.1) is a useful way of modeling the economy through the matrices C_1 and C_2 as the results of the influence of its unobserved common trends or driving forces. It is furthermore very convenient for describing the properties of the process since the mean and covariance functions are easily calculated and the cointegration properties are easily illustrated, as will now be discussed. Furthermore the asymptotic properties are simple consequences of (2.1).

Granger (1981) used this representation to note that if we take linear combinations β, such that $\beta' C_2 = 0$, and $\beta' C_1 \neq 0$, then the order of integration of the process is reduced from 2 to 1. He coined the phrase cointegration and denoted it C(2,1) in this case. The idea is to describe the "stable" relations in the economy by linear relations that are more stationary than the original variables.

DEFINITION 2.3 *The I(d) process* X_t *is called cointegrated C(d,b) with cointegrating vector* λ *if* $\lambda' X_t$ *is I(d−b), b = 1,...,d, d = 1,2,... .*

If we can find β such that $\beta' C_1 = \beta' C_2 = 0$, see (2.1), then clearly $\beta' X_t$ is stationary apart from its quadratic trend. Thus the stochastic variation has been reduced to stationarity. We call $\beta' X_t$ trend stationary (with a quadratic trend). If we find two vectors β_0 and β_1, such that $\beta_0' C_2 = 0$, and such that $\beta_0' C_1 + \beta_1' C_2 = 0$, then, assuming that $\mu_t = 0$,

$$\beta_0' X_t + \beta_1' \Delta X_t = \beta_0' C_2 \sum_{j=1}^{t} \sum_{i=1}^{j} \epsilon_i + (\beta_0' C_1 + \beta_1' C_2) \sum_{j=1}^{t} \epsilon_j + \beta_0' Y_t + \beta_1' C_1 \epsilon_t + \beta_1' \Delta Y_t,$$

which is stationary by the choice of β_0 and β_1. Thus the levels X_t are reduced to I(1) by the coefficients β_0, and these linear combinations cointegrate with the differences through the linear combinations $\beta_1' \Delta X_t$ which also form an I(1) process. This phenomenon is called multi–cointegration or polynomial cointegration, see Engle and Yoo (1989) and Gregoire and Laroque (1990).

The representation (2.1) models the variables by common trends, and the reduced rank of the coefficient matrices ensures that the variables cointegrate, since by suitable linear combinations the common trends can be eliminated, thereby creating "stable" economic relations. Another way of modeling cointegrating variables is through the so called error correction models.

As an example consider the model

(2.2) $$\Delta X_t = \alpha\beta' X_{t-1} + \mu + \epsilon_t,\ t = 1,...,T,$$

with initial value X_0, where α and β are p×r matrices. The motivation for this model is to consider the relation $\beta' X = 0$ as defining the underlying economic relations, and assume that the agents react to the disequilibrium error $\beta' X_{t-1}$ through the adjustment coefficient

α, to bring back the variables on the right track, that is, such that they satisfy the economic relations. If α and β have rank r we define $\alpha_\perp$ and $\beta_\perp$ as p×(p−r) matrices of full rank such that $\alpha'\alpha_\perp = \beta'\beta_\perp = 0$. If further $\alpha_\perp'\beta_\perp$ has full rank then one can solve (2.2) for X_t as a function of the initial values and $\epsilon_1,\ldots,\epsilon_T$ and find a representation of the form (2.1) with $C_2 = 0$, $\tau_2 = 0$, and

(2.3) $$C_1 = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' \text{ and } \tau_1 = C_1\mu,$$

whereas the value of τ_0 depends on whether we fix all of Z_0, or admit the stationary components to have their invariant distribution, that is, it depends on how we interpret the expectation in the calculation of the deterministic trend.

This is an instance of the celebrated representation theorem by Granger, see Engle and Granger (1987), which is a way of finding the MA representation from the AR representation, and vice versa, when there are I(1) variables in the system.

Granger also used model (2.2) to define common trends by the following observation: if we multiply the equation (2.2) by $\alpha_\perp'$ we obtain a set of variables that evolve without reacting to the disequilibrium error, and in this sense can be considered as modeling the driving forces of the economy.

DEFINITION 2.4 *The variables $\alpha_\perp' X_t$ are the common trends.*

The representation (2.1) contains the cumulative shock $\Sigma_1^t\epsilon_i$ with the coefficient matrix C_1, but the expression (2.3) shows that only the linear combinations $\alpha_\perp'\Sigma_1^t\epsilon_i$ enter the representation. Thus the driving forces of the process are modelled by $\alpha_\perp'\Sigma_1^t\epsilon_i$. From the expression

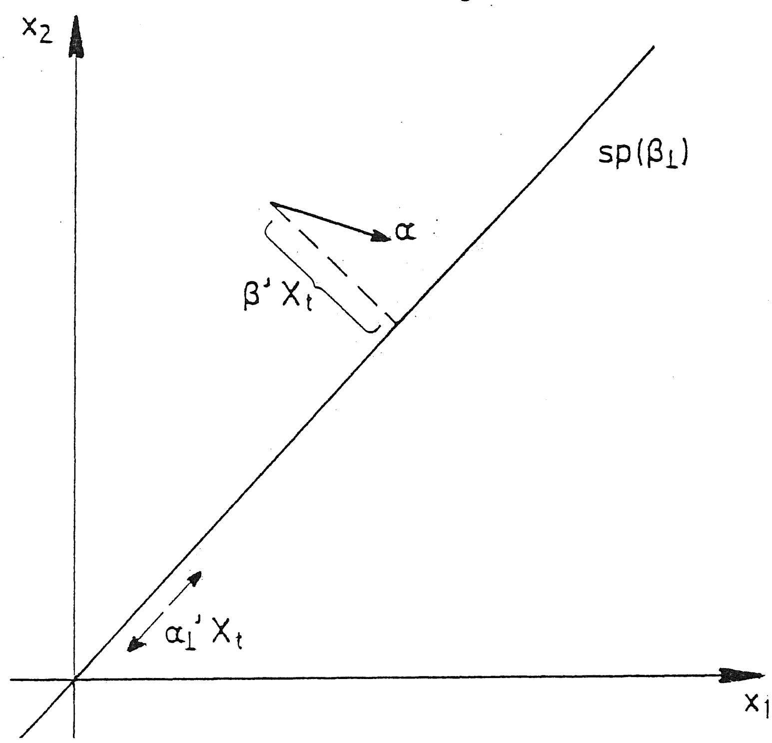
$$\alpha_\perp' X_t = \alpha_\perp'\Sigma_1^t\epsilon_i + \alpha_\perp'(\tau_0 + \tau_1 + Y_t)$$

it is seen that the common trends have the same non–stationary component as the driving force.

The error correction model, exemplified by (2.2), was introduced into econometrics by Sargan (1964), and has since been applied when modelling an economy where agents learn from the past when making their plans for the future.

The error correction model (2.2), which is constructed around the cointegrating relations as error correction terms, and the MA model (2.1) that is constructed in terms of unobserved common trends, are of course complementary, in the sense that the two approaches are mathematically equivalent, but they may appeal to different types of intuition.

Figure 1



Legend to Figure 1:

The process X_t is pushed along the attractor set $sp(\beta_\perp)$ by the common trends $\alpha_\perp' X_t$ and pushed away from this by the shocks to the system. The process reacts to the disequilibrium error $\beta' X_t$ via the adjustment coefficients and is pushed back towards the attractor set in the direction α.

The picture that one should have in mind is that the economic forces, $\alpha_{\perp}' X_t$, drive the economic variables to lie in the space spanned by $\beta_{\perp}$, termed the attractor set by Granger. The agents react to these forces and create economic variables that move around the common trends, following the economic "laws" or structural relations $\beta' X = 0$, in the sense that the variables react to the disequilibrium errors, $\beta' X_t$, through the adjustment coefficients, α, and are forced back towards the attractor set.

This closes our discussion of the fundamental concepts like trending variables, cointegration, and common trends, and the next sections will contain various attempts to model and analyze these phenomena. We conclude this section with some general comments on the purpose of building models for economic data.

2.2. Some comments on model building

First of all it must be emphasized that the purpose of constructing a statistical model, that is a parametrised family of probability distributions, is not to replace serious economic models with arbitrary statistical descriptions, but rather to provide a framework in which one can compare the economic theories with reality, as measured by the data series.

It seems that a proper statistical treatment of systems of trending variables should include the formulation of a statistical model where one can

1) Describe the stochastic variation of the data, such that inferences conducted concerning the various economic questions are valid.

2) Define precisely the economic concepts, like a "stable relation" and "driving force" as the statistical concepts cointegration and common trends.

3) Formulate interesting economic theories and questions in terms of the parameters of the model.

4) Derive estimators and test statistics as well as their (asymptotic) distributions such that useful inferences can be drawn.

If we have such a statistical model describing a relevant set of economic variables, the first task is to check for model specification, to make sure that property 1 is true, or at least not completely false. If it turns out that the model is valid from a statistical point of view, one can proceed to test that the formulated economic theory, or economic model, is consistent with the data. The reason that this point is important is that often the economic theory is developed for rather abstract concepts, whereas when it comes to the observations, one has to put up with actual data series that are not observations of the abstract concepts, but carefully selected proxies. Thus although the economic theory may be fine, the data chosen may not illustrate this. Hence a careful statistical analysis helps to support the economic conclusions.

In the case of cointegration and common trends there are a number of questions that need a statistical formulation and treatment. First of all the number of the cointegrating relations or common trends has to be determined, to compare with the number prescribed by the theory. Next economic hypotheses about the cointegrating relations or common trends have to be formulated and tested, and here the interpretation of the concepts becomes very important.

An economic theory is often formulated as a set of behavioral relations or structural equations between the levels of the variables, possibly allowing for lags as well. If the variables are I(1) it is convenient to reformulate it in terms of levels and differences, such that if a structural relation is modeled by a stationary relation then we are lead to consider stationary relations between levels, that is cointegrating relations.

It is well known that structural results are difficult to extract from the reduced form equations, but it is easy to see that the property of non–stationarity and stationarity can be deduced from the reduced form since they are basically statistical concepts, rather than economic notions.

The reason that cointegration is interesting is that the cointegrating relation captures the economic notion of a stable economic relation. And the reason that a statistical theory (rather than an economic theory) for the estimation and testing of cointegrating relations can be constructed rests on the fact that the reduced form suffices for the determination of the basic cointegrating relations. On the other hand a mindless attempt to finding cointegrating relations without knowing what they mean is not going to be fruitful, so I believe that the econometrician has to carefully choose the variables that should enter the study, and carefully discuss the economic theory that motivates this.

The reason for this discussion of the statistical model is that one often finds the opinion expressed that the purpose of a statistical analysis is to find the estimates of the parameters that one knows are the interesting ones. What I want to point out is that the statistical model offers a much richer basis for discussion of the relation between economic theory and economic reality.

In the above discussion we have focused entirely on statistical models that describe full systems, that is the joint stochastic behavior of all the processes observed. In situations where one has 25 or even 100 variables, this may not be feasible, since the interrelations between so many variables is extremely difficult to understand. It is customary to fix certain variables, which it is felt influence the main variables without being influenced by the variables of main concern. Thus assuming some sort of exogeneity one can construct a partial model.

It is obvious that if we do not specify the stochastic properties of the exogenous variables it is impossible to make statistical inferences for the parameters that have been estimated. A compromise is to model some variables carefully and some variables less carefully; that is, one can try to develop methods for the parameters of interest that are valid under a wide range of assumptions on the exogenous variables. This is the background for the semi parametric and regression type models considered in the next section and the topic will be discussed further in connection with the VAR model.

3. Statistical analysis of models for I(1) variables

The most widely used type of non–stationarity for economic series is I(1), and this section contains some results for this type of variables. We distinguish between regression type models and models for systems. This distinction is not quite clear cut and meant only as an attempt to structure a large body of work done by many authors.

3.1. *Regression models for the long–run parameters*

3.1.1 *Estimation of parameters*

Consider the model with variable, $X_t = (Y_{1t}', Y_{2t}')'$, of dimensions p_1 and p_2 respectively.

(3.1) $$Y_{1t} = BY_{2t} + U_{1t}$$

(3.2) $$\Delta Y_{2t} = U_{2t}.$$

where the process U_t is assumed to be an invertible stationary process. This is a model which has imbedded into it the cointegrating relations $\beta = (1,-B)'$ and the common trends $Y_{2t} = \Sigma_1^t U_{2i}$. The appealing property from the model point of view, is that one focuses on the interesting parameters B and considers the dynamics of the underlying stationary process as nuisance parameters. The model is treated in detail by Phillips (1990), and I shall here give a short summary of the results obtained.

As parameters in the model one can consider

(3.3) $$(B, f_u(.)),$$

where f_u denotes the spectral density of the process $\{U_t\}$. This is a model with minimal assumptions, many parameters, and no structure on the spectral density. It contains as sub–models almost any possible model in relation to the present questions, since it only requires that Y_{1t} and Y_{2t} cointegrate, and that the common trends Y_{2t} are known. We exemplify by considering some sub models, taken from the paper by Phillips.

The first example is purely pedagogical, since it involves specifying that

(3.4) $$U_t \text{ is i.i.d } N_p(0,\Omega).$$

This is in applications far from realistic, but the model serves the purpose of generating an estimator namely the regression of Y_{1t} on Y_{2t} and ΔY_{2t}, which can be derived as the maximum likelihood estimator. If in particular U_{1t} and U_{2t} are assumed independent, then the estimator generated is the regression of Y_{1t} on Y_{2t} as originally suggested by Engle and Granger (1987), and investigated by Stock (1987), under more general assumptions, as the first estimation method for cointegrating relations. The real advantage of model (3.4) is not its realism, but instead the fact that it is relatively easy to see that the asymptotic distribution of the maximum likelihood estimator is mixed Gaussian. Thus the model provides a simple framework in which one can convince oneself without too many technicalities that all this discussion about mixed Gaussian distributions is necessary even in the very simplest unrealistic cases as long as I(1) variables are involved.

A more realistic example is the sub model found by parameterizing the spectral density $f_u = f_{u,\vartheta}$ by a finite number of parameters ϑ such that the parameters become

$$(B,\vartheta) \in \Psi \times \Theta \tag{3.5}$$

The Whittle likelihood, see Dunsmuir and Hannan (1976)

$$L_W(B,\vartheta) = \ln|\Sigma(\vartheta)| + T^{-1}\Sigma_s \mathrm{tr}\{f_u^{-1}(\lambda_s;\vartheta)w(\lambda_s)w^*(\lambda_s)\} \tag{3.6}$$

with $\lambda_s = (2\pi s)/T$ and

$$f_u(\lambda,\vartheta) = (2\pi)^{-1}(\Sigma_0^\infty C_k(\vartheta)e^{ki\lambda})\Sigma(\vartheta)(\Sigma_0^\infty C_k'(\vartheta)e^{-ki\lambda})$$

and

$$w(\lambda) = (2\pi T)^{-1}\Sigma_1^T(Y_{1t}-BY_{2t-1},\Delta Y_{2t})'e^{it\lambda}$$

is a useful way of writing explicitly an expression for an approximation to the Gaussian likelihood function involving the spectral density and the empirical periodogram, thus avoiding the inversion of a large covariance matrix for the observations $(Y_{1t},Y_{2t},\ t = 1,\ldots,T)$.

This gives, at least in principle, a possibility of calculating a pseudo maximum likelihood estimator provided one can prove it exists. It is clear that the interpretation of the parameters ϑ is a bit difficult in such a general formulation but special cases may exist, where this is possible. One can also achieve an even more general type of model in which the parameter B also enters into the spectral density $f_u = f_{u,\vartheta,B}$, such that the parameters in the cointegrating relation and the spectral density are no longer variation independent. This last formulation has the advantage that it contains all VAR models, and in fact all ARMA and ARIMA models.

The main result of Phillips (1990) on the analysis of these models is that if one applies Gaussian maximum likelihood estimation in models like (3.6) then one achieves optimal inference in a sense discussed by Phillips (1990). For the present purpose we will think of optimal inference as the possibility of obtaining a mixed Gaussian distribution for $\hat{B}$, expressed in terms of two independent Brownian motions, and thus asymptotic χ^2 inference concerning hypotheses on B, our parameter of interest.

Thus it is a very general type of model, where the main conclusion is that in order to make sure that one gets optimal inference one should model the dynamics and calculate the maximum likelihood estimator. The model is defined in such general terms that it is more a framework in which usual inference works, than a prescription of which model to use, and how to analyze it.

The model (3.1) has given rise to a large number of different estimators of regression type. A direct regression of Y_{1t} on Y_{2t} leads to a limit distribution which is a mixture of Gaussian distributions, but expressed in terms of two dependent Brownian motions. It turns out that the variance matrix of these Brownian motions is just the long run variance matrix for the underlying stationary process. This has lead to a modified regression estimator, see Phillips and Hansen (1990) or Park (1988):

1) Estimate B by regression of Y_{1t} on Y_{2t}.

2) Estimate a long run variance matrix from the residuals by for instance a spectral estimator.

3) Correct Y_{1t} and Y_{2t} to Y_{1t}^* and Y_{2t}^* using the long run variance matrix and ΔY_{2t}.

4) Regress Y_{1t}^* on Y_{2t}^*.

The effect of the correction is that now the regression estimate will have a mixed Gaussian distribution, and asymptotic inference can be performed by Wald tests using the estimate of the long run variance matrix.

Phillips (1988) estimates B by regression in the frequency domain, and another way of achieving the independence between the two Brownian motions in the limit distribution is to regress Y_{1t} on Y_{2t} plus leads and lags of ΔY_{2t}, see Saikkonen (1989) and Stock and Watson (1991). These regression estimates have the property that the dynamics of the process U_t need not be modeled, only the long run variance needs to be estimated for optimal inference on B to be possible.

The reason for spending relatively little space on these regression type estimators, is that they are basically invented as extensions of the usual regression method with the purpose of deriving estimates without modeling the structure of the data in detail, and the main purpose of this paper is to discuss the analysis of systems.

3.1.2 Determination of cointegration rank

There is one aspect of the practical problem that has not been dealt with in the discussion of the above class of models. That is how to determine the cointegration rank, or the number of common trends. In fact the model formulation assumes that we can point out the common trends, as the variables belonging to Y_{2t}. Another way of saying this is that we have to assume that the variables Y_{1t} enter in the cointegrating relations in such a way that one can solve them for Y_{1t}.

From my experience with econometric data it is rather rare that ones prior guess of the number of cointegrating relations is correct, and it seems important to have a way of checking that one is not completely wrong. But why is this not possible in the above formulation? Usually when we want to test hypotheses, it is convenient to think of one model as a sub–model of another, that is, the models are nested. This gives the possibility of formulating likelihood ratio tests. The drawback of the model formulation (3.1) and (3.2) is that if we change the dimension of Y_{1t} we get non–nested hypotheses. That is, there is no value of the parameters in the model with $p_1 = 3$, say, for which the model reduces to a model with $p_1 = 2$. Thus the model with $p_2 = 2$ is not a sub–model of the model with $p_1 = 3$. This implies that there seems to be no simple way of testing if the original model formulation is correct with respect to the specification of the number of common trends, or cointegrating relations.

The first paper to deal systematically with this problem from the point of view of the analysis of a system is the paper by Stock and Watson (1988). They explicitly formulate the nested hypotheses $\mathcal{H}_k$ of k common trends, and suggest some methods for testing the $\mathcal{H}_k$ in $\mathcal{H}_m$, m < k. Their starting point is the model considered by Granger

$$\Delta X_t = C(L)\epsilon_t,$$

where ϵ_t are i.i.d with mean zero, say, and

$$C(z) = \Sigma_0^\infty C_i z^i.$$

The parameter space is not explicitly given, but one can supplement it by specifying a parametric dependence of the coefficients C_i on some parameters. This however is not used in the analysis that follows.

Cointegration is associated with reduced rank of the matrix $C(1) = \Sigma_0^\infty C_i$. If β is defined as the matrix of largest rank such that $\beta' C(1) = 0$, then Stock and Watson introduce the coordinate system defined by $(\beta, \beta_\perp)$, where $\beta_\perp$ is a matrix of full rank such that $\beta'\beta_\perp = 0$. They further define the common trends $W_t = \beta_\perp' X_t$. The idea is now that if W_t really contains all the common trends and they do not cointegrate, then ΔW_t is an invertible stationary process, and a regression of W_t on W_{t-1} will give a coefficient $\hat{\Phi}$ that converges toward the identity. It therefore seems natural to construct a test based upon the eigenvalues of $\hat{\Phi}$ to see if they are close to 1. Thus the test that they give for the null of r = p–k cointegrating relations against s = p–m cointegrating relations, m < k, is to investigate the m'th largest real part of the eigenvalues of $\hat{\Phi}$. While this explains the main idea, the actual procedure they suggest is to first estimate β, using principal components of the sums of squares matrix formed by the data, then use the estimated β to calculate W_t, and finally filter the series W_t by regressing it on its lagged value. The matrix $\hat{\Phi}_f$ is then constructed by regressing the residuals on the lagged residuals. They find the asymptotic distribution of the eigenvalues as those of the functional

$$[\int_0^1 BB' du]^{-1} \int_0^1 B(dB)',$$

where B is a p–r dimensional Brownian motion, and tabulate the critical values. Thus the procedure consists of fixing k, estimating β, filtering W_t and finding the (m+1)'th largest eigenvalue of $\hat{\Phi}_f$ as the test statistic.

It is a bit difficult to see how one can derive formally their analysis by an analysis of the model. Certainly the long–run variance is $C(1)Var(\epsilon_t)C(1)'$, which is singular if C(1) is, hence the smallest principle components of the sum of squares of the observations is a candidate for an estimate of the cointegrating relations β. It is clear that the filtering makes one think of an autoregressive model for the common trends. We shall see how the analysis of the VAR model in the next section leads to similar procedures as a result of an analysis of the likelihood function. Notice how the principal component analysis treats the variables on an equal footing, that is without solving the relation for any of the variables, and how one can choose any number of principal components to model the number of common trends, but note first of all how the formulation of the model allows the hypotheses of interest to be nested. Thus by exploiting the reduced rank of the MA representation that served as the starting point for Granger's original formulation of the cointegration problem, Stock and Watson manage to formulate the hypotheses of interest as nested hypotheses, thus making a formal test possible. We shall see how the VAR model allows a similar formulation exploiting the dual form of the representation, namely the autoregressive representation.

3.2 The autoregressive model

The interest in the VAR model formulation in econometrics seems to have been inspired by the paper by Sims (1980), and many of the estimation problems in the unrestricted VAR have been solved by Sims, Stock and Watson (1990). The reduced rank regression technique, which originally was proposed by an analysis of the likelihood function in the i.i.d case by Anderson (1951) was applied by Ahn and Reinsel (1988) for stationary processes, and by Johansen (1988) for non–stationary processes. The corresponding asymptotic theory was given in Johansen (1988) and was also developed by Ahn and Reinsel (1990) and Reinsel and Ahn (1990), and the results are generalized by Phillips (1990), in the above discussed framework.

What then is so special about the VAR model, when it is just a sub model of the general set up in section 3.1 ?

In my opinion the flexibility of the autoregressive formulation allows not only the statistical description of a large number of real data sets, but it also allows the embedding of interesting economic hypotheses into a general statistical framework, in which one can define the concepts of interest and actually prove results about them, as discussed in section 2. I shall therefore spend some more time on the autoregressive models.

What we gain by the autoregressive formulation should be apparent from what follows; what we loose is of course that we treat a rather small class of processes, such that in the scale between convenience and realism some may feel that we are a bit too far in the direction of convenience. Our experience with analyzing macro data, however, is that the models quite often provide an adequate fit of the data, especially after we have considered carefully which variables should be included in the analysis. It is our belief that seasonal dummies serve to diminish the required lag length considerably. It is also our belief that the multivariate information set allows one to decrease the lag length compared to a univariate analysis. It is an empirical finding that we have so far been able to manage with two lags for seasonally unadjusted data. This is important, since in VAR models one can easily get involved in estimating many parameters explicitly.

3.2.1 *Definition and representation of I(1) variables*

The VAR model with Gaussian errors is defined by the equations

$$X_t = \Sigma_1^k \Pi_i X_{t-i} + \mu_0 + \mu_1 t + \epsilon_t, \quad t = 1,...,T, \tag{3.7}$$

The parameter space is defined by the parameters

$$(\Pi_1,..., \Pi_k, \mu_0, \mu_1, \Omega) \tag{3.8}$$

varying freely. If ϵ_t are i.i.d $N_p(0,\Omega)$ then the maximum likelihood estimation of the parameters is just ordinary least squares, which of course is quite convenient since most computer packages can do this without hesitation and regression coefficients are well understood in econometrics.

The hypothesis of cointegration is defined as

$$\mathcal{H}_r: \quad \Pi = I - \Pi_1 - ... - \Pi_k = \alpha\beta' \tag{3.9}$$

where α and β are p×r matrices. Note that $\mathcal{H}_r$ specifies that the rank of Π is less than or equal to r which ensures that the models are nested: $\mathcal{H}_0 \subset ... \subset \mathcal{H}_p$. If we reparametrize the model (3.7) as a reduced form error correction model we can write it as

$$\Delta X_t = \alpha\beta' X_{t-1} + \Sigma_1^{k-1} \Gamma_i \Delta X_{t-i} + \mu_0 + \mu_1 t + \epsilon_t \tag{3.10}$$

with parameter space

$$(\alpha,\beta,\Gamma_1,...,\Gamma_{k-1}, \mu_0, \mu_1, \Omega), \tag{3.11}$$

with $\Gamma_i = -(\Pi_{i+1} + ... + \Pi_k)$. This parameterization has the advantage that the parameters vary unrestrictedly. Thus (3.10) defines a sub–model of the general VAR model, and this sub–model will be called an I(1) model. Note that since only the product $\alpha\beta'$ enters into the model the parameters α and β are not identified. What can be estimated is sp(β), the space spanned by the columns of β or equivalently the rows of Π, and similarly sp(α), the space spanned by the columns of Π.

The version of Granger's theorem we apply here is given in Johansen (1989,1991c). The properties of the process X_t under the conditions of reduced rank of Π are given in the next Theorem. In order to formulate the results we define the mean lag matrix $\Gamma = I - \Sigma_1^{k-1}\Gamma_i = \Sigma_1^k i\Pi_i$.

THEOREM. 3.1 *If the roots of the characteristic polynomial of (3.7) are outside the unit disk or at z = 1, if* $\Pi = \alpha\beta'$, *and if* $\alpha_{\perp}'\Gamma\beta_{\perp}$ *has full rank, then the processes* ΔX_t *and* $\beta' X_t$ *can be given initial distributions such that they become stationary. In this case the process* X_t *has the representation*

$$X_t = C\sum_{i=1}^{t} \epsilon_i + C_1(L)\epsilon_t + \tau_0 + \tau_1 t + \tfrac{1}{2}\tau_2 t^2 + P_{\beta}X_0 \qquad (3.12)$$

where $C = \beta_{\perp}(\alpha_{\perp}'\Gamma\beta_{\perp})^{-1}\alpha_{\perp}'$, *and* $\tau_2 = C\mu_1$, *and* $P_{\beta_{\perp}}$ *is the projection on the space spanned by* $\beta_{\perp}$, *that is* $P_{\beta_{\perp}} = \beta_{\perp}(\beta_{\perp}'\beta_{\perp})^{-1}\beta_{\perp}'$.

The autoregressive model allows us to define under which conditions we get I(1) variables and cointegration, thus allowing these assumptions to be checked against the data. We also get a possibility to define the attractor set, and the common trends, and discuss how these spaces can be interpreted. The adjustment coefficients α should be interpreted as the "force of adjustment" of the changes to the errors in the long–run relations, after the lagged changes have been taken into account. The "speed of adjustment" is determined by the roots of the characteristic polynomial, and is a complicated function of the parameters.

3.2.2 Likelihood inference

The analysis of the likelihood function under the assumption of reduced rank was carried out by Anderson (1951) in connection with independent errors, and applied to the analysis of stationary processes by Ahn and Reinsel (1988), and to non–stationary processes by Johansen (1988). The statistical calculations come down to a reduced rank regression of ΔX_t on X_{t-1} corrected for the lagged differences, constant and linear term. Since the parameters are variation free one can first eliminate the lagged differences, the constant and the linear term and obtain residuals R_{0t} and R_{1t}. Then we define the product moment matrices

$$S_{ij} = T^{-1}\Sigma_1^T R_{it} R'_{jt}, \quad i,j = 0,1.$$

The estimation procedure amounts to solving the eigenvalue problem:

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0.$$

We find $\lambda_1 > ... > \lambda_p$ and eigenvectors $v_1,...,v_p$. Note that this eigenvalue decomposition gives the partial canonical correlations between the levels and the changes in the process conditional on the lagged changes. The estimate of the cointegrating space is $sp(\hat{\beta}) = sp(v_1,...,v_r)$ and the likelihood ratio test statistic $Q(\mathscr{H}_r | \mathscr{H}_p)$ for $\mathscr{H}_r$ in $\mathscr{H}_p$, the full VAR model, is

$$Q_r = -2\ln Q(\mathscr{H}_r | \mathscr{H}_p) = -T\Sigma_{r+1}^p \ln(1-\lambda_i).$$

So far the algebra follows Anderson (1951). The new aspect that has interested Ahn and Reinsel (1990), Johansen (1988) and also Phillips (1990) is the non–standard limit distribution of this test statistic which can be expressed as

(3.13) $$\mathrm{tr}\{\textstyle\int_0^1 (dB)F'[\int_0^1 FF'du]^{-1}\int_0^1 F(dB)'\},$$

where B is a p–r dimensional Brownian motion, and F is B with the last component replaced by t^2, and then corrected for the trend. If instead the model with $\mu_0 = \mu_1 = 0$ is considered, then F = B. These distributions are then tabulated by simulation (see Johansen and Juselius (1990) and Johansen (1991f)). Note the similarity with the limit distribution appearing in the above mentioned work of Stock and Watson. The estimate of α is given by $\hat{\alpha} = S_{01}\hat{\beta}$, and $\hat{\Omega} = S_{00} - \hat{\alpha}\hat{\alpha}'$. More interesting is the estimate of $\alpha_\perp$, the common trend coefficients, which is given as $S_{00}^{-1}S_{01}(v_{r+1},...,v_p)$ or as the eigenvectors to the dual eigenvalue problem

$$|\lambda S_{00} - S_{01}S_{11}^{-1}S_{10}| = 0$$

corresponding to the p–r smallest eigenvalues, which are the same as for the other eigenvalue problem. Thus the dual economic concepts of cointegrating relations and common trends are treated in a statistically similar way in the estimation procedure.

The test statistic Q_r can either be used to check prior beliefs about the number of cointegration relations or to estimate the cointegrating rank. The rank can be estimated by the following procedure. Let c_r denote the quantiles calculated from the distribution (3.13), for some level 5% , say. Then we define the estimator $\hat{r}$ by

$$\{\hat{r} = r\} = \{Q_0 > c_0, \ldots, Q_{r-1} > c_{r-1}, Q_r < c_r\}.$$

It is not difficult to show that in the limit $\hat{r}$ takes on the correct value with probability 95%, see Johansen (1991e) for an application to cointegration of an idea due to Pantula (1989).

The next question that we shall discuss is the formulation and testing of hypotheses on the coefficients α and β.

Often the relevant economic question to ask concerns the value of the coefficients of one of the structural equations. How can one do this when the cointegrating relations are only defined up to linear transformations? The economic answer to this is that one has to identify the equations first, in one of the many senses of this word. We here used the word to mean that the cointegrating relations are just identified if we have chosen a coordinate system in which we want to express the results.

The formal definition is to take a matrix c of dimension $p \times r$ and define the normalized cointegrating relations as $\beta_c = \beta(c'\beta)^{-1}$. Thus the vectors are normalized such that $c'\beta_c = I$. If in particular we take $c = (I,0)'$ the normalization corresponds to solving the cointegrating relations for the first r variables. This corresponds to the formulation in (3.6), where the relations have been solved for Y_t. Other normalizations can be expressed by different choices of c.

Once this has been done, hypotheses can be tested on the coefficients as over identifying restrictions. The Wald tests are especially convenient since they do not require re–estimation of the parameters under the null, but of course one can also do the likelihood ratio test if the calculations can be performed.

It follows from the general results on the asymptotic distribution of $\hat{\alpha}$ and $\hat{\beta}$ that the likelihood ratio tests or Wald tests of hypotheses on either α or β are asymptotically distributed as χ^2.

A number of hypotheses have a structure that allows the estimation under the null to be carried out by various modifications of reduced rank regression. An example is that some economic questions can be formulated as linear restrictions on all of the beta vectors simultaneously. For instance the hypothesis of price homogeneity, would mean that coefficients on two prices were equal with opposite sign in all relations. Such a hypothesis can clearly be expresses as a restriction of β without first identifying or normalizing the coefficients. The hypothesis that a certain vector is cointegrating, that is if a given linear combination is stationary, is also an example of a hypothesis that can be tested without normalizing the coefficients first. Examples of the tests of such hypotheses are given in Johansen and Juselius (1991). Note that the hypothesis that a given unit vector is a cointegrating relation is really a test that a given variable is stationary. Thus by using this formulation we have decomposed the question of the stationarity of a given variable into two questions. The first is whether there are cointegrating relations at all, and the second is whether such a vector can be chosen as the given unit vector. Notice that in this last case the null is the null of stationarity, and the asymptotic inference is χ^2.

3.2.3 *The role of the constant term*

Our first goal is to interpret the parameters μ_0 and μ_1 and discuss their influence on the process X_t. From Granger's representation theorem, see (3.12), it follows that X_t has a quadratic trend with the coefficients $\tau_2 = \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp'\mu_1$. Thus the coefficients $\alpha_\perp'\mu_1$ give rise to the quadratic trend which influences the linear combinations $\beta_\perp'X_t$, but $\beta'X_t$ has no quadratic trend only a linear trend.

The model we get under the restriction $\alpha_\perp'\mu_1 = 0$, or equivalently $\mu_1 = \alpha\beta_1$ for some $r\times 1$ vector β_1, the model has the interesting property that it allows a linear trend in the cointegrating relations as well as in the common trends. The analysis of the model is again performed by a reduced rank regression since we can write

$$\alpha\beta'X_t + \alpha\beta_1 t = \alpha(\beta',\beta_1)(X_t',t)' = \alpha\beta^{*\prime}X_t^*,$$

say. Thus by a reduced rank regression of ΔX_t on X_t^* corrected for lagged differences and the constant we can derive estimators of α, β and β_1. The likelihood ratio test for the restriction $\alpha_\perp'\mu_1 = 0$ is found by comparing the eigenvalues from the calculations with and without the restriction, and the asymptotic distribution is given by the χ^2 distribution, see Johansen (1991f).

Another hypothesis that is of interest is concerned with trend stationarity. In the model with no quadratic trend, the cointegrating relations are allowed to have a linear trend. Thus if we want to investigate if a given linear combination of X_t, or any of its components, is trend stationary we can formulate this as a hypothesis on the cointegrating relations as follows: Let b $(p \times r_0)$ denote the linear combinations believed to be trend stationary. The hypothesis is then expressed as

$$\beta^* = (H\varphi, \psi),$$

where

$$H = \begin{bmatrix} b & 0 \\ 0 & 1 \end{bmatrix}$$

is of dimension $(p+1)\times(r_0+1)$, φ is $(r_0+1)\times r_0$ and ψ is $(p+1)\times(r-r_0)$. This way of writing the hypothesis of trend stationarity shows that the calculations can be performed as a switching algorithm between two reduced rank regressions, see Johansen and Juselius (1991) for a discussion of this. The asymptotic distribution of the likelihood ratio test for this hypothesis is χ^2.

Finally one can test if the trend stationary relations $\beta' X_t$ in fact have no trend, by analyzing the model with $\mu_1 = 0$. Again a comparison of the eigenvalues derived by reduced rank regression will give the likelihood ratio test which is asymptotically χ^2.

3.2.4 Partial models

In the autoregressive formulation of $X_t' = (Y_t', Z_t')$ we can interpret a partial model as a conditional model of Y_t given Z_t, that is, as the model given in error correction form

$$\Delta Y_t = \omega \Delta Z_t + \alpha_y \beta' X_{t-1} + \Sigma_1^{k-1} \gamma_i X_{t-i} + \epsilon_{y.zt}. \tag{3.14}$$

where $\omega = \Omega_{yz}\Omega_{zz}^{-1}$, such that $\epsilon_{y.zt} = \epsilon_{yt} - \omega\epsilon_{zt}$ is independent of ϵ_{zt}. If Z_t is weakly exogenous or equivalently if $\alpha_z = 0$ then the analysis of the full VAR model will give the same estimator as the analysis of the partial model, see Johansen (1991g). Thus under some conditions one can determine the cointegration rank from the partial system. The condition is that the number of equations analyzed is larger than the number of cointegration relations and that the variables that we condition on are weakly exogenous. Note that the conditioning variables can cointegrate among themselves.

The problem about weak exogeneity is that it is clearly not enough to assume it, one should also check that it holds. This of course is a bit difficult without analyzing the full system. A simple solution is to perform an F test on the coefficients of $\hat{\beta}' X_{t-1}$ in the z–equations, where $\hat{\beta}$ is the estimate derived from the equation for Y given Z. If the test rejects we are still faced with the real problem, and that is, that there is more information about the cointegrating relations in the equations we are leaving out. If one tries the maximum likelihood estimation of β in the partial system we can still find the asymptotic distribution, but it turns out that the asymptotic distribution depends on nuisance parameters, and that inference is difficult to say the least.

4. The statistical analysis of models for I(2) variables.

While the non–stationarity of economic variables can often be described by I(1) variables there are some variables that are perhaps best described by I(2) processes. Thus if X_t is the log of a price variables, it is clearly non–stationary, but even the inflation rate ΔX_t can be non–stationary. If ΔX_t is described by an I(1) variable, then X_t is I(2). Inference for I(2) variables is being developed and not many results exist. Below we sketch some of the regression type results, and some results for the VAR model for systems.

4.1. Regression models for the long–run parameters

The regression model (3.1) and (3.2) can be formulated for I(d) variables see Stock and Watson (1991). We shall discuss the results for d = 2. For I(2) variables they suggest that the variables in X_t be split into Y_{1t}, Y_{2t} and Y_{3t} such that

$$\Delta^2 Y_{1t} = U_{1t} \tag{4.1}$$

$$\Delta Y_{2t} = \Theta_1 \Delta Y_{1t} + U_{2t} \tag{4.2}$$

$$Y_{3t} = \Theta_2 Y_{2t} + \Theta_3 Y_{1t} + \Theta_4 \Delta Y_{1t} + U_{3t} \tag{4.3}$$

where U_t is an invertible stationary process.

Thus the model assumes that we know the common I(2) trends, Y_{1t} , and that we know that the variables Y_{1t} can cointegrate with Y_{2t}. Finally the variables Y_{3t} can cointegrate with levels Y_{1t}, Y_{2t} and differences ΔY_{2t}.

They focus on the estimation of the last equation and suggest regression estimates based upon (4.3) augmented by leads and lags of ΔY_{1t} and ΔY_{2t}. They find that their estimator not only gives consistent estimation of the parameters of interest but the asymptotic distribution of the estimated coefficients is mixed Gaussian such that usual inference can be performed.

4.2 The autoregressive model

Thus provided the structure of the data is so well understood that a model of the type (4.1), (4.2) and (4.3) can be built, regression type estimators can be used for making inference on hypotheses on the parameters Θ. The VAR model discussed below allows one to analyze the structure of I(2) variables and allows one to check the structure against the data.

4.2.1 Cointegration and representation of I(2) variables

The autoregressive model allows for the analysis of I(2) variables, by suitably restricting the parameter space. Consider the general model (3.7) reparametrized in the error correction form

(4.4) $$\Delta^2 X_t = \Gamma \Delta X_{t-1} + \Pi X_{t-2} + \Sigma_1^{k-2} \Phi_i \Delta^2 X_{t-i} + \mu + \epsilon_t,$$

where we have allowed a constant term in order to discuss various forms of trending behavior. Our first task is to find under what conditions the model allows for I(2) variables. In order to formulate the results we need some notation. We let $\bar{a} = a(a'a)^{-1}$ for any matrix a of full rank, such that $a'\bar{a} = I$, and the projection onto the space spanned by a is $P_a = a\bar{a}' = a(a'a)^{-1}a'$.

The I(1) theory rests on the assumption that $\alpha_\perp' \Gamma \beta_\perp$ has full rank, so for the equations to generate I(2) variables we assume that $\alpha_\perp' \Gamma \beta_\perp = \varphi \eta'$ has reduced rank and define $\alpha_1 = \bar{\alpha}_\perp \varphi$ and $\beta_1 = \bar{\beta}_\perp \eta$, together with $\alpha_2 = \alpha_\perp \varphi_\perp$ and $\beta_2 = \beta_\perp \eta_\perp$. Then $(\alpha, \alpha_1, \alpha_2)$ are mutually orthogonal and span R^p. The same holds for $(\beta, \beta_1, \beta_2)$. Finally we let $\Phi = I - \Sigma_1^{k-2} \Phi_i$. The results are given in Johansen (1991a)

THEOREM 4.1 *If the roots of the characteristic polynomial are either outside the unit disc or at $z = 1$, if there exist matrices α and β* both *$p \times (p-r)$, of full rank and matrices φ and η* both *$(p-r) \times s$, of full rank such that*

(4.5) $$\Pi = \alpha\beta', \ \alpha_\perp' \Gamma \beta_\perp = \varphi\eta', \text{ and } \alpha_2'(\Phi - \Gamma \bar{\beta} \bar{\alpha}' \Gamma)\beta_2 \text{ has full rank,}$$

then $\Delta^2 X_t$, $\beta_1' X_t$ and $\beta' X_t + \bar{\alpha}' \Gamma \bar{\beta}_2 \beta_2' \Delta X_t$ can be given initial distributions such that they become stationary, and X_t has the representation

(4.6) $$X_t = C_2 \sum_{j=1}^{t} \sum_{i=1}^{j} \epsilon_i + C_1 \sum_{i=1}^{t} \epsilon_i + C_2(L)\epsilon_t + \tau_0 + \tau_1 t + \tfrac{1}{2}\tau_2 t^2 + A + Bt,$$

with $C_2 = \beta_2(\alpha_2'(\Phi - \Gamma \bar{\beta} \bar{\alpha}' \Gamma)\beta_2)^{-1}\alpha_2'$, and $\tau_2 = C_2 \mu$. It also holds that $\beta' C_1 \in sp(\alpha_2')$, and $\beta_1' C_1 \in sp(\alpha_1', \alpha_2')$. Here A and B can be determined by the initial conditions and $(\beta, \beta_1)' B = 0$ and $\beta' A + \bar{\alpha}' \Gamma \bar{\beta}_2 \beta_2' B = 0$.

Note that the example (2.1) is a special case of (4.6) with $A = B = 0$ and $Y_t = C_2(L)\epsilon_t$. This representation readily shows the asymptotic properties of the process and how it should be normalized in the different directions exploiting the reduced rank of the matrices C_1 and C_2. Thus the reduced rank of the total multiplier matrix, Π, provides the cointegrating relations $\beta' X_t$, which for I(2) variables are only I(1), but they cointegrate with the differences $\Delta\beta_2' X_t$ with the coefficients $\bar{\alpha}'\Gamma\bar{\beta}_2$. Another way of expressing this result is that the residuals R_t of X_{t-1} corrected for lagged differences are asymptotically stationary. This fits nicely in with the empirical findings in Johansen and Juselius (1990), where it was found that the plots of $\hat{\beta}' R_t$ appear more stationary then the plots of $\hat{\beta}' X_t$.

The second condition in (4.5) can be given the following interpretation; multiply the equation (4.4) by $\alpha_\perp'$ to obtain

$$\Delta^2 \alpha_\perp' X_t = \alpha_\perp' \Gamma \Delta X_{t-1} + \Sigma_1^{k-2} \alpha_\perp' \Phi_i \Delta^2 X_{t-i} + \alpha_\perp' \mu + \alpha_\perp' \epsilon_t.$$

Thus the combinations $\alpha_\perp' X_t$ evolve without taking into account the disequilibrium errors $\beta' X_t$, hence can be called common trends. Now apply the identity

$$I = \beta\bar{\beta}' + \beta_\perp \bar{\beta}_\perp'$$

to decompose $\alpha_\perp' \Gamma$ into

$$\alpha_\perp' \Gamma = \alpha_\perp' \Gamma(\beta\bar{\beta}' + \beta_\perp \bar{\beta}_\perp') = \alpha_\perp' \Gamma \beta\bar{\beta}' + \varphi\eta' \bar{\beta}_\perp'.$$

giving the equation

$$(4.7) \qquad \Delta^2 \alpha_\perp' X_t = (\alpha_\perp' \Gamma\beta)\bar{\beta}' \Delta X_{t-1} + \varphi\eta' \bar{\beta}_\perp' \Delta X_{t-1} + \Sigma_1^{k-2} \alpha_\perp' \Phi_i \Delta^2 X_{t-i} + \alpha_\perp' \mu + \alpha_\perp' \epsilon_t.$$

Thus φ and η are given a similar interpretation as α and β only for the differences of the original process.

Thus the structure of the I(2) model can be summarized by saying hat $\beta_2' X_t$ are the common I(2) trends, while $\beta_1' X_t$ are the common I(1) trends. The combinations $\beta' X_t$ represent cointegrating relations although only of type C(2.1). They can be made stationary by introducing the common trends $\beta_2' \Delta X_t$. Comparing with the model (4.1), (4.2), and (4.3) we see that we can choose $Y_1 = \beta_2' X_t$, $Y_{2t} - \Theta_1 Y_{1t} = \beta_1' X_t$ and $Y_{3t} - \Theta_2 Y_{2t} - \Theta_3 Y_{3t} = \beta' X_t$.

4.2.2. The role of the constant term

The model (4.4) allows for a quadratic trend with leading term $\frac{1}{2}C_2\mu t^2$, and the hypothesis $C_2\mu = 0$ or $\alpha_2'\mu = 0$ is the hypothesis of no quadratic trend. Note, however, that under this hypothesis it also holds that $\beta' X_t$ has no linear trend, since $\beta' C_1 \in sp(\alpha_2')$. Thus under the hypothesis that $\alpha_2'\mu = 0$, the quadratic trend disappears, and the linear trend is restricted to the combinations β_1 and β_2. If also $\alpha_1'\mu = 0$, then the linear trend in β_1 disappears, and only $\beta_2' X_t$ has a linear trend. A similar discussion can be carried out if a linear term is inserted in the model, but this will not be attempted here.

4.2.3. Likelihood analysis

The analysis of the likelihood function corresponding to the model (4.4) is not so simple, see Johansen (1990). We shall here present an analysis that suggests itself on the basis of equation (4.7). If one introduces $U_t = \Delta X_t$, and considers α and β known, then (4.7) is a cointegrating model for I(1) variables U_t and the analysis of reduced rank after some preliminary regressions is as discussed in section 3. The only problem of course is that in practice α and β are unknown. It can be proved, that if α and β are estimated by the reduced rank regression described in section 3, that is, as if Γ is unrestricted by (4.5), then one can proceed with the analysis of (4.7) as if α and β were known and equal to the estimates just obtained.

Thus a simple analysis of the cointegration ranks r and s is the following:

1) First determine the rank r and the parameters α and β by a reduced rank regression of $\Delta^2 X_t$ on X_{t-2} corrected for lagged differences and the constant. That is, analyze the model (4.4) with Γ unrestricted.

2) For the given estimates of r, α and β determine s, φ and η by reduced rank regression of $\alpha_\perp' \Delta^2 X_t$ on $\Delta \beta_\perp' X_{t-1}$ corrected for lagged second differences, $\bar{\beta}' \Delta X_{t-1}$, and

the constant.

3) It then holds that asymptotic inference about α, β, φ and η can be conducted by means of the χ^2 distribution. The details can be found in Johansen (1991b) and an illustrative example is given in Johansen (1991d).

5. References

Ahn, S.K. & G.C. Reinsel (1988): "Nested reduced–rank autoregressive models for multiple time series," *Journal of the American Statistical Association* 83, 849–856.

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.

Campbell, J.Y. and P. Perron. (1991): "Pitfalls and opportunities: What macroeconomists should know about unit roots," Princeton Discussion paper.

Dunsmuir, W. and E.J. Hannan (1976): "Vector linear time series models," *Advances in Applied Probability* 8, 339–364.

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

Engle, R.F. & B.S. Yoo, (1989): "Cointegrated economic time series: A survey with new results," UCSD Discussion paper.

Granger, C.W.J. (1981): " Some Properties of Time Series Data and their Use in Econometric Model Specification," *Journal of Econometrics* 16, 121–130.

Granger, C.W.J. (1983): "Cointegrated Variables and Error Correction Models," UCSD Discussion Paper, 83–13a.

Granger, C.W.J. & T–H. Lee (1989): "Investigations of Production, Sales and Inventory Relationships using multi cointegration and non symmetric error correction models," *Journal of Applied Econometrics* 4, 145–159.

Gregoire, S. & G. Laroque (1991): "Multivariate integrated time series: A general error correction representation with associated estimation and test procedures," Discussion paper, INSEE

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, University of Copenhagen.

Johansen, S. (1990): "An algorithm for estimating the cointegration relations in vector autoregressive processes allowing for I(2) variables," University of Copenhagen.

Johansen, S. (1991a): "A representation of vector autoregressive processes integrated of order 2," to appear in *Econometric Theory.*

Johansen, S. (1991b): "A statistical analysis of cointegration for I(2) variables," University of Copenhagen.

Johansen, S. (1991c): "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," to appear in *Econometrica.*

Johansen, S. (1991d): "Test of weak exogeneity for cointegration parameters and test for the order of non–stationarity in the UK money demand data," to appear in *Journal of Policy Modeling.*

Johansen, S. (1991e): "Determination of cointegration rank in the presence of a linear trend," to appear in *Oxford Bulletin of Economics and Statistics.*

Johansen, S. (1991f): "The role of the constant term in cointegration analysis of non stationary variables," Preprint University of Copenhagen.

Johansen, S. (1991g): "Cointegration in partial systems and the efficiency of single equation analysis," to appear in *Journal of Econometrics.*

Johansen, S. & 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. & K. Juselius (1991): "Structural tests in a multivariate cointegration analysis of the PPP and UIP for UK," to appear in *Journal of Econometrics.*

Pantula, S.G. (1989): "Testing for unit roots in time series data," *Econometric Theory* 5, 256–271.

Phillips, P.C.B. (1988): "Spectral Regression for cointegrated Time Series," Cowles Foundation No. 872.

Phillips, P.C.B. (1990): "Optimal inference in cointegrated systems," Econometrica 59, 283–306.

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

Park, J.Y.(1988): "Canonical cointegrating regressions," Cornell University

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.

Saikkonen, P. (1989): "Asymptotically efficient estimation of cointegrating regressions," Preprint University of Helsinki

Sims, C.A. (1980): "Macroeconomics and Reality," *Econometrica* 48, 1–48..

Sims, C.A., J.H. Stock, and M.W. Watson (1990): " Inference in Linear Time Series Models with some Unit Roots," *Econometrica* 58, 113–144

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.

Stock, J.H., and M.W. Watson (1991): "A simple estimator of cointegrating vectors in higher order integrated systems," Discussion paper Berkeley.

Reinsel, G.C. and Ahn, S.K. (1990): "Vector AR models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting," University of Wisconsin–Madison

PREPRINTS 1990

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, DK-2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 31 35 31 33.

No. 1 Johansen, Søren and Juselius, Katarina: Some Structural Hypotheses in a Multivariate Cointegration Analysis of the Purchasing Power Parity and the Uncovered Interest Parity for UK.

No. 2 Tjur, Tue: Analysis of Variance and Design of Experiments.

No. 3 Johansen, Søren: A Representation of Vector Autoregressive Processes Integrated of Order 2.

No. 4 Johansen, Søren: Cointegration in Partial Systems and the Efficiency of Single Equation Analysis.

PREPRINTS 1991

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, DK-2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 31 35 31 33.

No. 1 Johansen, Søren: Determination of Cointegration Rank in the Presence of a Linear Trend.

No. 2 Johansen, Søren: A Statistical Analysis of Cointegration for I(2) Variables.

No. 3 Jacobsen, Martin: Homogeneous Gaussian Diffusions in Finite Dimensions.

No. 4 Johansen, Søren: Testing Weak Exogeneity and the Order of Cointegration in UK Money Demand Data.

No. 5 Johansen, Søren: An I(2) Cointegration Analysis of the Purchasing Power Parity between Australia and USA.

No. 6 Johansen, Søren: Estimating Systems of Trending Variables.