# Forecasting and Interpolation Using Vector Autoregressions with Common Trends

F. J. FERNÁNDEZ MACHO,
A. C. HARVEY, J. H. STOCK *

**ABSTRACT.** – A modification of the vector autoregressive model is to include a stochastic trend component in each equation. It is argued that this formulation will lead to a more parsimonious model than traditional vector autoregressions formulated in terms of levels or differences. Common trends, or factors, may be introduced into the model. This leads to certain of the variables being co-integrated and, as shown in GRANGER and ENGLE [1987], the model then has an error correction representation. Estimation of the model can be carried out by casting it in state space form and applying the Kalman filter. This enables estimation to be carried out for a very general situation in which observations may be missing, temporally aggregated or observed at different time intervals. The common trends may also be extracted using smoothing techniques. Missing observations can also be estimated and the model is likely to be useful if this is the main objective.

## Prévision et interpolation utilisant des modèles autorégressifs vectoriels avec des tendances communes

**RÉSUMÉ.** – Une modification du modèle VAR consiste à introduire une tendance aléatoire dans chaque équation. Il est montré que cette formulation conduit à un modèle plus parcimonieux que les traditionnels modèles VAR en niveaux ou en accroissements. Des tendances communes, ou des facteurs, peuvent être introduits dans le modèle. Certaines variables sont co-intégrées et le modèle a une représentation à corrections d'erreurs. L'estimation du modèle est faite à l'aide de la forme espace d'états et du filtre de Kalman. Cela permet une estimation dans un cadre général (observations manquantes, agrégation temporelle, observations à intervalles différents). Les tendances communes peuvent aussi être estimées par des techniques de lissage. Les observations manquantes peuvent être estimées.

* F. J. FERNÁNDEZ MACHO, A. C. HARVEY: London School of Economics, Houghton Street, London WC2A 2AE, U.K.; J. H. STOCK: Harvard University, Cambridge MA 02138, USA.
HARVEY would like to thank the ESRC for support under the grant Multivariate Aspects of Structural Time Series Models'. STOCK's work was carried out under NSF Grant SES-84-08797. FERNÁNDEZ MACHO was supported by the Suntory-Toyota Foundation. We would like to thank an anonymous referee for helpful comments on an earlier draft.

# 1 Introduction

Vector autoregressions have, in recent years, proved very popular for forecasting economic time series. However the formulation of such models is not altogether satisfactory for dealing with nonstationary time series. If differences are taken the length of the autoregressions may be quite long, leading to models with a very large number of parameters. Conversely, if the model is estimated in levels, the roots of the autoregressive polynomial will lie on the unit circle and this can be difficult to handle statistically. More importantly it may again lead to a model with a large number of parameters. Alternatively, vector ARIMA models could be used. However it may be argued that the ARIMA class becomes too big when considering vector processes so that in practice model specification can be complicated and possibly misleading. Besides, the way in which nonstationarity is dealt with in the ARIMA framework is not satisfactory in the present context since after differencing the series, it is difficult to obtain estimates of the trend components. Furthermore, if some of the trends are common, important relationships between the variable levels will be lost, while the appearence of noninvertibility can complicate the statistical analysis.

In this article we introduce a class of multivariate models which contain stochastic trend as well as autoregressive components. Such models may provide a more satisfactory and more parsimonious treatment of nonstationary time series. Furthermore, it is possible to introduce an additional feature into the model, namely common trends. If common trends are present in the model, some, or all, of the variables are co-integrated and an error correction representation is possible; see GRANGER and ENGLE [1987].

Time domain evaluation of the likelihood function for vector autoregressive models with (common) stochastic trends is carried out by means of the Kalman filter. This likelihood function must then be maximised numerically. While this estimation procedure is clearly more complicated than the OLS regressions required for a pure vector autoregression, it does at least allow missing and temporally aggregated observations to be handled at little extra cost. This in turn means that it can cope with situations where, for example, some variables may be measured on a monthly basis, while others are quarterly. The purpose of Section 3 is to set down an algorithm for dealing with these features for data which include both stocks and flows.

The ability to allow for missing and temporally aggregated observations widens the range of data sets that can be used in multivariate forecasting models. It also gives the models another purpose, namely that they can be used for *interpolation* when there are missing observations on stock variables, and *distribution* when flow variables have been subject to temporal aggregation or stock variables have been averaged over several time periods. Once a model has been estimated, interpolation and distribution is carried out by a smoothing algorithm.

Earlier treatments of the estimation of missing observations, such as CHOW and LIN [1971], are based on single equation methods and make strong, and in our view arbitrary, assumptions about exogeneity. The method suggested here does not rely on such assumptions.

# 2 Vector Autoregressions with Stochastic Trends

A vector autoregression with stochastic trends may be written as

$$y_t = \mu_t + \Phi^{-1}(L)\,\varepsilon_t, \qquad t = 1, \ldots, T, \tag{1}$$

$$\mu_t = \mu_{t-1} + \beta + \eta_t, \tag{2}$$

where $y_t$ is an $N \times 1$ vector of observations, $\Phi(L)$ is an $N \times N$ matrix polynomial in the lag operator, i. e.

$$\Phi(L) = I + \Phi_1 L + \ldots + \Phi_p L^p, \tag{3}$$

$\varepsilon_t$ and $\eta_t$ are multivariate white noise processes with zero mean vectors and covariance matrices $\Sigma_\varepsilon$ and $\Sigma_\eta$ respectively, $\mu_t$ is an $N \times 1$ vector of stochastic trends and $\beta$ is an $N \times 1$ vector of constants. It is assumed that $\varepsilon_t$ and $\eta_t$ are uncorrelated with each other in all time periods. It is also assumed that the roots of $\Phi(L)$ lie outside the unit circle, so that if $\mu_t$ were not present in the model, $y_t$ would be stationary. The model may be extended to allow for $\beta$ being time-varying. This is done in sub-section 2.2, but for many macroeconomic time series on real variables $\beta$ can be taken to be constant and so we will concentrate on this case. Seasonal components can also be brought into the model but we will not deal with this feature explicitly, since the general principles remain the same.

Model (1)-(2) expresses the observations as the sum of a trend component and a short term component, which may exhibit pseudo-cyclical behaviour due to the interactions between the variables induced by $\Phi(L)$. Such pseudo-cyclical behaviour may also be incorporated in the trend. Thus (2) may be extended to:

$$\mu_t = \mu_{t-1} + \beta + \Phi_\mu^{-1}(L)\,\eta_t, \tag{4}$$

where $\Phi_\mu(L)$ is an $N \times N$ polynomial of order $p_\mu$ defined in a similar way to (3). The analysis of univariate series, as in Harvey [1985], suggests that cyclical behaviour may well be incorporated within the trend itself.

It is worth drawing attention to some special cases of (1)-(4). First, if $\Sigma_\eta = 0$, the model collapses to a stationary vector autoregression about a set of deterministic linear trends. Secondly, if $\Sigma_\varepsilon = 0$, the differenced observations follow a stationary vector autoregression. Both cases may be handled by the Kalman filter.

## 2.1. Common Trends

Common trends may be introduced into (1) by defining an $N \times K$ matrix, $\Theta$, where $K \leqslant N$, and re-writing (1) as

(5) $$y_t = \Theta \mu_t + \mu_0 + \Phi^{-1}(L)\,\varepsilon_t.$$

The $N \times 1$ vector $\mu_0$ has zeroes for its first K elements while its last N-K elements consists of an unconstrained vector $\bar{\mu}$. The trend components, $\mu_t$, are still generated by (2) or (4) but with the difference that $\mu_t$, $\beta$ and $\eta_t$ are now of length K. This is the number of common trends in the model. As it stands, the model represented by (5) is not identifiable since for any nonsingular $K \times K$ matrix P the $N \times K$ matrix $\Theta$ and the trend components $\mu_t$ could be redefined as $\Theta P^{-1}$ and $P\mu_t$ respectively. In other words there is an infinite number of parameter sets for which the model would generate identical time series $\{ y_t \}$, and, therefore, it is necessary to choose one member within each equivalence class so that the structure of the model can be estimated.

In order for (5) to be identifiable, it is necessary to place restrictions on $\Sigma_\eta$ and $\theta$. There are a number of ways in which this may be done. We propose setting $\Sigma_\eta$ equal to a diagonal matrix while $\Theta$ is such that $\theta_{ij}=0$ for $j>i$, and $\Theta_{ii}=1$, $i=1, \ldots, K$. Note that when $K=N$ the model reduces to (1) since $\Theta \mu_t$ in (5) may be re-defined as $\mu_t$ and the variance of the corresponding $\eta_t$ is then the p. s. d. symmetric matrix $\Theta\Sigma_\eta\Theta'$.

The common trends model has the important property that N-K linear combinations of the $y_t$ vector are stationary even though all the individual elements of $y_t$ are only stationary in first differences. In the terminology of GRANGER and ENGLE [1987] the model is said to be co-integrated of order (1,1). GRANGER and ENGLE [1987] show in turn that this implies an error-correction representation of the kind adopted by SARGAN [1964] and DAVIDSON *et al.* [1978].

The co-integrating vectors are the N-K rows of an $(N\text{-}K) \times N$ matrix A which have the property that $A\Theta = 0$. Hence

$$A\,y_t = A\,\Phi^{-1}(L)\,\varepsilon_t + A\,\mu_0$$

and $A\,y_t$ is an $(N\text{-}K) \times 1$ stationary process.

A model with common trends is not only more parsimonious than a model without common trends, but also preserves certain levels relationships between the variables in forecasting. In addition the trends themselves may have an interesting interpretation. For this purpose it may be useful to consider a rotation of the estimated trends. Suppose that $\Phi_\mu(L)=I$ and that the estimated $\Sigma_\eta$ is p. d. and define $\mu_t^+ = H\Sigma_\eta^{-1/2}\mu_t$, $\eta_t^+ = H\Sigma_\eta^{-1/2}\eta_t$, $\Theta^+ = \Theta\Sigma_\eta^{1/2}H'$, and $\beta^+ = H\Sigma_\eta^{-1/2}\beta$, where H is an orthogonal matrix. The model may now be written

(6.*a*) $$y_t = \Theta^+\mu_t^+ + \Phi^{-1}(L)\,\varepsilon_t,$$

(6.*b*) $$\mu_t^+ = \mu_{t-1}^+ + \beta^+ + \eta_t^+,$$

with $\mathrm{Var}(\eta_t^+) = I$ for any choice of H. This H may be used to redefine the common trends so as to give the desired interpretation. The trends remain independent of each other for any choice of H.

## 2.2. Slope Components

The specifications of the common trends can be extended so as to include stochastic slopes. Thus

$$(7.a)\qquad \mu_t = \mu_{t-1} + \Theta_\beta \beta_{t-1} + \beta_0 + \eta_t,$$

$$(7.b)\qquad \beta_t = \beta_{t-1} + \zeta_t,$$

where $\beta_t$ is $K_\beta \times 1$, $0 \leqq K_\beta < K$, $\mathrm{Var}(\zeta_t) = I$, $\Theta_\beta$ is a $K \times K_\beta$ matrix with $\theta_{\beta,ij} = 0$ for $j > i$. The disturbance vectors $\eta_t$ and $\zeta_t$ are mutually uncorrelated. Finally the $K \times 1$ vector $\beta_0$ has its first $K_\beta$ elements zero, so that $\beta_0 = (0' \bar{\beta}')'$.

Combining (7) with (5) leads to a model which is co-integrated of order (2,2). Provided $\Theta$ and $\Theta_\beta$ contrain no null rows, each element of the $y_t$ vector needs to be differenced twice to make it stationary; in Granger's terminology it is integrated of order two, I(2). Thus there is no necessity to require that $\Sigma_\eta$ be p. d. If some of the rows of $\Theta_\beta$ are null the system can still be handled in the time domain. Furthermore since some variables are I(2) while others are I(1) it follows that there must be at least two common trends.

## 2.3. A Simple Illustration

Consider two variables, say income and consumption, which are known to be co-integrated. This constitutes a system with $N=2$ and $K=1$. With a stationary first-order vector AR process this yields the model

$$(8.a)\qquad \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 \\ \theta \end{bmatrix} \mu_t + \begin{bmatrix} 1-\varphi_{11}L & -\varphi_{12}L \\ -\varphi_{21}L & 1-\varphi_{22}L \end{bmatrix}^{-1} \times \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{\mu} \end{bmatrix},$$

$$(8.b)\qquad \mu_t = \mu_{t-1} + \beta + \eta_t,$$

with $\mathrm{Var}(\eta_t) = \sigma_\eta^2$. Note that even with $p=1$, the vector AR process is capable of generating pseudo-cyclical behaviour.

The co-integrating vector, A, may be normalized as $A = (1\ \alpha)$. It must be such that

$$1 + \alpha\theta = 0$$

and so

$$\alpha = -1/\theta.$$

Multiplying (8) through by A gives

$$(9)\qquad y_{1t} = (1/\theta)\, y_{2t} + (-1/\theta)\,\bar{\mu} + u_t,$$

where $u_t$ is a stationary ARMA (2,1) process.

Thus there is a levels relationship between $y_{1t}$ and $y_{2t}$. Applying OLS to (9) gives a consistent estimator of $\theta$ even though $y_{2t}$ is endogenous. This is essentially because $y_{1t}$ and $y_{2t}$ are both I(1) while the error term is stationary; see Stock [1984].

# 3 Statistical Treatment

It is relatively straightforward to put model (1)-(4) in state space form. The likelihood function may then be evaluated in the prediction error decomposition form using the Kalman filter, and maximised by numerical optimisation with respect to the unknown parameters in $\Theta$, $\Sigma_\varepsilon$, $\Phi(L)$ and $\Phi_\mu(L)$. An alternative approach is to carry out estimation in the frequency domain along the lines suggested in FERNÁNDEZ MACHO [1986]. Once the parameters have been estimated, prediction and smoothing is straightforward to carry out using standard techniques; see HARVEY [1981] or ANDERSON and MOORE [1979].

The treatment below generalises the model to situations where observations may be missing, temporally aggregated or observed at different timing intervals. A general form of the Kalman filter is then employed. Such techniques are also applicable in pure vector autoregressions without the trend components and so are of quite general interest.

Consider model (5) but suppose that the full $y_t$ vector is not necessarily observed in all time periods. The variables generated by (5) will therefore be denoted by $y_t^+$, so that (5) is actually written

$$y_t^+ = \Theta\,\mu_t + \Phi^{-1}(L)\,\varepsilon_t \tag{10}$$

with $\mu_t$ defined, as before, in (2). At any particular point in time $t$, $t=1,\ldots,T$, we may not observe some of the variables in $y_t^+$. In the case of flow variables or time averaged stock variables, we may only observe the sum of the current and some past values of a variable. This latter phenomenon is known as temporal aggregation. In order to simplify matters we will assume that when temporal aggregation takes place for different variables it is always over the same $\delta$ time periods. Thus if we have mixed monthly and quarterly observations $\delta$ is equal to three.

At time $t$ we observe an $N_t \times 1$ vector $y_t$. This can be written $y_t = (y_t^{s\prime}\; y_t^{f\prime})'$ where the $N_t^s \times 1$ vector $y_t^s$ and the $N_t^f \times 1$ vector $y_t^f$ are defined as follows:

(1) $y_t^s$ contains all the variables, stocks and flows, which are observed in all time periods, together with the variables, normally stocks, which contain missing observations. The relationship between $y_t^s$ and $y_t^+$ is determined by an $N_t \times N_t^s$ selection matrix $Z_t^s$, i. e.

$$y_t^s = Z_t^s y_t^+, \qquad t=1,\ldots,T; \tag{11.a}$$

(2) $y_t^f$ contains all the variables which are subject to temporal aggregation. In this case

$$y_t^f = Z_t^f y_t^c, \qquad t=1,\ldots,T, \tag{11.b}$$

where $y_t^c$ is an $N^c \times 1$ vector which contains the cumulated values of all the $N^c$ variables subject to temporal aggregation and $Z_t^f$ is an $N_t^f \times N^c$ selection matrix. If cumulated variables are observed at the points $t=\delta, 2\delta, 3\delta, \ldots$, then

$$(12) \qquad y_t^c = Z^c \sum_{j=0}^{\delta-1} y_{t-j}^+, \qquad t=\delta, 2\delta, \ldots,$$

where $Z^c$ is an $N^c \times N$ selection matrix. If all the variables in $y_t^c$ are observed at $t=\delta, 2\delta, \ldots$, then $N_t^f = N^c$ and

$$Z_t^f = \begin{cases} I, & t=\delta, 2\delta, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

If some of the elements of $y_t^c$ are missing at a given point in time then $N_t^f < N^c$ and the appropriate rows are removed from $Z_t^f$ at $t=\delta, 2\delta, \ldots$

Some variables may initially be subject to temporal aggregation, but may later be observed in every time period. At the point where they are observed every time period such variables switch from being in $y_t^f$ and $y_t^c$ to being in $y_t^s$. This could be accounted for by putting time subscripts on $N^c$ and $Z^c$.

We now consider the state space representation for the model defined by (2), (5), (11) and (12). The stochastic trend, (2), can be generalised to include autoregressive components, as in (4), quite straightforwardly and so will not be dealt with explicitly.

The $p$-th order vector autoregressive process $\Phi^{-1}(L)\varepsilon_t$ in (5) can be written as a first-order vector autoregression by defining an $N_p \times 1$ state vector $\alpha_t^+$ such that

$$(13) \qquad \alpha_t^+ = T^+ \alpha_{t-1}^+ + R^+ \varepsilon_t,$$

where $T^+$ is an $N_p \times N_p$ matrix containing the element of $\Phi_1, \ldots, \Phi_p$ and $R^+ = (I_N\, 0)'$; i. e.

$$(14) \qquad \alpha_t^+ = \begin{bmatrix} \Phi_1 \ldots \Phi_p \\ \text{- ---- --} \\ \quad\quad 0 \\ I \quad \vdots \\ \quad\quad 0 \end{bmatrix} \alpha_{t-1}^+ + \begin{bmatrix} I \\ \text{---} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \varepsilon_t.$$

The first N elements of $\alpha_t^+$ are the elements of $\Phi^{-1}(L)\varepsilon_t$ and these are extracted from $\alpha_t^+$ by defining $Z^\alpha = (I \quad 0)$ and writing

$$(15) \qquad \Phi^{-1}(L)\varepsilon_t = Z^\alpha \alpha_t^+.$$

The state vector for the full model is given by

$$(16) \qquad \alpha_t = (\mu_t' \; \alpha_t^{+\prime} \; y_t^{c\prime})'.$$

Equation (2) is already effectively in the form of a transition equation. The transition equation for $y_t^c$ can readily be derived from (12), (14) and (2). First note that (12) can be written as

$$(17) \qquad y_t^c = Z^c y_t^+ + \psi_t y_{t-1}^c, \qquad t=1, \ldots, T,$$

where

$$\psi_t = \begin{cases} 0, & t=\delta+1,\ 2\delta+1,\ 3\delta+1, \ldots, \\ 1, & \text{otherwise,} \end{cases}$$

and $y_0^c = 0$.

Using (10), (13), (2) and (17), $y_t^c$ can be written,

$$y_t^c = Z^c \Theta \beta + Z^c \Theta \mu_{t-1} + Z^c Z^\alpha T^+ \alpha_{t-1}^+ + Z^c \Theta \eta_t + Z^c Z^\alpha R^+ \varepsilon_t + \psi_t y_{t-1}^c, \tag{18}$$

and combining (2), (13) and (18),

$$\begin{bmatrix} \mu_t \\ \alpha_t^+ \\ y_t^c \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & T^+ & 0 \\ Z^c \Theta & Z^c Z^\alpha T^+ & \psi_t \end{bmatrix} \times \begin{bmatrix} \mu_{t-1} \\ \alpha_{t-1}^+ \\ y_{t-1}^c \end{bmatrix} + \begin{bmatrix} \beta \\ 0 \\ Z^c \Theta \beta \end{bmatrix} + \begin{bmatrix} I & 0 \\ 0 & R \\ Z^c \Theta & Z^c Z^\alpha R^+ \end{bmatrix} \times \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix}. \tag{19}$$

Using obvious notation, (19) can be rewritten more compactly as

$$\alpha_t = T \alpha_{t-1} + \gamma + R \upsilon_t, \tag{20}$$

where

$$E(\upsilon_t \upsilon_t') = \begin{bmatrix} \Sigma_\eta & 0 \\ 0 & \Sigma_\varepsilon \end{bmatrix} \tag{21}$$

and $\Sigma_\eta$ is a diagonal matrix, as discussed previously.

The measurement equation, relating the state vector to the observed variable $y_t$, is obtained by rewriting (11) in terms of the state vector $\alpha_t$. First, note that

$$y_t^+ = \Theta \mu_t + \mu_0 + Z^\alpha \alpha_t^+. \tag{22}$$

Thus the measurement equation is

$$y_t = \mu_{0\,t}^s + Z_t \alpha_t, \tag{23}$$

where

$$\mu_{0\,t}^s = \begin{bmatrix} Z_t^s \mu_0 \\ \text{-----} \\ 0 \end{bmatrix}$$

and

$$Z_t = \left[\begin{array}{ccc} Z_t^s \Theta & Z_t^s Z^\alpha & 0 \\ \hline 0 & 0 & Z_t^f \end{array}\right].$$

Predictions of future observations can be made on the basis of the state space form (20) and (23). Estimates of the common trends can be made by a smoothing algorithm. Initial conditions for the filter can be obtained by using the unconditional mean and covariance matrix of $\alpha_t$, by noting that $y_0^c=0$, and by imposing a diffuse prior on $\mu_0$; cf. ANSLEY and KOHN [1985]. As regards estimation of the unknown parameters in $\Theta$, $\Phi$, $\Sigma_\eta$ and $\Sigma_\varepsilon$, this may be carried out *via* the prediction error decomposition, using the Kalman filter. The drift parameters, $\beta$, may be concentrated out of the likelihood function by noting that, conditional on $\Theta$, the ML estimator of $\beta$ is

$$\tilde{\beta}(\Theta)=(\Theta'\Theta)^{-1}\Theta'\overline{\Delta y} \tag{24}$$

where $\overline{\Delta y}=(y_T-y_1)/(T-1)$; see FERNÁNDEZ MACHO [1986]. The elements of $\bar{\mu}$ may also be concentrated out of the likelihood function, either by using the device suggested in KOHN and ANSLEY [1985] or by including $\bar{\mu}$ in an augmented state vector.

## • References


ANDERSON, B. D. O. and MOORE, J. B. (1979). – *Optimal Filtering*, Englewood Cliffs: Prentice-Hall.

ANSLEY, C. F. and KOHN, R. (1985). – "Estimation, Filtering and Smoothing in State Space Models with incompletely Specified Initial Conditions", *Annals of Statistics*, 13, p. 1286-1316.

CHOW, G. C. and LIN, A. (1971). – "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series", *Review of Economics and Statistics*, 53, p. 372-375.

DAVIDSON, J. E. H., HENDRY, D. F., SRBA, F. and YEO, J. S. (1978). – "Econometric Modelling of the Aggregate Time Series Relationship Between Consumers' Expenditure and Income in the United Kingdom", *Economic Journal*, 88, p. 661-692.

ENGLE, R. F. and GRANGER, C. W. J. (1987). – "Co-Integration Error-Correction: Representation, Estimation and Testing", *Econometrica*, 55, p. 251-276.

FERNÁNDEZ MACHO, F. J. (1986). – "Estimation and Testing of Multivariate Structural Time Series Models", *Ph. D. Thesis*, University of London.

HARVEY, A. C. (1981). – *Time Series Models*, Philip Allan (ed.), Deddington, Oxford and Humanities Press, Atlantic Highlands, N.J.

HARVEY, A. C. (1985). – "Trends and Cycles in Macroeconomic Time Series", *Journal of Business and Economic Statistics*, 3, p. 216-227.

KOHN, R. and ANSLEY, C. (1985). – "Efficient Estimation and Prediction in Time Series Regression Models", *Biometrika*, 72, p. 694-697.

SARGAN, J. D. (1964). – "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", *Econometric Analysis for National Economic Planning*. P. E. HART, MILLS, G. and WHITAKER, J. K. Eds., Butterworths, London.

STOCK, J. (1984). – "Asymptotic Properties of Least Squares Estimators of Co-integrating Vectors", *manuscript*, Harvard University.