Haavelmo's Probability Approach and the Cointegrated VAR*

Katarina Juselius

University of Copenhagen, Department of Economics

May 7, 2013

Abstract

Some key econometric concepts and problems of great importance to Trygve Haavelmo and Ragnar Frisch are discussed within the general framework of a cointegrated VAR. The focus is on problems typical of time-series data such as multicollinearity, spurious correlation and regression, time dependent residuals, model selection, missing variables, simultaneity, autonomy and identification. The paper argues that the more recent development of unit root econometrics has been instrumental for a solution to the above problems.

*Valuable comments from Kevin Hoover, Roman Frydman, Søren Johansen and two anonymous referees are gratefully acknowledged. I also thank Olav Bjerkholt for his insightful comments and for making me aware of unpublished work by T. Haavelmo.

Keywords: Haavelmo, CVAR, autonomy, identification, passive observations

JEL classification: B16, B31, B41, C32, C82

1 Introduction

Haavelmo's Nobel prize winning Econometrica monograph "The Probability Approach to Econometrics" from 1944 is considered to have laid the foundations for modern econometrics. See for example Morgan (1995) and the survey articles in Aldrich (1989), Hendry, Spanos, and Ericsson (1989), and Anderson (1992). In today's econometric world, new concepts, tests and estimators are developed side by side with empirical applications. This was less so when Haavelmo developed his probability approach. Empirical analyses which now can be done within seconds would then have required years of work, if at all possible. Given these obstacles, his vision about econometric modelling using time series data obtained by "passive observation" is truly remarkable.

The idea of the paper is to address a number of key econometric concepts in Haavelmo's work and discuss them in the context of a Cointegrated VAR (CVAR) model (Johansen, 1996). The focus is on problems typical of macroeconomic data such as multicollinearity, spurious correlation and regression, time dependent residuals, normalization, reduced rank, simultaneity, autonomy and identification. The paper argues that these problems were well understood but not satisfactorily solved and that it is the more recent development of the theory of nonstationary processes (Phillips,1987, Johansen, 1988, 1996) that has allowed a major econometric breakthrough in this respect. In particular unit-root econometrics seems to have been instrumental for a solution of the above problems. This was an important reason for choosing the CVAR, while recognizing that other econometric approaches may provide equally good solutions. Another reason is that the CVAR can represent Haavelmo's vision of a joint probability formulation of the observables and, hence, of a likelihood based approach (Johansen, 1996). It also offers a practical and well-worked out methodology for analyzing nonstationary data (Juselius, 2006).

Haavelmo's work was strongly influenced by Ragnar Frisch and it is difficult to ignore Frisch when discussing the above problems. Many of the basic concepts were first formulated by Frisch and later reformulated by Haavelmo within his probability approach. In the words of Aldrich (1989, p.1): "The concepts relating to structure were devised by Frisch and they passed into classical econometrics through Trygve Haavelmo's Probability Approach in Econometrics (1944)." The fact that Frisch was not fully convinced that a joint probability formulation is a solution to the above econometric problems makes it even more interesting to re-address them today in the probability framework of a CVAR.

One of Haavelmo's important contributions is to have provided us with a coherent framework for addressing methodological problems relevant for the analysis of data by passive observations. This has in particular been important for macroeconomic data. The discussions in this paper are strictly restricted to such problems.

2 Haavelmo's probability approach

Consider a time series of p variables, $x_{1,t} = f(x_{2,t}, \cdots, x_{p,t}),\ t = 1,..,T$, where all variables, $x_{i,t}\ \ i = 1,..,p$, are stochastic. From the economic theory point of view, $x_{1,t}$ is the variable of interest and $x_{2,t}, \cdots, x_{p,t}$ are the explanatory variables, some of which might be exogenously given. At each point in time, t, there is just one realization, $x_t' = [x_{1,t}, x_{2,t}, \cdots, x_{p,t}]$ of the underlying stochastic process. The sample of observations is given by:

$$X = \begin{bmatrix} x_{1,1} & x_{2,1} & \cdots & x_{p,1} \\ x_{1,2} & x_{2,2} & \cdots & x_{p,2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,T} & x_{2,T} & \cdots & x_{p,T} \end{bmatrix} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{bmatrix}$$

Both Haavelmo and Frisch were concerned about the fact that successive observations of typical macroeconomic variables were highly dependent rendering the ordinary regression model less suitable. Haavelmo's solution was to formulate the joint probability, $P(X|X_0;\theta)$, of the sample point X given the initial value X_0 and then estimate the parameters θ based on maximum likelihood rather than OLS.

> ... it has been argued, e.g., that most economic time series do not conform well to any probability model, "because the successive observations are not independent." But it is not necessary that the observations should be independent and that they should all follow the same one-dimensional probability law. It is sufficient to assume that the whole set of, say n, observations [where n means $T \times p$ above, my addition] may be considered as one observation of n variables (or a "sample point") following an n-dimensional joint probability law, the "existence" of which may be purely hypothetical. Then, one can test hypotheses regarding this joint probability law, and draw inference as to its possible form, by means of one sample point (in n dimensions). [Haavelmo, 1944, Preface, *iii*]

The next section will discuss the conditions under which the VAR model can be considered a suitable description of such a sample point.

2.1 Deriving a stationary VAR assuming multivariate normality

We first consider the joint probability of $X|X_0$:

$$P(X|X_0) = P(x_1, x_2, ..., x_T|X_0) \tag{1}$$

and assume that the multivariate normal distribution is a reasonable approximation to the probability law, P. The joint probability of $(X|X_0)$ can be expressed as the probability of the stacked process $Z' = [x'_1, x'_2, x'_3, ..., x'_T] \sim N(M, S)$. Since M is $Tp \times 1$ and S is $Tp \times Tp$, there are far more parameters than observations and, without simplifying assumptions, it is not possible to find unique estimates of M and S. Therefore, we make the assumption that x_t is a stationary process, implying it has a constant mean and constant covariances over time.

Following Hendry and Richard (1983) we decompose the joint probability into conditional probabilities:

$$P(x_1, x_2, x_3, ..., x_T|X_0) = P(x_1|X_0)P(x_2|x_1, X_0) \cdots P(x_T|X^0_{T-1}) \qquad (2)$$

where $X^0_{T-1} = [x_{T-1}, x_{T-2}, ..., x_1, X_0]$. We use the notation $x_t - E(x_t|X^0_{t-1}) = \varepsilon_t$ where ε_t is $NID(0, V(x_t|X^0_{t-1}))$ and $E(x_t|X^0_{t-1})$ and $V(x_t|X^0_{t-1})$ are complicated expression of the mean and the covariances of the marginal processes. Because of the multinormality assumption, the conditional mean is linear, $E(x_t|X^0_{t-1}) = \mu_t + \Pi_{1,t}x_{t-1} + \Pi_{2,t}x_{t-2} + ... + \Pi_{t-1,t}x_1$, but is cumbersome, partly because it contains too many parameters to be estimable, partly because its parameters depend on the length of the conditioning set. To simplify, we assume that only the first k matrices $\Pi_1, \Pi_2, ..., \Pi_k \neq 0$, and the remaining $\Pi_{k+1}, \Pi_{k+2}, ... \simeq 0$. If, in addition $X_0 = [x_0, x_{-1}, ..., x_{-k+1}]$, then we arrive at the k^{th} order vector autoregressive (VAR) model with constant parameters:

$$x_t = \mu_0 + \Pi_1 x_{t-1} + \cdots + \Pi_k x_{t-k} + \varepsilon_t,\ t = 1, ..., T \tag{3}$$

where ε_t is $NID(0, \Omega)$, $x_0, x_{-1}, ...x_{-k+1}$ are assumed fixed, and $\Omega \simeq V(x_t|X^0_{t-1})$. Under these assumptions, ε_t is independent, the OLS estimates of $\{\Pi_1, ..., \Pi_k, \mu_0, \Omega\}$ are Maximum Likelihood estimates, and (3) is essentially a reformulation of the $Tp-$dimensional sample point. It can be considered a first characterization of the unknown data generating process (Hendry and Mizon, 1993).

Two caveats are needed when discussing the usefulness of (3) as a valid characterization of economic data: (i) It is derived for the particular sample window [1,T] and there is no guarantee that other sample periods produce the same linear estimates. This is the question of structural invariance to be discussed in Section 6. (ii) The assumption of stationarity of x_t is seldom empirically tenable for economic time series. But, as discussed in Section 3, nonstationarity of x_t can be accounted for by subjecting the VAR model (3) to nonlinear reduced rank restrictions on the matrices, $\Pi_1, ..., \Pi_k$. Thus, the nonstationary VAR can be considered a submodel of the more general baseline VAR.

2.2 Is the multivariate normality assumption plausible?

Many economists would consider multivariate normality a convenient assumption that one would not expect to find in observed data.[1] Not least Frisch was sceptical about assuming normality:

> Frisch's convictions about the structure of economic reality paired with 'passive observations' left him in no doubt that normality or other reasonable distributions were unlikely to be fulfilled, as required by standard method of statistical analysis. [Bjerkholt, 2011, p. 9]

and

> Frisch was here and in other projects where he hunted for alternative approaches, very skeptical about falling back on the least squares method and perhaps even more about making unwarranted assumptions about normality [Bjerkholt, 2011, p. 12].

But, even though there is no *a priori* reason to expect the VAR residuals to be normally distributed, Haavelmo provided some arguments for such an assumption:

[1] Therefore, the VAR model is often derived under much looser conditions using the Wold representation theorem.

> ... if we consider a set of related economic variables, it is, in general, not possible to express any one of the variables as an exact function of the other variables only. There will be an "unexplained rest," and, for statistical purposes, certain stochastic properties must be ascribed to this rest, a priori. Personally I think that economic theorists have, in general, paid too little attention to such stochastical formulation of economic theories. For the necessity of introducing "error terms" in economic relations is not merely a result of statistical errors of measurement. It is as much a result of the very nature of economic behavior, its dependence upon an enormous number of factors, as compared with those which we can account for, explicitly, in our theories. We need a stochastical formulation to make simplified relations elastic enough for applications. [Haavelmo, 1944, p. 1]

When the residuals are considered a catch-all for everything else that is not included in the empirical model and 'everything else' comprises an 'enormous number of factors' the central limit theorem suggests that normality could be approximately valid, provided these factors are independent. But, as there is no a priori reason to expect independence, normality is an assumption that needs to be checked and when checked it is often rejected. Thus, one could conclude like Frisch, that normality is an assumption which is not necessarily warranted in the data. But instead of giving up on normality, another possibility, which seems more in line with Haavelmo's methodological thinking, is to control for the causes of non-normality, for example, by conditioning on factors in the environment that have changed. Experience shows that it is often extraordinary events during the sample period, such as economic reforms and political interventions, which are the reason for residual skewness and excess kurtosis. In Haavelmo's words:

> Purely empirical investigations have taught us that certain things in the real world happen only very rarely, they are "miracles," while others are "usual events." The probability calculus has developed out of a desire to have a formal logical apparatus for dealing with such phenomena of real life. The question is not whether probabilities exist or not, but whether - if we proceed as if they existed - we are able to make statements about real phenomena that are "correct for practical purposes." [Haavelmo, 1944, p. 43]

The "usual events" can often be adequately described by a normal distribution, whereas the "miracles" tend to fall outside the normal range. Without the normality assumption one would be inclined to ignore these important "miracles" which can be highly informative, for example, about the effect of changes in policy. Failure to properly control for such events is likely to cause residuals to be autocorrelated and inference to be biased. For example, a non-modeled shift in the equilibrium mean and/or average growth rates is likely to cause residual autocorrelation and may (incorrectly) suggest longer lags in the VAR. Such extraordinary events also have strong implications for the model's forecasting performance. See, for example, Clements and Hendry (1999).

Thus, the reason for assuming multivariate normality, is not because we necessarily believe economic data follow the multivariate normality rule, but because it allows us to check that all systematics have been included in the model. It is a safeguard against relying on conclusions from a model which is basically misspecified (Hoover et al., 2009, Hoover, 2012) and ensures that our estimates are full information maximum likelihood.

Today it is straightforward to control for the effect of "miracles" by adequate use of various dummies (see the illustrative example in Section 4.2). Numerous tests for parameter constancy and structural change have been implemented in user-friendly software packages (see for example, Doornik and Hendry, 2006, Hansen and Johansen, 1999 and Dennis et al. 2006) and can be used to detect such "miracles". Experience shows that multivariate normality is seldom supported unless all major changes in the environment have been corrected for. As Section 4.2 shows, this is also crucial for parameter constancy. Frisch's scepticism about the usefulness of the normality assumption might have been a result of problems with parameter nonconstancy without proper tools to cope with them.

3 Correlation, confluence and cointegration

Both Frisch and Haavelmo were concerned about the fact that correlation and regression analysis of trending data tend to produce nonsense results, a fact already demonstrated by Yule (1926). One solution was to use trend-adjusted variables to reduce strong multicollinearity, but this was not sufficient. For instance, Frisch and Waugh (1933) showed (by applying the Gaussian algorithm for solving the normal equations) that the estimated coefficients are identical in regression models with trend-adjusted data or with a linear trend included as a regressor variable. Thus, accounting for a linear trend does not solve the inherent problem of multicollinearity between economic variables. Today we know that the multicollinearity problem is not just due to deterministic but also stochastic trends in the data.

To solve the multicollinearity problem, Frisch developed confluence analysis in the 1930s as a tool for unravelling (identifying) different linear relationships that might hold between a set of (trend-adjusted) variables. The major tool for discovering confluent relations among the variables was called 'bunch maps'. In Haavelmo's lecture notes on confluence analysis, he emphasizes that

> the main purpose of bunch maps was to discover possible multicollinearity in the linear relation to be studied, and that it was necessary to settle this question *before* any attempt to find "best" estimates. ... It is intended to prevent the adoption of a *model of*

estimation which might lead to meaningless results". [Haavelmo and Staehle, 1941, p.28].

The bunch maps is essentially a sophisticated analysis of empirical correlation coefficients. But when data are difference stationary, rather than (trend)-stationary, the empirical correlation coefficient is a misleading estimate of the true association between two variables (Phillips, 1986 and Johansen, 2012). This is because the average value of a unit root variable is an inconsistent estimate of its mean. Since unit root econometrics is a more recent development, this problem could obviously not have been addressed by Frisch and Haavelmo. Nonetheless, hidden unit roots in the variables may very well have affected the empirical performance of bunch maps. Today we know that multicollinearity, whether due to stochastic or deterministic trends, can be easily solved by formulating the VAR model in error-correction form:

$$\begin{aligned} \Delta x_t &= \mu_0 + \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \varepsilon_t,\ t = 1, ..., T \\ \varepsilon_t &\sim NID(0, \Omega). \end{aligned} \tag{4}$$

The hypothesis that $x_t \sim I(1)$ is formulated as a reduced rank condition

$$\Pi = \alpha\beta' \tag{5}$$

where α and β are $p \times r$ matrices $(r < p)$ and the r relations, $\beta' x_t$, define stationary linear relationships between p nonstationary variables. The advantage of the error correction formulation compared to the VAR in levels is that by transforming the trending variables, x_t, into stationary differences, Δx_t, and stationary cointegration relations, $\beta' x_t$, the multicollinearity problem which was of such great concern to Frisch and Haavelmo is more or less solved. This is because:

1. Multicollinearity between the x variables does not lead to imprecise estimates of the cointegration relations, $\beta' x_t$. This is because two variables are cointegrated only if they share a common stochastic trend defined as the cumulation of all permanent shocks that have pushed the variables out of equilibrium. While, for example, cointegration between two unrelated random walks will be rejected with high probability, they may have a correlation coefficient close to one in small samples (see Johansen, 2012). Also, the cointegration coefficients are "canonical" in the sense of being invariant to increasing the information set, or to changing the direction of minimalization.

2. The removal of trends either by differencing or by cointegration is likely to make the multicollinearity between Δx_t and $\beta' x_t$ small enough not to be a problem. When $x_t \sim I(1)$, Δx_t and $\beta' x_t$ are stationary, standard inference on $(\alpha, \Gamma_1, \Sigma)$ applies for given β.

Thus, it is the explicit separation between short-run and long-run effects made possible by cointegration that makes all the difference between the

CVAR type of models and the Haavelmo-Cowles Commission models. As a matter of fact, the interest in confluence analysis (and bunch maps) subsided after the breakthroughs of the Cowles Commission. Hendry and Morgan (1989) gives a number of reasons why this was the case and argues that the more recent advances in cointegration has solved some of the problems associated with identifying structural economic relationships in the data.

4 Structural models, the environment, and the CVAR

Haavelmo discussed the meaning of a structural relation in an unpublished paper prepared for the ESEM-16 meeting in 1954[2]:

> It has little meaning to talk about economic relations that exist without some notion of "environment" in which the relation may be expected to hold good. The totality of properties of the experimental conditions under which a particular economic relation is valid is often called the structure of the economy considered, and the relation itself is called a structural relation. Thus

[2] Herman Wold chaired the program committee with Haavelmo, Tinbergen et al. as members. The main theme of the ESEM meeting had been decided to be "The possibilities and limitations of econometric models; recursive versus structural systems" (Econometrica 22(1), 139). In the tentative program Haavelmo is not listed as presenter of a paper, only for opening comments on the second day for which the theme was given as "General discussion on the possibilities and limitations of econometric models." This may explain the shortness of the paper (6 pp and a 2-page Appendix) and why he never made an effort to publish it. [Personal communication with Olav Bjerkholt.]

> a structural economic relation is not actually a particular kind of economic relation, but rather any economic relation associated with and valid for a specified real economic structure that could conceivably be reproduced experimentally.[Haavelmo, 1954, p. 2]

In the Appendix of the paper he used a simple model describing the demand for a commodity, y_t, as a function of its price, p_t, to illustrate his ideas[3]. The postulated true relationship is first introduced together with a design of experiment based on which the unknown parameter of interest, β_1, could be estimated:

$$y_t = \beta_1 p_t + \beta_0 + u_t \tag{6}$$

where (i) p_t can be deliberately fixed for experimental purposes, (ii) for every fixed value of p_t, u_t is an unobservable random variable with a known distribution which does not depend on the value of p_t, (iii) $u_t's$ are independent in repeated trials, (iv) $E(y_t) = \beta_1 p_t + \beta_0$, (v) β_1 and β_0 are unknown parameters.

He then discussed a situation where instead of an experimental set-up, there are only time-series data available which were not collected in accordance with the designed experiment postulated above:

[3] To facilitate the comparison with the CVAR model, I have changed Haavelmo's (1954) notation as follows: In Haavelmo's equation (A), I have changed α to β_1, in (a)-(b) I have changed $w_{1,t}$ to $\varepsilon_{1,t}$, $w_{2,t}$ to $\varepsilon_{2,t}$, w_t to $\varepsilon_{3,t}$, and β to b. I have also explicitly added a constant β_0 to (6) which is only implicitly assumed in Haavelmo.

$$y_t = \beta_1 p_t + \beta_0 + h\varepsilon_{3,t} + \varepsilon_{1,t} \tag{7}$$

$$p_t = bp_{t-1} + k\varepsilon_{3,t} + \varepsilon_{2,t} \tag{8}$$

where (i) $\varepsilon_{1,t},\ \varepsilon_{2,t}$ and $\varepsilon_{3,t}$ are assumed to be mutually and serially independent random variables, (ii) β_1 is the same unknown parameter as in (6), and (iii) b, h, k are unknown constants. Haavelmo notes that in this case p_t is also a random variable and, unless h or k, or both are zero, $Ey_t \neq (\beta_1 p_t + \beta_0)$, due to the simultaneous effect of $\varepsilon_{3,t}$ on both y_t and p_t.

4.1 Translating Haavelmo's model to a CVAR

The CVAR model is based on the assumption that some of its characteristic roots are unit roots. Even though for obvious reasons nonstationarity was not discussed by Haavelmo, a CVAR translation of Haavelmo's model requires the variables to be nonstationary. By assuming that $b = 1$ in (8), p_t and, hence, also y_t becomes unit root nonstationary:

$$p_t = k\sum_{t=1}^{t} \varepsilon_{3,i} + \sum_{t=1}^{t} \varepsilon_{2,i} + X_0, \qquad t = 1, ..., T \tag{9}$$

where X_0 is a catch-all for initial components. The Haavelmo model contains three stochastic shocks but only two variables. As the CVAR is supposed to account for all sources of stochastic variation, we define $z_t = \sum \varepsilon_{3,t}$ and:

$$\Delta x_t = \alpha\beta' x_{t-1} + \mu_0 + u_t, \qquad u_t \sim NID(0, \Omega) \tag{10}$$

where $x'_t = [y_t, p_t, z_t]$, $u'_t = [u_{1,t}, u_{2,t}, u_{3,t}]$, and $\mu_0 = \alpha\beta_0$, consistent with the assumption that there are no deterministic trends in the variables. To allow for the current effects in the equations (7) and (8), (10) is premultiplied by a matrix A_0:

$$A_0 \Delta x_t = a_1 \tilde{\beta}' \tilde{x}_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim NID(0, \Sigma) \tag{11}$$

where $\tilde{x}'_t = [y_t, p_t, z_t, 1]$, $\tilde{\beta}' = [\beta_1, \beta_2, \beta_3, \beta_0]$, $\varepsilon'_t = [\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t}] = A_0 u_t$, and Σ is a diagonal matrix. The following coefficients of the matrices A_0 and a_1 reproduce Haavelmo's model:

$$\begin{bmatrix} 1 & -\alpha & -h \\ 0 & 1 & -k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta y_t \\ \Delta p_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} [y_{t-1} - \beta_1 p_{t-1} + \beta_0] + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix} \tag{12}$$

$$t = 1, ..., T$$

The system has one cointegration relation corresponding to the true relation (6). It has two exogenous variables, p_t and z_t, of which the former is weakly exogenous because $\alpha_2 = 0$ and the latter is strongly exogenous because $A_{0,31} = A_{0,32} = \alpha_3 = 0$. The first equation in the system corresponds to (7), the second to (8) with $b = 1$, whereas the third is just an auxiliary equation defining $\Delta z_t = \varepsilon_{3,t}$. Thus, the distinction between endogenous and exogenous variables which plays a prominent role in Haavelmo's work is fully compatible with the CVAR formulation. Furthermore, the assumption that p_t is exogenous with respect to y_t is easily testable by the two zero restrictions in the equation for p_t in (12), as is the assumption of lag length and error independence. Thus the CVAR formulation allows us to check the empirical adequacy of the assumptions made and, if needed, make the model more flexible. As discussed in Hoover and Juselius (2012), (12) can be thought of as a "design of experiment for passive observations".

The moving average representation of (12) is given by:

$$\begin{bmatrix} y_t \\ p_t \\ z_t \end{bmatrix} = \begin{bmatrix} \beta_1 & k\beta_1 \\ 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{t} \varepsilon_{2,i} \\ \sum_{i=1}^{t} \varepsilon_{3,i} \end{bmatrix} + C^*(L)\varepsilon_t + \tilde{X}_0, \quad t = 1, ..., T$$

where $C^*(L)\varepsilon_t$ is a lag polynomial of stationary components describing impulse response functions of shocks to the system and $\tilde{X}_0$ contains the initial values, x_0,x_{-1}, of the process and the initial value of the short-run dynamics $C^*(L)\,\varepsilon_0$. The estimate of β_1 is super consistent (Phillips and Durlauf, 1986, and Stock, 1987) and not biased by the appearance of the common shock, $\varepsilon_{3,t}$, in (7) and (8), i.e. $E(\hat{\beta}_1) = \beta_1$. The two common stochastic trends, $\sum_{i=1}^{t} \varepsilon_{2,i}$ and $\sum_{i=1}^{t} \varepsilon_{3,i}$, are cancelled in the relation $y_t - \beta_1 p_t$ and $E(y_t - \beta_1 p_t + \beta_0) = 0$. While y_t and p_t are likely to be strongly multicollinear (measured by the empirical correlation coefficient), this is not a problem in (12) where all components in Δx_t and $\beta' x_t$ are nontrending.

4.2 Changes in the environment

The question whether a cointegration relation is likely to remain constant when there are changes in the environment is crucial for its structural interpretability. In Haavelmo, Chapter II, about the "degree of permanence of economic laws" Haavelmo raised the question "whether or not we might hope to find elements of invariance in economic life, upon which to establish permanent 'laws'":

> When we use the terms "constant relationships", or "unstable, changing relationships", we obviously refer to the behavior of some real economic phenomena, as compared with some behavior that we expect from theoretical considerations. The notion of constancy or permanence of a relationship is, therefore, not one of pure theory. It is a property of real phenomena as we look upon them from the point of view of a particular theory. [Haavelmo, 1944, p. 13]

As an illustration, let us assume that the environment changes at time t_1 in Haavelmo's model, for example as a result of a political reform that leads to an unanticipated shift in the level of y_t, but not in p_t. As a consequence there is an extraordinary change in Δy_t and a shift in the equilibrium mean of the relation between y_t and p_t. To account for these changes the CVAR is respecified as:

$$\Delta x_t = \alpha\beta' x_{t-1} + \mu_0 + \Phi_1 D_{s,t_1} + \Phi_2 D_{p,t_1} + u_t, \quad u_t \sim NID(0, \Omega), \quad t = 1, ..., T$$

where $D_{s,t_1} = 1$ for $t = t_1, ..., T$, 0 otherwise, $D_{p,t_1} = 1$ for $t = t_1$, 0 otherwise, and $\Phi_1 = \alpha\beta_{01}$ implies that the mean $E(\beta' x_t) = \beta_0 + \beta_{01}$ after t_1. The CVAR with current effects becomes:

$$A_0 \Delta x_t = a_1 \tilde{\beta}' \tilde{x}_{t-1} + \Phi_2 D_{p,t_1} + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \Sigma)$$

where $\tilde{x}'_{t-1} = [y_t, p_t, z_t, 1, D_{s,t_1}]$, $\tilde{\beta}' = [\beta_1, \beta_2, \beta_3, \beta_0, \beta_{01}]$. Haavelmo's model can now be represented by:

$$\begin{bmatrix} 1 & -\alpha & -h \\ 0 & 1 & -k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta y_t \\ \Delta p_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} [y_{t-1} - \beta_1 p_{t-1} - \beta_0 - \beta_{01} D_{s,t_1}] + \Phi_2 D_{p,t_1} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix},$$

and the corresponding moving average representation is

$$\begin{bmatrix} y_t \\ p_t \\ z_t \end{bmatrix} = \begin{bmatrix} \beta_1 & k\beta_1 \\ 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{t} \varepsilon_{2,i} \\ \sum_{i=1}^{t} \varepsilon_{3,i} \end{bmatrix} + \begin{bmatrix} \beta_0 & \beta_{01} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ D_{s,t} \end{bmatrix} +$$

$$C^*(L)(\varepsilon_t + \Phi_2 D_{p,t_1}) + X_0, \quad t = 1, ..., T$$

By accounting for the unanticipated effect of the reform with the impulse dummy and the shift in the equilibrium mean with a step dummy, the estimate of the parameter of interest, β_1, is unaffected by the change in environment, $E(\hat{\beta}_1) = \beta_1$ and $E(\tilde{\beta}'\tilde{x}_{t-1}) = 0$. Also, the residuals continue to obey the multivariate normality rule, illustrating the discussion in Section 2.2. That such changes in the environment played an important role in Haavelmo's methodological thinking is evident from:

> Very often our theories are such that we think certain directly observable series would give adequate experimental results for a verification, provided other things did not change. What bearing may such theories have upon reality, if we simply neglect the influences of these "other things"? This, again, is connected with the following problem: Are we interested in describing what actually does happen, or are we interested in what would happen if we could keep "other things" unchanged? In the first case we construct theories for which we hope Nature itself will take care

> of the necessary ceteris paribus conditions, knowing, e.g., that this has been approximately so in the past. In the second case we try to take care of the *ceteris paribus* conditions ourselves, by statistical devices of clearing the data from influences not taken account of in the theory (e.g., by multiple-correlation analysis). [Haavelmo, 1944, p. 16-17]

5 Simultaneous equations and identification

The problem of identifying simultaneous economic relationships, for example a demand relation from a supply relation, was first addressed by R.A. Lehfeldt in 1914, followed by a number of other scholars. See Hendry and Morgan (1995) for a survey and discussion. Haavelmo's important contribution was to address the statistical implications of such simultaneous equations when the variables and the relations are stochastically determined.

> if we consider a set of related economic variables, it is, in general, not possible to express any one of the variables as an exact function of the other variables only. There will be an"unexplained rest," and, for statistical purposes, certain stochastical properties must be ascribed to this rest ... We need a stochastical formulation to make simplified relations elastic enough for applications. [Haavelmo, 1943, p.1]

A related issue was how to estimate the parameters of interest in a simultaneous equation system. Haavelmo pointed out that regressing one endogenous variable on exogenous and other endogenous variables would in general give rise to a simultaneity bias in the parameters of interest:

> if one assumes that the economic variables considered satisfy, simultaneously, several stochastic relations, it is usually not a satisfactory method to try to determine each of the equations separately from the data, without regard to the restrictions which the other equations might impose upon the same variables. That this is so is almost self-evident, for in order to prescribe a meaningful method of fitting an equation to the data, it is necessary to define the stochastical properties of all the variables involved. [Haavelmo, 1943, p.2]

To be able to estimate the parameters of a simultaneous equations system, the problem of identification has first to be solved. Johansen (1994) and Johansen and Juselius (1994) discuss three different concepts: (1) generic identification which is related to the specification of a simultaneous model and is necessary for model parameters to be uniquely determined, (2) economic identification which requires that the model has identified the economic parameters of interest and, finally, (3) empirical identification which requires that generic identification is not lost by setting a statistically insignificant coefficient to zero. The discussion below will center around the first two concepts and how they can be understood when data are nonstationary.

5.1 Identification when data are nonstationary

A typical Haavelmo - Cowles Commission simultaneous equations model was defined by endogenous variables being a function of other endogenous variables, exogenous variables and lagged variables and identified by imposing a sufficient number of (exclusion) restrictions on the parameters. Contrary to the traditional Haavelmo model, the CVAR does not distinguish between endogenous and exogenous variables: all stochastic variables are modelled and exogeneity of a variable is tested as a zero row restriction on the α matrix and not assumed from the outset. The separation between the r pulling and the $p-r$ pushing forces, implies that the CVAR is inherently consistent with $p-r$ exogenous trends, $\alpha'_{\perp}\sum_{i=1}^{t}\varepsilon_i$, where $\alpha'_{\perp}$ is a $p-r\times p$ matrix orthogonal to α. Unless the $p-r$ vectors in $\alpha_{\perp}$ are unit vectors, the exogenous trends do not correspond directly to any of the p variables. Instead the exogenous trends are cumulations of latent "structural shocks" to the system, such as a demand and a supply shock, estimated as a linear combination of the CVAR residuals.

The dichotomy of pulling and pushing forces allows us to address identification in four dimensions: the identification of (1) the long-run cointegration structure, (2) the short-run adjustment structure, (3) the exogenous driving shocks, and (4) the dynamics of the impulse responses. See Juselius (2006) for a detailed discussion. Only the first two will be discussed here as they are most closely related to the identification problem of a traditional system of simultaneous equations.

To illustrate the relationship between long-run and short-run identification, the 'reduced form' (4) subject to (5) is pre-multiplied by the current effects matrix A_0:

$$A_0 \Delta x_t = A_1 \Delta x_{t-1} + a_1 \beta' x_{t-1} + v_t, \quad v_t \sim NID(0, \Sigma). \tag{13}$$

where $A_1 = A_0 \Gamma_1$, $a_1 = A_0 \alpha$, $\mu_{0,a} = A_0 \mu_0$, $v_t = A_0 \varepsilon_t$, $\Sigma = A_0 \Omega A_0'$. It appears that β is the same in the "reduced form" and the "contemporaneous form" and can, therefore, be estimated based on either form. The fact that the estimate of β is super consistent, while the estimate of the short-run adjustment parameters are $\sqrt{T}$ consistent, allows the identification to be performed in two steps: (1) the identification of the long-run parameters, β, and (2) the identification of the short-run structure conditional on the identified β (Johansen, 1995).

Generic identification of the r (simultaneous) long-run relations requires at least $r(r-1)$ restrictions, and the short-run adjustment equations at least $p(p-1)$ restrictions. In both cases the restrictions have to satisfy the identification rank conditions derived for the CVAR model by Johansen (1995) and Johansen and Juselius (1994).[4] Thus, the reduced rank restrictions $\Pi = \alpha\beta'$ are helpful for identification as they allow a separation of long-run from

[4] Similar rank conditions was already established for the traditional simultaneous equation system by Koopmans, Rubin and Leipnik (1950) and Wald (1950).

short-run effects.

The identification of the long-run structure, $\beta' x_t$, is similar to the traditional simultaneous equations, describing relationships among endogenous, exogenous, and lagged variables, except that no lags are involved in $\beta' x_t$. Identification of the short-run adjustment structure is basically about how to identify causal links in the data by imposing restrictions on the contemporaneous matrix A_0, given lagged changes of the process, Δx_{t-1}, and lagged equilibrium errors, $\hat{\beta}' x_{t-1}$, where $\hat{\beta}$ is an estimated long-run structure. Economic identification of the short-run structure generally requires the residuals to be uncorrelated. Large off-diagonal elements of the covariance matrix Ω arise when the current changes of the system variables are strongly correlated, or when the residuals are simultaneously being affected by omitted variables. As an illustration of the former case, the residual covariance matrix is a nondiagonal matrix in the reduced form of Haavelmo's model (10), whereas it is a diagonal matrix in the structural form (12).

5.2 Irreducible cointegration and coflux relations

An identified cointegration structure consists of r irreducible cointegration relations, where irreducibility implies that stationarity is lost if one of the variables is omitted from the relation (Davidson, 1998). Hence, they contain exactly the right number of variables needed to make the relation stationary, no less, no more. There is, however, no reason to expect the number of irreducible relations to be same as the number of postulated economic relations.

The consequence is that a cointegration relation does not necessarily correspond to a hypothetical economic relation. To illustrate this we assume that the relation of economic interest is a Phillips curve with a Phelpsian natural rate (Phelps 1994, Juselius and Juselius, 2013):

$$\Delta p_t = -b_1 u_t + u_t^* \tag{14}$$

where Δp_t is the inflation rate, u_t is the unemployment rate, u_t^* is the natural rate as a function of the real interest rate, r_t:

$$u_t^* = b_2 r_t + b_0. \tag{15}$$

The sign of the coefficients b_i define *a priori* expected effects, the three variables are collected by passive observation,and all of them are stochastic. A CVAR model for $x_t' = [\Delta p_{t,} u_t, r_t]$ is given below where for simplicity we disregard the transitory effects Γ_1 and focus exclusively on the long-run component $\alpha\beta'$.

For rank one, implying two stochastic exogenous trends, we get:

$$\begin{bmatrix} \Delta p_t \\ \Delta u_t \\ \Delta r_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} [\beta_1' x_t] + \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}.$$

where $\beta_1' x_t = \beta_1 \Delta p_t + \beta_2 u_t + \beta_3 r_t$.With just one cointegration relation, normalizing on Δp_t is sufficient for generic identification[5]. If, in addition, $\beta_2 = b_1$ and $\beta_3 = -b_2$ then there is a one to one correspondence between the cointegration relation and the Phillips curve (14) with a Phelpsian natural rate (15), that is we have achieved economic identification. In this case inflation and unemployment are not cointegrated, nor is the real interest rate stationary. Furthermore, if $\alpha_1 < 0$ then Δp_t is equilibrium correcting, and if $\alpha_2 = \alpha_3 = 0$, then u_t and r_t are exogenous.

For rank two implying one stochastic exogenous trend, we get:

$$\begin{bmatrix} \Delta p_t \\ \Delta u_t \\ \Delta r_t \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} \beta_1' x_t \\ \beta_2' x_t \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}.$$

With two cointegration relations, generic identification requires at least one restriction on each relation in addition to normalization. Let the identified relations be $\beta_1' x_t = \beta_{11} \Delta p_t + \beta_{12} u_t$ and $\beta_2' x_t = \beta_{23} r_t$ where $\beta_{11} = 1$, $\beta_{12} = b_1$ and $\beta_{23} = 1$, implying one overidentifying restriction on each relation. In this case, inflation and unemployment are cointegrated and the real interest rate is stationary. The economic relation (14) can be recovered by combining the two cointegration relations:

$$\alpha_{11}\beta_1' x_t + \alpha_{12}\beta_2' x_t = \alpha_{11} \left\{ \Delta p_t + b_1 u_t + \alpha_{12}/\alpha_{11} r_t \right\}. \tag{16}$$

[5]While normalization can be on any of the variables x_i without changing the relative values β_i/β_j, an economically meaningful normalization requires that the corresponding α_i is significant.

To label (16) a Phillips curve relation usually requires that inflation is significantly equilibrium correcting to the relation and the condition for economic identification is that $\alpha_{11} < 0$ and $\alpha_{12}/\alpha_{11} = -b_2$.

The example illustrates that economic identification is generally incomplete without combining irreducible cointegration relations with the short-run adjustment coefficients. This is different from a traditional simultaneous equation model associating a number of endogenous variables with a number of exogenous variables and lagged endogenous and exogeous variables. Identification is then mostly achieved by exclusion restrictions and causality is implicitly assumed by normalizing on a postulated endogenous variable in each equation. One may say that the adjustment coefficients α in the CVAR play a similar role for identification as the lagged variables in the simultaneous equations model.

The above example points to strong parallels between cointegration and Frisch confluence analysis (Frisch, 1934). Both represents statistical methods developed as a means to uncover structure among correlated variables. In particular, the correspondence between an irreducible cointegration relation and what Frisch called a coflux equation is rather striking:

> An equation which is irreducible with respect to the set of functions which forms the actual solution of the complete system we shall call a "coflux" equation. ... The notion of coflux relations is fundamental when we ask what sorts of equations it is possible to determine from the knowledge of the time shapes that are actu-

> ally produced. The answer is obviously that all coflux equations and no other equations are discoverable from the time shapes of the functions that form the actual solutions. [Frisch, 1938, p.14]

Coflux equations were important because they could be estimated (in the deterministic case, solved for) from the data - "they were discoverable through passive observations." In modern terms, coflux equations are the ones that are identified. But, while coflux equations are estimable, they are not necessarily the most interesting equations (Aldrich,1989, p. 24).

The same can also be said about cointegration relations. Because any linear combination of r cointegration relations is also a stationary relation there are usually many ways of identifying a structure of irreducible relations. For example, if $x_{1,t} - x_{2,t}$ and $x_{2,t} - x_{3,t}$ are stationary, then $x_{1,t} - x_{3,t}$ is also stationary and the long-run structure $(\beta_1' x_t, \beta_2' x_t)$ can be identified by either $(x_{1,t} - x_{2,t}, x_{2,t} - x_{3,t})$ or $(x_{1,t} - x_{2,t}, x_{1,t} - x_{3,t})$ and one of the sets may not be economically interesting.

To summarize, a generically identified structure of r irreducible cointegration relations, $\beta' x_t$, can be thought of as building blocks that can be used to construct meaningful economic relations with the help of the α coefficients.

6 Structural invariance, autonomy, and cointegration

> In trying to establish relations with high degree of autonomy we take into consideration various *changes* in the economic structure which might upset our relations, we try to dig down to such relationships as actually might be expected to have a great deal of invariance with respect to certain changes in structure that are "reasonable" [Haavelmo, 1944, p.28].

Can we use cointegration to dig down to such invariant relationships? Cointegration possess certain invariance properties that can be exploited when searching for structure. For example, the cointegration property is invariant to extensions of the information set. If cointegration is found between a set of variables in small CVAR model, the same cointegration relation will be found in a CVAR model with a larger set of variables. Adding new variables to the CVAR model is, however, likely to increase the cointegration rank and, hence, new cointegration relations would have to be identified. The invariance property of a cointegration relation does not, however, extend to the short-run adjustment coefficients. For example, a variable found to be exogenous in a smaller model may no longer be so in a larger model. Also allowing for simultaneous effects among the endogenous variables is likely to change α and Γ_1 in (4). While this suggests that economic identification should be based on a fairly complete CVAR model, experience shows that identification of the long-run structure tends to become increasingly difficult as the number of variables increases. Fortunately, the invariance property of a cointegration relation, allows us to gradually expand the CVAR by building on previously found cointegration relations when adding variables the model. Such a procedure allows us to systematically exploit the effect of the *ceteris paribus* assumption in the theory model on the empirical conclusions. For an illustration, see Juselius (2006).

But even though cointegration analysis is a powerful method for uncovering genuine relationships among variables, it is basically a statistical regularity that may break down if conditions change. Therefore, cointegration is no guarantee for structural invariance in the sense that its coefficients might change when other parts of the structure change.[6] Consider for example, the rank two case of the Phillips curve of the previous section where the inflation-unemployment and the real interest relations were both stationary. Assume now that economic conditions change in a way that introduces a new stochastic trend causing the cointegration relations to become nonstationary. An economist estimating a standard Phillips curve between inflation and unemployment without accounting for the effect of the real interest rate would now conclude that the Phillips curve has broken down. But the basic feature of the Phillips curve, $\beta_{12} = b_1$, could still be unchanged in the combined relation (16). In Haavelmo's words: "The construction of systems of autonomous relations is, therefore, a matter of intuition and factual knowledge; it is an

[6] This is closely related to the concept of super exogeneity in Engle et al. (1983).

art" (Haavelmo, 1944, p.29).

Another example is the Danish money demand relation which was shown to have been constant from 1975 to 2003 except for a shift in the equilibrium mean in connection with a major financial deregulation in 1983 (Juselius 2006). The question is whether the relation would be stable under other hypothetical reforms that change the economic structure. Would Haavelmo/Frisch have considered it a structural (autonomous) relation despite the equilibrium mean shift? Haavelmo seemed to have had a fairly pragmatic view on this issue:

> It has little meaning to talk about economic relations that exist without some notion of "environment" in which the relation may be expected to hold good. The totality of properties of the experimental conditions under which a particular economic relation is valid, is often called the *structure* of the economy considered, and the relation itself is called a *structural relation* [Haavelmo, 1954, p.2].

Still it is easy to share Frisch conclusion that autonomy and structural invariance are theoretical concepts which are empirically elusive.

> The question of what connection there is between the relations we work with in theory and those we get by fitting curves to actual statistical data is a very delicate one. I think it has never been exhaustively and satisfactorily discussed. [Frisch, 1938]

7 A concluding discussion

The idea of this paper was to demonstrate that many econometric problems which were discussed by Haavelmo and his contemporaries have been given a practical solution within the general framework of a *well* specified CVAR model. It was chosen because it can be derived from a joint probability model for all observables, and therefore can represent Haavelmo's probability approach to economics. By exploiting the unit root feature, typical of many economic variables, the CVAR model was shown to solve the problem of (1) time dependent residuals by conditioning on sufficiently many lags and controlling for a changing environment when needed, (2) spurious correlation and regression results, (3) multicollinearity, (4) normalization, and (5) reduced rank.

The paper also argues that the unit root property of economic data allows us to address identification, simultaneity, and structural invariance in a much richer context than was possible for Frisch and Haavelmo and their contemporaries. Whether the CVAR model can be assumed to produce autonomous or structurally invariant results depends on whether the economist is able to "unravel the relationships of interest from the ones which were a characteristic of the data set, but of no interest to the economist (Hendry and Morgan, 1989)". It also depends on whether the sample period represents a reasonably constant environment or, when this is not the case, whether such changes can be controlled for. The frequent finding that the normality assumption and parameter constancy are acceptable only *ex post* after we have allowed for shifts in the equilibrium mean (or shifts in the growth rates) due to extraordinary institutional events implies that the CVAR is not likely to produce unbiased forecast errors over periods potentially subject to structural changes and location shifts in the probability distribution (Castle et al. 2010). This caveat is of course not specifically relevant for the CVAR, but applies generally to empirical models. It is likely to have important implications for the choice of theoretical models that can be claimed to be empirically relevant.

As a final conclusion, likelihood based cointegration seems to be able to combine the basic ideas of Frisch's confluence analysis with Haavelmo's probability approach into a rich methodological approach for making inference based on passive observations.

8 References

Aldrich, J. (1989) Autonomy. *Oxford Economic Papers*, New Series, Vol. 41, No. 1, History and Methodology of Econometrics, pp. 15-34.

Anderson, T.W. (1951) Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. *Annals of Mathematical Statistics,* 22, 327-351.

Andersen, T.W. (1991) Trygve Haavelmo and Simultaneous Equation Models. *Scandinavian Journal of Statistics*, 18, 1-19.

Bjerkholt, O. (2007) Writing "the Probability Approach" with Nowhere to go: Haavelmo in the United States, 1939-1944. *Econometric Theory*, 23, pp. 775-837.

Bjerkholt, O. (2011) Trygve Haavelmo, Frisch and the Probability Approach. *The International Journal of Applied Economics and Econometrics*, 21, 3, 96-117.

Castle, J.L., N.W.P. Fawcett, D.F. Hendry (2010) Forecasting with Equilibrium-correction Models during Structural Breaks. *Journal of Econometrics*, 158, 25-38.

Clements, M. P. and D. F. Hendry (1999) *Forecasting Non-stationary Economic Time Series*. MITPress: Cambridge, Mass.

Davidson, J. (1988) Structural Relations, Cointegration and Identification: Some Simple Results and their Application. *Journal of Econometrics*, 87(1), 87-113.

Dennis, J. G., H. Hansen, S. Johansen, and K. Juselius (2006) CATS in RATS. Cointegration Analysis of Time Series. Version 2. Estima: Evanston, Illinois, USA.

Doornik, J. A. and D. F. Hendry (2006). Empirical Econometric Modelling Using PcGive. Volumes I, II and III. Timberlake Consultants Press: London.

Engle, R.F., D.F. Hendry, and J-F Richard (1983), "Exogeneity", *Econometrica*, 51, 2, 277-304.

Frisch, R. (1934) Statistical confluence analysis by means of complete regression systems. University Institute of Economics, Oslo, pp. 5-8.

Frisch, R. (1938) Statistical Versus Theoretical Relations in Economic Macrodynamics. Unpublished memorandum, University of Oslo.

Frisch, R. and F. Waugh (1933) Partial Time Regressions as Compared with individual Trends. *Econometrica*, 1, 4, pp. 387-401

Haavelmo, T. (1943) The Statistical Implications of a System of Simultaneous Equations. *Econometrica*, 11, 1-12.

Haavelmo, T. (1944) The Probability Approach to Econometrics. *Econometrica*, 12 (Supplement), 1-118.

Haavelmo, T. (1950) Remarks on Frisch's Confluence Analysis and its Use in Econometrics. In *Statistical Inference in Dynamic Models* (ed.) T. C. Koopmans, Cowles Commission Monograph, Vol. 10, John Wiley & Sons, New York, 258-265.

Haavelmo, T. (1954) Structural Models and Econometrics. Unpublished paper presented at the Econometric Society Meeting 1954 in Stockholm. http://www.sv.uio.no/econ/english/research/networks/haavelmo-network/ publications/files/th19
es-1954-uppsala.pdf

Haavelmo, T. & H. Staehle (1941) The Elements of Frisch's Confluence Analaysis. Hecto., Harvard University, Cambridge, Massachusetts.

Hansen, H. and S. Johansen (1999) Some Tests for Parameter Constancy in Cointegrated VAR Models. *The Econometrics Journal*, 2(2), 306–333.

Hendry, D.F. and G.E. Mizon (1993) Evaluating Econometric Models by Encompassing the VAR. In Phillips, P.C. (ed.) *Models, Methods and*

Applications of Econometrics, Blackwell, Oxford, 272-300.

Hendry, D.F. and M.S. Morgan (1989) A Re-Analysis of Confluence Analysis. *Oxford Economic Papers*, Vol. 41, No. 1, History and Methodology of Econometrics, pp. 35-52

Hendry, D.F. and M.S. Morgan (1995) *The Foundations of Econometric Analysis*. Cambridge University Press, Cambridge.

Hendry, D.F. and J.F. Richard (1983) The Econometric Analysis of Economic Time-Series (with discussion). *International Statistical Review*, 51, 111-163.

Hendry, D.F., A. Spanos, and N.R. Ericsson (1989) The Contributions to Econometrics in Trygve Haavelmo's 'The Probability Approach in Econometrics'. *Sosialøkonomen*, 43(11), 12-17.

Hoover, K. D. (2012) The Role of Hypothesis Testing in the Molding of Econometric Models. CHOPE Working Paper No. 2012-03; Available at SSRN: http://ssrn.com/abstract=2001481.

Hoover, K. D., S. Johansen, and K. Juselius (2009) Allowing the Data to Speak Freely: The Macroeconometrics of the Cointegrated Vector Autoregression. *American Economic Review* 98, pp. 251-55.

Hoover, K. D. and Juselius, K. (2012) Experiments, Passive Observation and Scenario Analysis: Trygve Haavelmo and the Cointegrated Vector Autoregression. Univ. of Copenhagen Dept. of Economics Discussion Paper No. 12-16. Available at SSRN: http://ssrn.com/abstract=2171700

Johansen, S. (1988) Statistical Analysis of Cointegration Vectors. *Journal* *of Economic Dynamics and Control*, 12(213), 231–254.

Johansen, S. (1995) Identifying Restrictions of Linear Equations. With Applications to Simultaneous Equations and Cointegration. *Journal of Econometrics*, 69(1), 111-132.

Johansen, S. (1996) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford*. Oxford University Press.

Johansen, S. (2012) The Analysis of Nonstationary Time Series Using Regression, Correlation and Cointegration. *Contemporary Economics*, Vol. 6, pp. 40-57.

Johansen, S. and K. Juselius (1994) Identification of the Long-Run and Short-Run Structure: An Application to the ISLM Model. *Journal of Econometrics*, 63, 7-36.

Juselius, K. (2006), *The Cointegrated VAR Model: Methodology and Applications*. Oxford: Oxford University Press.

Juselius, K. and M. Juselius (2013) Balance sheet recessions and time-varying coefficients in a Phillips curve relationship: An application to Finnish data. In eds.) N. Haldrup, M. Meitz, and P. Saikkonen, *Essays in Nonlinear Time Series Econometrics: Festschrift in Honour of Timo Teräsvirta*. Oxford University Press. Forthcoming

Koopmans, T.C., H. Rubin, and R.B. Leipnik (1950) Measuring the Equation Systems of Dynamic Economics. In T.C. Koopmans (ed.) *Statistical inference in dynamic economic models,* Cowles Commission Research, John Wiley & Sons, Inc. New York.

Lehfeldt, R.A. (1914) The Elasticity of the Demand for Wheat. *Economic Journal*, Vol. 24, pp. 212-17.

Morgan, M.S. (1990) *The History of Econometric Ideas. Cambridge University Press.*

Nielsen, H. B. (2008) Influential Observations in Cointegrated VAR Models: Danish Money Demand 1973-2003. *The Econometrics Journal*, 11 (1): 1-19.

Phelps, E. (1994) *Structural Slumps*, Princeton University Press, Princeton.

Phillips, P.C.B., (1986) Understanding Spurious Regressions in Econometrics. *Journal of Econometrics* 33, 311-340.

Phillips, P.C.B., (1987) Time Series Regression with a Unit Root. *Econometrica*, 55, 277-3011.

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

Romer, D. (1996) *Advanced Macroeconomics.* McGraw Hill: New York.

Stock, J.H. (1987) Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors. *Econometrica*, 55, 1035-1056.

Yule, U. (1926) Why do we Sometimes Get Nonsense-Correlations Between Time Series? – A Study in Sampling and the Nature of Time Series. *Journal of the Royal Statistical Society* 89, 1–63.

Wald, A. (1950) A Note on the Identification of Economic Relations. in T.C. Koopmans (ed.) *Statistical inference in dynamic economic models,*

Cowles Commission Research