

Editorial

# A Conversation with Søren Johansen

Rocco Mosconi <sup>1</sup>  and Paolo Paruolo <sup>2,\*</sup> 

<sup>1</sup> Dipartimento di Ingegneria Gestionale, Politecnico di Milano, Via Lambruschini 4/B, 20156 Milano, Italy; rocco.mosconi@polimi.it

<sup>2</sup> European Commission, Joint Research Centre (JRC), Via E. Fermi 2749, 21027 Ispra, Italy

\* Correspondence: paolo.paruolo@ec.europa.eu

**Abstract:** This article was prepared for the Special Issue “Celebrated Econometricians: Katarina Juselius and Søren Johansen” of *Econometrics*. It is based on material recorded on 30 October 2018 in Copenhagen. It explores Søren Johansen’s research, and discusses inter alia the following issues: estimation and inference for nonstationary time series of the I(1), I(2) and fractional cointegration types; survival analysis; statistical modelling; likelihood; econometric methodology; the teaching and practice of Statistics and Econometrics.

**Keywords:** cointegration; fractional (co-)integration; statistical model; survival analysis; VAR; I(1); I(2)

**JEL Classification:** C32; B41; C01; C10; C30; C52

## Introduction

On 30 October 2018 the authors sat down with Søren Johansen in Copenhagen to discuss his wide-ranging contributions to science, with a focus on Econometrics. Figure 1 reports a photo of Søren taken on the day of the conversation; other recent photos are reported in Figure 2. The list of his publications can be found at the following link: <http://web.math.ku.dk/~sjo/>.<sup>1</sup>

In the following, frequent reference is made to vector autoregressive (VAR) equations of order  $k$  for a  $p \times 1$  vector process,  $X_t$ , for  $t = 1, \dots, T$ , of the following form:

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t, \quad (1)$$

where  $\Pi$  and  $\Gamma_i$  are  $p \times p$  matrices, and  $\Delta = 1 - L$  and  $L$  are the difference and the lag operators, respectively.

Various models of interest in cointegration are special cases of (1), in particular the cointegrated VAR (CVAR), defined by restricting  $\Pi$  in (1) to have reduced rank, i.e.,  $\Pi = \alpha\beta'$  with  $\alpha$  and  $\beta$  of dimension  $p \times r$ ,  $r < p$ . Another matrix of interest is the  $p \times p$  matrix  $\Gamma = I - \sum_{i=1}^{k-1} \Gamma_i$ , see Johansen (1996, chp. 4) for further reference. For any matrix  $\alpha$ ,  $\alpha_{\perp}$  indicates a basis of the orthogonal complement to the span of  $\alpha$ ; this orthogonal complement is the set of all vectors orthogonal to any linear combinations of the column vectors in  $\alpha$ .

In the rest of the article, questions are in bold and answers are in Roman. Text additions are reported between [ ] or in footnotes. Whenever a working paper was later published, only the published paper is referenced. The sequence of topics covered in the conversation is as follows: cointegration and identification; survival analysis and convexity; model specification.



**Citation:** Mosconi, Rocco, and Paolo Paruolo. 2022. A Conversation with Søren Johansen. *Econometrics* 10: 21. <https://doi.org/10.3390/econometrics10020021>

Received: 4 April 2022

Accepted: 6 April 2022

Published: 13 April 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



**Figure 1.** Søren Johansen, 30 October 2018 in Copenhagen.

### **What is your current research about?**

I worked on several projects. With Bent Nielsen [referred to as Bent hereafter] I have studied some algorithms and estimators in robust statistics including M-estimators, see [Johansen and Nielsen \(2019\)](#), and with Morten Ørregaard Nielsen [referred to as Morten hereafter] I have worked on fractional cointegration and other topics in cointegration, see for instance the paper on a general formulation for deterministic terms in a cointegrated VAR model [Johansen and Nielsen \(2018\)](#).

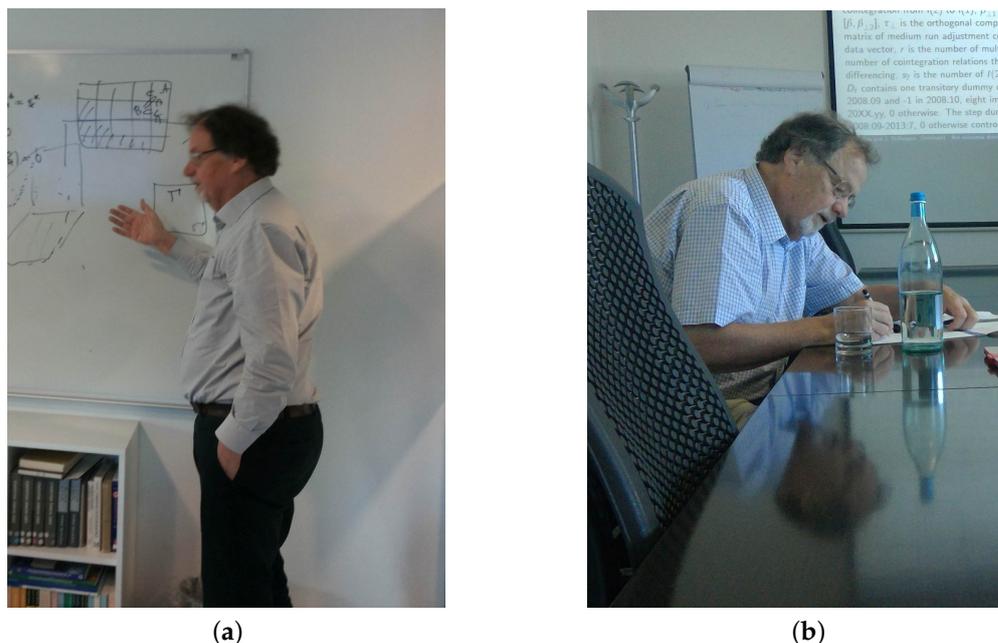
I have collaborated with Kevin Hoover on the analysis of some causal graphs, and just written a paper for this Special Issue ([Johansen 2019](#)) on the problem that for a CVAR the marginal distribution of some of the variables is in general an infinite order CVAR, and one would like to know what the  $\alpha$  coefficients in the marginal model are.

I have also recently worked with Eric Hillebrand and Torben Schmith ([Hillebrand et al. 2020](#)) on a cointegration analysis of the time series of temperature and sea level, for the Special Issue for David Hendry in the same journal. We compare the estimates for a number of different models, when the sample is extended. There has been a growing interest in using cointegration analysis in the analysis of climate data, but the models have to be built carefully taking into account the physical models in this area of science.

### **The notion of cointegrating space was implicit in Engle and Granger’s 1987 paper. You mentioned it explicitly in a paper of yours in 1988.<sup>2</sup> Could you elaborate on this?**

When you realize that linear combinations of cointegrating vectors are again cointegrating, it is natural to formulate this by saying that the cointegrating vectors form a vector space. That of course implies that you have to call the zero vector “cointegrating”, even if there are no variables involved. Moreover a unit vector is also cointegrating, even though only one variable is involved. I sometimes try to avoid the word “cointegration”, which obviously has connotations to more than just one variable, and just talk about stationary linear combinations.

This lack of acceptance, that a cointegrating vector can be a unit vector, is probably what leads to the basic misunderstanding that almost every applied paper with cointegration starts with testing for unit roots with univariate Dickey-Fuller tests, probably with the consequence that stationary variables will not be included in the rest of the analysis. It is, I think, quite clear that analysing the stationarity of individual vectors in a multivariate framework by testing for a unit vector in the cointegrating space is more efficient than trying to exclude variables from the outset for irrelevant reasons.



**Figure 2.** Søren Johansen: (a) 24 February 2016 in Copenhagen; (b) 3 October 2016 in Milan.

Going back to the cointegrating space, it is a natural concept in the following sense. The individual cointegrating relations are not identified, and one has to use restrictions from economic theory to identify them. But the cointegrating space itself is identified, thus it is the natural object to estimate from the data in the first analysis.

Hence the cointegrating space is a formulation of what you can estimate without having any special knowledge (i.e., identifying restrictions) about the individual cointegrating relations. The span of  $\beta$  (which is the cointegrating space) is therefore a useful notion.

#### **Estimation and testing for cointegration are sometimes addressed in the framework of a single equation.**

When estimating a cointegrating relation using regression, you get consistent estimates, but not valid  $t$ -statistics. Robert Engle [referred to as Rob hereafter] worked out a three-step Engle-Granger regression which was efficient, see [Engle and Yoo \(1991\)](#). Later Peter [Phillips \(1995\)](#) introduced the fully modified regression estimator, where the long-run variance is first estimated and then used to correct the variables, followed by a regression of the modified variables. If there are more cointegrating relations in the system, and you only estimate one, you will pick up the one with the smallest residual variance. It is, however, a single equation analysis and not a system analysis, as I think one should try to do.

#### **How were your discussions on cointegration with the group in San Diego?**

My contact with the econometric group in San Diego started when I met Katarina Juselius [referred to as Katarina hereafter]. She had met David Hendry [referred to as David hereafter] while on sabbatical at London School of Economics in 1979. She was one of the first to use PcGive.<sup>3</sup> Rob was visiting David in those days. That meant that in 1985, when we went for a month to San Diego, we met Clive Granger [referred to as Clive hereafter], Helmut Lütkepohl and Timo Teräsvirta. So when I started to work on cointegration we knew all the right people.

We were well received and discussed all the time. Clive was not so interested in the technicalities I was working on, but was happy to see that his ideas were used. Rob, however, was more interested in the details. When we met a few years later at the 1987 European Meeting of the Econometric Society in Copenhagen, he spent most of his lecture talking about my results, which is the best welcome one can receive.

So I was certainly in the inner group from the beginning. In 1989, we spent three months in San Diego with Clive, Rob, David, Timo Teräsvirta and Tony Hall. That was really a fantastic time we had. There was not any real collaboration, but lots of lectures and discussions.

I later collaborated with David on the algorithms for indicator saturation he had suggested. His idea was to have as many dummy regressors as you have observations. By including first one half and then the other half you get a regression estimator, and we found the asymptotic properties of that, see Santos et al. (2008).

Later I continued to work on this with Bent, see Johansen and Nielsen (2009); that led to a number of papers on algorithms, rather than likelihood methods. We analysed outlier detection algorithms and published it in Johansen and Nielsen (2016b), and a paper on the forward search, Johansen and Nielsen (2016a).

### How was cointegration being discussed in the early days?

Clive in Engle and Granger (1987) was the first to suggest that economic processes could be linear combinations of stationary as well as nonstationary processes, and thereby allowing for the possibility that linear combinations could eliminate the nonstationary components. That point of view was a bit difficult to accept for those who worked with economic data. I think the general attitude was that each macroeconomic series had its own nonstationary component.

In Engle and Granger (1987) they modelled the multivariate process as a moving average process with a non-invertible impact matrix, and they showed the surprising result that this “non-invertible” system could in fact be inverted to an autoregressive model (with infinite lag length). Thus a very simple relation was made to the error correction (or equilibrium correction) models studied and used at London School of Economics.

David was analysing macroeconomic data like income and consumption using the equilibrium correcting models, see Davidson et al. (1978). He realized very early that some of the results derived from the model looked more reasonable if you include the spread between income and consumption (for instance) rather than the levels of both. He did not connect it to the presence of nonstationarity.

One of the first applications of the ideas of cointegration was Campbell and Shiller (1987), who studied the present value model in the context of a cointegrating relation in a VAR. The first application of the CVAR methodology was Johansen and Juselius (1990). Here the model is explained in great detail, and it is shown how to test hypotheses on the parameters. Everything is exemplified by data from the Danish and Finnish economies.

Another early paper of the CVAR was an analysis of interest rates, assumed to be nonstationary, while still the spreads could be stationary, as discussed in Hall et al. (1992). These papers contain examples where one can see directly the use and interpretation of cointegration.

### How did you start thinking about identification of cointegrating vectors?

The identification problem for cointegrating relations is the same as the identification problem discussed by the Cowles Commission, who modelled simultaneous equations for macro variables and needed to impose linear restrictions to identify the equations. We were doing something similar, but trying to model nonstationary variables allowing for linear cointegrating relations, and we needed linear restrictions on the cointegrating coefficients  $\beta$  in (1) in order to distinguish and interpret them.

Then one can use the Wald condition for identification, which requires that the matrix you get by applying the restrictions of one equation to the parameters of the other linear equations should have full rank  $r - 1$ , see e.g., Fisher (1966) Theorem 2.3.1. This condition, however, contains the Data Generating Process (DGP) parameter values. This implies that the rank condition cannot be checked in practice, because the DGP is unknown. I asked David what he would do, and he said that he checks the Wald rank condition using

uniform random numbers on the interval  $[0, 1]$  instead of the true unknown parameters. This approach inspired me to look for the mathematics behind this.

### How did you derive the explicit rank conditions for identification?

For simultaneous equations, the restrictions  $R_i$  imposed on the parameters  $\theta_i$  of equation  $i$ ,  $R_i'\theta_i = 0$ , define also a parametrization using the orthogonal complement  $H_i = R_{i\perp}$  and the parameter is  $\theta_i = H_i\phi_i$ . The classical Wald result is that if  $\theta$  denotes the matrix of coefficients of the DGP for the whole system, then  $\theta$  is identified if and only if the rank of the matrix  $R_i'\theta$  is  $r - 1$  for all  $i$ .

I realized soon that I should apply the restrictions not to the parameters but to the parametrizations as given by the orthogonal complements of the restrictions, and the Wald condition can be formulated as the condition  $\text{rank}(R_i'(H_{i_1}, \dots, H_{i_k})) \geq k$  for any set of  $k$  indices not containing  $i$ . This condition does not involve the DGP values and, if identification breaks down, it can be used to find which restrictions are ruining identification.

I reformulated the problem many times and my attention was drawn to operations research, so I asked Laurence Wolsey, when I was visiting the University of Louvain, who suggested the connection to Hall's Theorem (for zero restrictions) and Rado's Theorem (for general linear restrictions), see [Welsh \(1976\)](#). The results are published in [Johansen \(1995a\)](#).

The solution found was incorporated in the computer programs we used when we developed the theory for cointegration analysis. With a moderate amount of equations, the results can be useful to modify the restrictions if they are not identifying, by finding out which restrictions cause the failure of identification.

The value added of this result is the insight: we understand the problem better now, and finding where these conditions fail can help you reformulate better exclusion restrictions. Katarina has developed an intuition for using these conditions, which I do not have. You need to have economic insight to see what is interesting here; for me, it is a nice mathematical result.

I also discussed the result with Rob and he said that it's interesting to see the identification problem being brought back into Econometrics. After Sims' work, identification of systems of equations had been sort of abandoned, because in Sims' words, you had "incredible sets of restrictions".

### You introduced reduced rank regression in cointegration. How did this come about?

In mathematics, you reformulate a problem until you find a solution, and then you sometimes find that someone else has solved the problem—this is what happened with reduced rank regression in cointegration, which I worked out as the Gaussian maximum likelihood estimation in the cointegrated VAR model.

When I first presented the results—later published in [Johansen \(1988b\)](#)—at the European Meeting of the Econometric Society in 1987 in Copenhagen, I was fortunate to have Helmut Lütkepohl in the audience who said: "isn't that just reduced rank regression?". This helped me include references to [Anderson \(1951\)](#), [Velu et al. \(1986\)](#) and to the working paper version of [Ahn and Reinsel \(1990\)](#). Finally, reduced rank regression is also used in limited information calculations, which can be found in many textbooks.

I used Gaussian maximum likelihood to derive the reduced rank estimator, but Bruce Hansen in this Special Issue, [Hansen \(2018\)](#), makes an interesting point, namely that reduced rank regression is a GMM-type estimator, not only a Gaussian Maximum Likelihood solution.

Finally, my analysis revealed a kind of duality between  $\beta$  and  $\alpha_{\perp}$  which can be exploited to see how many models can be analysed by reduced rank regression. As summarized in my book ([Johansen 1996](#)) reduced rank regression can be used to estimate quite a number of different submodels, with linear restrictions on  $\beta$  and/or  $\alpha$  and allowing different types of deterministic terms. But of course it is easy to find sub-models, where one has to use iterative methods to find the maximum likelihood estimator.

### How did you start working on Granger-type representation theorems?

In 1985 I was shown by Katarina the original working paper by Clive before it was published; this was when I started working on cointegration. I started with an autoregressive representation of a process, and found its moving average representation that Clive used as the starting point. I find that a more satisfactory formulation, trying to understand the structure of what he was working on, and I produced the paper on the mathematical structure, [Johansen \(1988a\)](#).

I was looking for something simple in the very complicated general case with processes integrated of any integer order, and I settled to focus on what I called the “balanced case”, that is a relation between variables that are all differenced the same number of times. The balanced case is very simple, and was a way of avoiding a too complicated structure. However, I was focusing on the wrong case, because it is the unbalanced case which is of importance in the I(2) model.

The mathematical structure paper, however, contains “the non-I(2) condition” (see Theorem 2.5 there), which states that  $\alpha'_{\perp} \Gamma \beta_{\perp}$  need to be full rank in I(1) VAR systems in (1) with  $\Pi = \alpha\beta'$ . That came out as just one small result in this large paper, but that was the important result which was missed in the [Engle and Granger \(1987\)](#) paper.

### This links to the I(2) model and its development.

In 1990 Katarina obtained a grant from the Joint Committee of the Nordic Social Sciences Research Council. The purpose was to bring together Ph.D. students in Econometrics together with people working in private and public institutions in the Nordic Countries to teach and develop the theory and the applications of cointegration. We had two to three workshops a year for 6 or 7 years. The work we did is documented in [Juselius \(1994\)](#) [see Figure 3].

In the beginning, Katarina and I would be doing the teaching and the rest would listen, but eventually they took over and presented various applications. It was extremely inspiring to have discussions on which direction the theory should be developed. One such direction was the I(2) model, and I remember coming to a meeting in Norway with the first computer programs for the analysis of the I(2) model on Katarina’s portable Toshiba computer with a liquid crystal screen.

It was a very inspiring system we had, where questions would be raised at one meeting and I would then provide the answers at the next meeting half a year later. Identification was discussed, I(2) was discussed, and computer programs were developed, and people would try them out. I kept the role as the “mathematician” in the group all the time and decided early on that I would not try to go into the Economics.

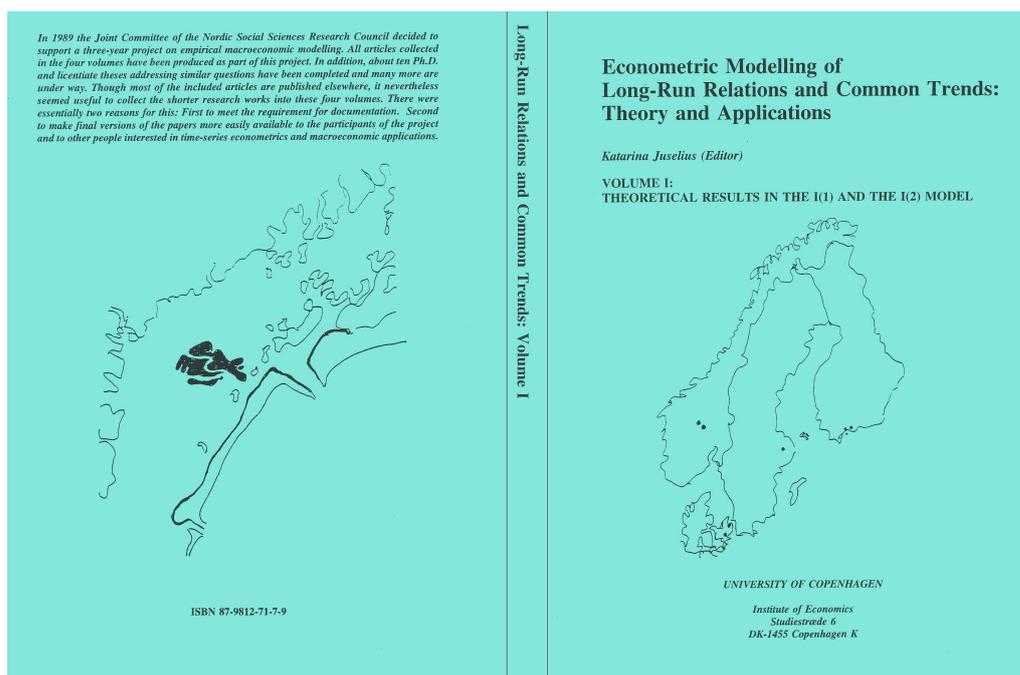
### Which I(2) results came first?

The I(2) model was developed because we needed the results for the empirical analyses in the group, and the first result was the representation theorem, [Johansen \(1992\)](#). This contained the condition for the process generated by the CVAR to have solutions which are I(2), generalizing “the non-I(2) condition” to “the non-I(3) condition”.

The next problem I took up was a systematic way of testing for the ranks of the cointegrating spaces, which I formulated as a two stage analysis for ranks, [Johansen \(1995b\)](#). This problem was taken up by Anders Rahbek and Heino Bohn Nielsen who took over and analysed the likelihood ratio test for the cointegration ranks, [Nielsen and Rahbek \(2007\)](#).

The likelihood analysis for the maximum likelihood estimation of the parameters is from [Johansen \(1997\)](#). When I developed the I(2) model, I realized that the balanced case is not the interesting one. You need relationships for the I(2) processes of the type  $\beta'X_t + \varphi'\Delta X_t$  to reach stationarity, and this is the so-called “multi-cointegration” notion.

I realized from the very beginning that Clive’s structure with the reduced rank matrix in the autoregressive model  $\Pi = \alpha\beta'$  in (1) is an interesting structure. So one wants to see how one can generalize it. This of course can be done in many ways but the collaboration with Katarina on the examples was very inspiring.



**Figure 3.** Front and back cover of Juselius (1994), vol. I (of IV). Areas in black indicate sites where the Nordic workshops took place between 1990 and 1993.

One such example is to take two log price indices  $p_{it}$ ,  $i = 1, 2$ , where each one is I(2), but  $p_{1t} - p_{2t}$  is I(1); one could then have that  $p_{1t} - p_{2t} + \varphi \Delta p_{1t}$  comes down to stationarity, where  $\Delta p_{1t}$  is an inflation rate and  $\varphi$  is some coefficient. She pointed out that the important part of the I(2) model was that it allowed for the combination of levels and differences in a single equation, and this is exactly the unbalanced case. In order to understand this I needed to go back and first work out the representation theory, and then start on the statistical analysis.

### What asymptotic results did you derive first?

The asymptotics for the rank test in the I(1) model came first. I attended a meeting at Cornell in 1987, where I presented the paper on the mathematical structure of error correction models (Johansen 1988a). I included one result on inference, the test for rank. For that you need to understand the likelihood function and the limits of the score and information. I could find many of the results, but the limit distribution of the test for rank kept being very complicated.

At the conference I met Yoon Park who pointed out that the limit distributions had many nuisance parameters, and that one could try to get rid of them. This prompted me to work through the night to see if the nuisance parameters would disappear in the limit. I succeeded and could present the results in my lecture the next day.

So the mathematical structure paper Johansen (1988a) had the rank test in it and its limit distribution, see Section 5 there. The most useful result was that the limit distribution of the test for rank  $r$  is the same as if you test that  $\Pi = 0$  in the CVAR with one lag and  $p - r$  dimensions, that is, a multivariate setup for the analogue of the Dickey-Fuller test.

The limit distribution for the rank test with Brownian motions is something I always showed as a nice result when I lectured on it, but it is in a sense not so useful for analysis, because we don't know its mean, variance, or quantiles. So to produce the tables of the asymptotic distribution you must go back to the eigenvalue problem with random walks and then simulate the distribution for a sufficiently large value of  $T$ .

I think, the next result I worked on was the limit distribution for  $\hat{\beta}$ . It was derived using the techniques that Peter Phillips had developed, see Phillips (1986). He had picked

the right results on Brownian motion from probability and used them to analyse various estimators, and I could simply use the same techniques.

Ted Anderson's reduced rank regression, Peter Phillips' Brownian motions, Phil Howlett's results (about which I found out much later) on the non-I(2) condition (Howlett 1982) were all fundamental to my work, but the reason that I could exploit all these methods and results was my basic training in probability theory, and I am very grateful for the course Patrick Billingsley gave in Copenhagen in 1964–1965.

### **What are recent related results that you find interesting?**

The paper by Onatski and Wang (Onatski and Wang 2018) has some very nice results. They consider a multivariate Dickey Fuller test, testing that  $\Pi = 0$  in the VAR in (1). They let the dimension  $p$  of the system go to infinity proportionally to the number of observations  $T$ , and they get an explicit limit distribution. This is based on results on the eigenvalues of matrices of i.i.d. observations in large dimensions, which has been studied in Mathematics and Statistics. Onatski and Wang have an explicit expression for the limit distribution of the multivariate Dickey Fuller test, called the Wachter distribution.

They refer to the paper Johansen et al. (2005) where we do the simulations to discuss Bartlett's correction. Part of that is simply simulating the multivariate Dickey Fuller test for different dimensions and different  $p$ . And they show that their asymptotic formula fits nicely with our simulations. Extensions to cases with deterministic terms and breaks, and the ones for rank different from 0 should be carefully considered.

### **Tell us about your contribution to fractional cointegration.**

Morten wrote his thesis on fractional processes 2003 at Aarhus University, and I was asked to sit on his committee. Some years later I had formulated and proved the Granger representation theorem for the fractional CVAR (FCVAR) in Johansen (2008), where the solution is a multivariate fractional process of order  $d$ , which cointegrated to order  $d - b$ . We decided to extend the statistical analysis from the usual CVAR to this new model for fractional processes.

The fractional processes had of course been studied by many authors including Peter Robinson and his coauthors, like Marinucci, Hualde and many others. There are therefore many results on the stochastic behaviour of fractional process on which we could build our statistical analysis.

The topic had mostly been dealt with by analyzing various regression estimators and spectral density estimators, where high level assumptions are made on the data generating process. I thought it would be interesting to build a statistical model, where the solution is the fractional process, so one can check the assumptions for the model.

We had the natural framework in the VAR model, and we just needed to modify the definition of differences and work out properties of the solution. From such a model one could then produce (likelihood) estimators and tests, and mimic the development of the CVAR.

We decided, however, to start with the univariate case, simply to get used to the analysis and evaluation of fractional coefficients. We published that in Johansen and Nielsen (2010), and our main results on the FCVAR, that is the fractional CVAR, are in Johansen and Nielsen (2012).

It helped the analysis that for given fractional parameters  $b$  and  $d$ , the FCVAR model can be estimated by reduced rank regression. We found that inference on the cointegrating relations is mixed Gaussian, but now of course using the fractional Brownian motion, so basically all the usual results carry over from the CVAR.

We are currently working on a model where each variable is allowed its own fractional order, yet after suitable differencing, we can formulate the phenomenon of cointegration. The analysis is quite hard with some surprising results. It turns out that inference is asymptotically mixed Gaussian both for the cointegrating coefficients, but also for the difference in fractional order.

**For fractional cointegration, you appear to be attracted more by the beauty of the model and the complexity of the problem, rather than the applications. Is this the case?**

You are absolutely right. There is not a long tradition for the application of fractional processes in Econometrics, even though some of the examples are financial data, where for instance log volatility shows clear sign of fractionality and so do interest rates when measured at high frequency, see [Andersen et al. \(2001\)](#).

Clive and also other people have tried to show that fractionality can be generated by aggregation. [Granger \(1980\)](#) takes a set of AR(1) autoregressive coefficients with a cross-sectional beta distribution between  $-1$  and  $+1$ ; then integrating (aggregating) he gets fractionality of the aggregate. However, if you choose some other distribution, you do not get fractionality.

As another source of fractionality, [Parke \(1999\)](#) considered a sum of white noise components  $\varepsilon_t$  which are dropped from the sum with some given probability. If you choose some specific waiting time distribution, you obtain the spectrum or auto-covariance function of a fractional process. There is also another result by [Diebold and Inoue \(2001\)](#) who show that a Markov switching model generates fractionality. Still we lack economic questions that lead to fractionality.

I read about an interesting biological study of the signal from the brain to the fingers. The experiment was set up with a person tapping the rhythm of a metronome with a finger. After some time the metronome was stopped and the person had to continue tapping the same rhythm for a quarter of an hour. The idea was that the brain has a memory of the rhythm, but it has to send a signal to the fingers, and that is transmitted with an error. The biologist used a long memory process (plus a short memory noise) to model the signal.

**Have you ever discussed fractional cointegration with Katarina?**

No, she refuses to have anything to do with it, because she is interested in Macroeconomics. She feels strongly that the little extra you could learn by understanding long memory, would not be very interesting in Macroeconomics. It will also take her interest away from the essence, and I think she's right. In finance, something else happens. Here you have high frequency data, and that seems a better place for the fractional ideas.

**Tell us about your contributions in survival analysis.**

I spent many years on developing the mathematical theory of product integration, which I used in my work on Markov chains, [Johansen and Ramsey \(1979\)](#). I later collaborated with Richard Gill on a systematic theory of product integration and its application to Statistics, [Gill and Johansen \(1990\)](#). The interest in the statistical application of product integration came when I met Odd Aalen in Copenhagen. He had just finished a Ph.D. on the theory of survival analysis using counting processes, with Lucien Le Cam from Berkeley, and was spending some time in Copenhagen.

Towards the end of his stay, he presented me with a good problem: he asked me if I could find the asymptotic distribution of the Kaplan-Meier estimator, which estimates the distribution function for censored data. As I had worked with Markov chains, I could immediately see that I could write the estimator as a product integral.

Of course this doesn't help anyone, but a product integral satisfies an obvious differential equation. And once you can express the estimator as the solution of a differential equation, you can find the asymptotic distribution, by doing the asymptotics on the equation instead of the solution. So we found the asymptotic distribution of what has later been called the Aalen-Johansen estimator, see [Aalen and Johansen \(1978\)](#).

**How did this come about?**

The breakthrough in this area of Statistics came with the work of David Cox, who in 1972 presented the Cox survival model ([Cox 1972](#)) in which you model the hazard rate. That is, the intensity of the event under consideration, unemployment for instance, in a small interval around time  $t$ , given the past history. The hazard function is allowed to

depend on explanatory regressors. The expression for the likelihood then becomes a special case of the product integral.

In our department Niels Keiding worked with statistical methods applied to medical problems. He got interested in survival analysis and wanted to understand the mathematical theory behind it, so he was teaching the theory of point processes and martingales. A typical example of such problems is to follow a group of patients for a period to see, for instance, how a treatment is helping cure a disease. Ideally you follow all patients for as long as it takes, but in practice you have to terminate the study after a period, so the data is truncated.

The data is made more complicated to work with, because people can leave the study for other reasons, and hence the data is censored. Such data consists of a sequence of time points, and is therefore called a point process. Niels was very active with this type of data and he and his colleagues wrote the book [Andersen et al. \(1992\)](#), describing both applications and the theory of the analysis, including some of my work with product integration with Richard Gill.

### **Did this research have practical implications?**

At the University of Copenhagen a retrospective study of the painters syndrome was conducted. The reason for and time point of retirement were noted for a group of painters, and as control group the same data were recorded for bricklayers. Such data is typically made more complicated by individuals changing profession or moving, or dying during the period of investigation.

One way of analysing such data is to draw a plot of the estimated integrated intensity of retirement due to brain damage (painters' syndrome), which can take into account the censoring. It was obvious from that plot that the risk of brain damage was much higher for painters than for brick layers. This investigation was just a small part of a larger investigation which resulted in changing working conditions for painters, and much more emphasis on water based paint.

### **Tell us about your work on convexity.**

The topic was suggested to me by Hans Brøns shortly after I finished my studies and I had the opportunity to go to Berkeley for a year. The purpose was to write a thesis on the applications of convexity in probability. The important result in functional analysis was the theorem by Hewitt and Savage ([Hewitt and Savage 1955](#)) about representing points in a convex set as a mixture of extreme points. We hoped to find some applications of this result in probability theory.

The simplest example of such a result is that a triangle is a convex set with three extreme points, and putting some weights on the extreme points, we can balance the triangle by supporting it at its center of gravity, which is the weighted average of the extreme points. Another simple example is the set of Markov probability matrices, with positive entries adding to one in each row. The extreme points are of course the matrices you get by letting each row be a unit vector.

A more complicated example is the following: in probability theory there is a well known Lévy-Khintchine representation theorem, which says that the logarithm of the characteristic function of an infinitely divisible distribution is an integral of a suitable kernel with respect to a measure on the real line. It is not difficult to show that these functions form a compact convex set. One can identify the extreme points to be either Poisson distributions or the Gauss distribution. The representation theorem then follows from the result of Hewitt and Savage. This provided a new understanding and a new proof of the Lévy-Khintchine result.

Another result I worked on I still find very intriguing. If you consider a non-negative concave continuous function on the unit circle, normalized to have an integral of 1, then such functions form a convex compact set. The challenge is to find the extreme points. I found a large class of extreme points, which have the property that they are piecewise flat.

I needed a further property: that at each corner of the function, where the flat pieces meet, there are only three pieces meeting. Imagine a pyramid with four sides, so that four lines meet at the top. This function is not an extreme point, but if you cut the tip off the pyramid, then at each of the four corners created will have only three sides meeting, and then it is an extreme point.

The set of functions has the strange property that each point in the set (a concave function) can be approximated uniformly close by just one extreme point.

### **Tell us about other models you worked on.**

I once collaborated with a group of doctors who were investigating the metabolism of sugar, say, by the liver in order to find a good measure of liver capacity. The data was the concentration of sugar in the blood at the inlet and the outlet of the liver. There were three models around at the time. One modelled the measurement at the inlet and the other at the outlet.

In developing the model we used an old idea of August Krogh—Winner of the Nobel Prize in Physiology or Medicine in 1920 “for his discovery of the capillary motor regulating mechanism”—of modelling the liver as a tube lined with liver cells on the inside, such that the concentration of sugar at the inlet would be higher than the concentration at the outlet. This physiological model gave the functional form of the relation between the inlet and outlet concentrations, which we used to model the data.

We used the data to compare the three models and found out that ours was the best. I worked on this with Susanne Keiding, see [Keiding et al. \(1979\)](#) and [Johansen and Keiding \(1981\)](#). We analyzed the data by nonlinear regression that used the mathematics of the model, the so called Michaelis-Menten kinetics.

### **It is not so common to check model assumptions as suggested by David Hendry. What is your view on this?**

In my own training in mathematics, I could not use a theorem without checking its assumptions. This is obviously in the nature of mathematics. Our education in Statistics was based on Mathematics, so for me it was natural to check assumptions when you have formulated a model for the data.

At the Economics Department of the University of Copenhagen Katarina held for 9 years a “Summer School in the Cointegrated VAR model: Methodology and Applications”. In total we had about 300 participants. I would give the theoretical lectures and Katarina would tell them about how they need to model the data in order to investigate the economic theories.

The main aspect of the course, however, was that they brought their own data and had a specific economic question in mind concerning their favourite economic theory. They spent all afternoons for a month doing applied work, choosing and fitting a model, checking assumptions of the model, and comparing the outcome with economic knowledge they had. Katarina would supervise the students, and they were encouraged to discuss among themselves. They had never tried such a thing and learned a tremendous amount.

On a smaller scale, most courses should include some software for doing econometric analysis. Such programs would often produce output for different models (different lag length, cointegration rank, and deterministic terms) as well as misspecification tests. It seems a good idea to include the interpretation of such output in a course, so one can have a discussion of what it means for a model to be wrong, and how one can react to change it for the better.

### **Is the ability to check assumptions related to likelihood models—i.e., models with a likelihood?**

A very simple regression model, that everyone knows about, is to assume for two series  $X_t, Y_t$  that they are linearly related  $Y_t = \beta X_t + \varepsilon_t$ , and the error terms  $\varepsilon_t$  are mutually independent and independent of the  $X_t$ s. Obviously, without specifying a precise family of distributions for the error term, one cannot talk about likelihood methods. So what do we

gain by assuming Gaussian errors for example? We can derive the least squares method, but in fact Gauss did the opposite. He derived the distribution that gave you least squares.

There is another application of a parametric model that is also useful. Suppose you realize, somehow, that the regression residuals are autocorrelated. Then you would like to change the estimation method, and a method for doing that is to build a new model, which can tell you how to change the method. This is where an autoregressive model for  $\varepsilon_t$  would, after a suitable analysis of the likelihood, the score, and information, tell you what to do.

So I think the answer is that the likelihood method tells you how to get on with the analysis, and what to do when your assumptions fail. In this light, one can see that the failure of conducting inference using a cointegrating regression can be remedied by formulating the CVAR with Gaussian errors and then derive the methods from the likelihood.

### **How did your training help you, and what does this suggest for education needs in the econometric profession?**

I think what helped me in Econometrics is the basic training I received in Mathematical Statistics. At the University of Copenhagen, the degree in Statistics (“candidatus statisticae”) was introduced in 1960, when Anders Hjorth Hald was appointed professor of Statistics. He appointed Hans Brøns as the second teacher.

Anders Hald had been working for a number of years as statistical consultant and later as professor of Statistics at the Economic Department at the University of Copenhagen. He was inspired by the ideas of R. A. Fisher at Cambridge, and our Statistics courses were based on the concept of a statistical model and analysis of estimators and test statistics derived from the likelihood function. The purpose was to educate statisticians to do consulting with other scientists, but also to develop new statistical methods. The teaching was research-based and included many courses in mathematics.

The teaching attracted very good students. In those days, if you had a background in mathematics, there was essentially only one thing you could use it for, and that was teaching at high school. I was very interested in mathematics but did not want to teach at high school, so I became a statistician. This would allow me to collaborate with scientists from other fields, something that I would enjoy a lot.

Our department grew over the years to about 10 people and we discussed teaching and research full time. It was a very inspiring environment for exchanging ideas and results. We regularly had visitors from abroad, who stayed for a year doing teaching and research. For my later interest in Econometrics the course by Patrick Billingsley in 1964–1965 was extremely useful, as it taught me advanced probability theory. He was lecturing on what was to become the now classical book on convergence of probability measures (Billingsley 1968) while he was visiting Copenhagen.

### **What should one do when the model doesn't fit?**

There does not seem to be an easy set of rules for building models, so it is probably best to gain experience by working with examples. Obviously a model should be designed so it can be used for whatever purpose the data was collected. But if the first attempt fails, because it does not describe the data sufficiently well, it would possibly be a good idea to improve the model by taking into account in what sense it broke down.

You could look for more explanatory variables, including dummies for outliers, different variance structure or perhaps study related problems from other countries, say, to get ideas about what others do. It is my strong conviction that the parametric model can help you develop new estimation and test methods to help to find a model which better takes into account the variation of the data.

As students, we only analysed real life data and sometimes even had a small collaboration with the person who had taken the measurements. Our role would be to help building a statistical model and formulate the relevant hypotheses to be investigated in collaboration with the user. Then we would do the statistical analysis of the model based

on the likelihood function. With this type of training we learned to discuss and collaborate with others.

### How and why should models be built?

I do not think that there are general rules for model building, partly because models can serve so many different purposes. By considering many examples, it is my opinion that you can develop a feeling for what you do with the kind of problems you are investigating. But if you change field, you probably have to start from scratch. Thus the more experience you have with different types of models, the more likely it is that you can find a good model next time you need it.

I personally find that the main reason for building and analyzing models is that you want to be able to express your own understanding of the phenomenon to other people. The mathematical language has this nice property that you can communicate concepts in a precise way. I think about the model as a consistent way of formulating your understanding of the real world.

It is interesting to consider an average of measurements as something very relevant and useful in real life. The model for i.i.d. variables includes the nice result of the law of large numbers, and gives us a way of relating an average to an abstract concept of expectation in a model. But perhaps more important than that, is that it formulates assumptions, under which the result is valid, and that gives you a way of checking if the average is actually a good idea to calculate for the data at hand.

Another practically interesting concept is the notion of spurious correlation, which for nonstationary data can be very confusing, if you do not have a model as a basis for the discussion, see for instance the discussion of [Yule \(1926\)](#). It was the confusion about the notion of correlation (for nonstationary time series variables) that inspired the work of Clive on the concept of cointegration.

### Could you elaborate on the theory and practice of likelihood methods in Econometrics?

Econometric textbooks often contain likelihood methods, but they do not have a prominent position. There are only few books which are based on likelihood methods from the beginning, as for instance [Hendry and Nielsen \(2007\)](#). In the space between models and methods, the weight is usually on the methods and how they perform under various assumptions. There are two good reasons to read textbooks, one is that you can then apply the methods, and the other is that you can then design new methods.

When R. A. Fisher introduced likelihood analysis, the starting point was obviously the model, and the idea is that the method for analysing the data should be derived from the model. In fact it is a unifying framework for deriving methods that people would be using anyway. Thus instead of remembering many estimators and statistics, you just need to know one principle, but of course at the price of some mathematical analysis.

By deriving the method from first principles you also become more aware of the conditions for the analysis to hold, and that helps checking for model misspecification, which again can help you modify the model if it needs improvement. It is clear that the likelihood requires a model, and the likelihood analysis is a general principle for deriving estimators and test statistics; yet it usually also requires a lot of mathematical analysis, and the solutions often need complicated calculations.

It is, however, not a solution to all problems, there are counter examples. In particular, when the number of parameters increases, the maximum likelihood estimator can be inconsistent. A standard example is to consider observations  $(X_i, Y_i)$ ,  $i = 1, \dots, n$  which are independent Gaussian with mean  $(\mu_i, \mu_i)$  and variance  $\sigma^2$ . In this simple situation  $\hat{\sigma}^2 \rightarrow \frac{1}{2}\sigma^2$  in probability, the so-called “Neyman-Scott Paradox”.

**What are the alternative approaches with respect to a well-specified statistical model?**

The simple regression model, where the calculations needed to find the estimator is the starting point, is an example of an algorithm which is often taken as the starting point, and which does not require a statistical model. The statistical model is needed, when you want to test hypotheses on the coefficients, and the parametric statistical model is useful if you want to derive new methods.

Of course there exists many methods, expert systems, based on complicated nonlinear regressions. I am not an expert on these, but I note that the people behind them collaborate with statisticians.

**So what needs to be avoided is the use of Statistics without knowledge of it. Correct?**

Sounds like a good idea! Many people think that Statistics is a set of well developed methods that we can just use. I think that can be a bit dangerous, and highly unsatisfactory for the users. It would of course be lovely, but a bit unrealistic, that all users should have a deep understanding of Statistics, before they could use a statistical method. I explained elsewhere the summer course we had in Copenhagen, where the students are put in a situation where they have to make up their minds about what to do, and that certainly improves learning.

**Are statisticians especially trained to collaborate?**

As a statistician, you study all the classical models about Poisson regression and two-way analysis of variance, survival analysis and many more. If the exercises contain real data, you will learn to formulate and build models and choose the right methods for analyzing them. It is of course in the nature of the topic that if you are employed later in a medical company doing controlled clinical trials, then you will have to collaborate with the doctors.

The education should therefore also try to put the students in situations, where such skills can be learned. The problem is of course that if you end up in an insurance company or in an economics department you probably need different specializations. So in short I think the answer to your question is Yes! the students should be trained to collaborate.

**Hence, is Statistics a science at the service of other sciences?**

Of course Statistics as a field has a lot of researchers working at Universities on teaching and developing the field, but most statisticians work in industry or public offices, pharmaceutical companies, insurance companies, or banks.

Another way of thinking about it was implemented by my colleague Niels Keiding. In 1978 he started a consulting service for the medical profession at the University of Copenhagen, using a grant from the Research Council. The idea was to help the university staff in the medical field getting expert help with their statistical problems, from planning controlled clinical trials, to analysing data of various sort. This has been a tremendous success and now is a department at the University with around 20 people working full time in this, as well as some teaching of Statistics for the doctors.

**Any message on the publication process?**

I remember when I was in Berkeley many years ago, in 1965, I took a course with Lester Dubins, who had just written a book called "How to gamble if you must", [Dubins and Savage \(1965\)](#). I was then working with the coauthor on my first paper, [Johansen and Karush \(1966\)](#). I must have been discussing publications with Lester, and he kindly told me "But you have to remember, Søren: every time you write a paper and get it published, it becomes slightly more difficult for everybody else to find what they want".

This carried a dual message on the benefit of advancing knowledge and the associated increased cost in retrieving information. Fortunately, this cost has been greatly reduced by the current powerful internet search engines available.

**Author Contributions:** R.M. and P.P. have contributed to all phases of the work. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Authors gratefully acknowledge useful comments and correction from the interviewee on the draft of the article.

**Conflicts of Interest:** The authors declare no conflict of interest. Information and views set out in this paper are those of the authors and do not necessarily reflect the ones of the institutions of affiliation.

## Notes

<sup>1</sup> Accessed on 3 April 2022.

<sup>2</sup> See Engle and Granger (1987) and Johansen (1988b).

<sup>3</sup> For a description and the current version of PcGive see: <https://www.doornik.com/doc/PcGive/>, accessed on 3 April 2022.

## References

- Aalen, Odd O., and Søren Johansen. 1978. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* 5: 141–50.
- Ahn, Sung K., and Gregory C. Reinsel. 1990. Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association* 85: 813–23. [CrossRef]
- Andersen, Per K., Ørnulf Borgan, Richard D. Gill, and Niels Keiding. 1992. *Statistical Models Based on Counting Processes*. New York: Springer.
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens. 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61: 43–76. [CrossRef]
- Anderson, Theodore Wilbur. 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics* 22: 327–51. [CrossRef]
- Billingsley, Patrick. 1968. *Convergence of Probability Measures*. New York: John Wiley & Sons.
- Campbell, John Y., and Robert J. Shiller. 1987. Cointegration and tests of present value models. *Journal of Political Economy* 95: 1062–88. [CrossRef]
- Cox, David R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34: 187–202.
- Davidson, James E. H., David F. Hendry, Frank Srba, and Stephen Yeo. 1978. Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom. *The Economic Journal* 88: 661–92. [CrossRef]
- Diebold, Francis X., and Atsushi Inoue. 2001. Long memory and regime switching. *Journal of Econometrics* 105: 131–59.
- Dubins, Lester E., and Leonard Savage. 1965. *How to Gamble if You Must: Inequalities for Stochastic Processes*. New York: McGraw-Hill.
- Engle, Robert F., and Byung Sam Yoo. 1991. Cointegrated time series: An overview with new results. In *Long Run Economic Relationships: Readings in Cointegration*. Edited by Robert F. Engle and Clive W. J. Granger. Oxford: Oxford University Press, chp. 12, pp. 237–66.
- Engle, Robert F., and Clive W. J. Granger. 1987. Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55: 251–76. [CrossRef]
- Fisher, Franklin M. 1966. *The Identification Problem in Econometrics*. New York: McGraw-Hill.
- Gill, Richard D., and Søren Johansen. 1990. A Survey of Product-Integration with a View Toward Application in Survival Analysis. *The Annals of Statistics* 18: 1501–55. [CrossRef]
- Granger, Clive W. J. 1980. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14: 227–38. [CrossRef]
- Hall, Anthony D., Heather M. Anderson, and Clive W. J. Granger. 1992. A cointegration analysis of treasury bill yields. *The Review of Economics and Statistics* 74: 116–26. [CrossRef]
- Hansen, Bruce E. 2018. Johansen's reduced rank estimator is GMM. *Econometrics* 6: 26. [CrossRef]
- Hendry, David F., and Bent Nielsen. 2007. *Econometric Modeling: A Likelihood Approach*. Princeton: Princeton University Press.
- Hewitt, Edwin, and Leonard J. Savage. 1955. Symmetric measures on cartesian products. *Transactions of the American Mathematical Society* 80: 470–501. [CrossRef]
- Hillebrand, Eric, Søren Johansen, and Torben Schmith. 2020. Data revisions and the statistical relation of global mean sea level and surface temperature. *Econometrics* 8: 41. [CrossRef]
- Howlett, Phil G. 1982. Input retrieval in finite dimensional linear systems. *Journal of Australian Mathematical Society (Series B)* 23: 357–82. [CrossRef]
- Johansen, Søren. 1988a. The mathematical structure of error correction models. *Contemporary Mathematics* 80: 359–86.

- Johansen, Søren. 1988b. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12: 231–54. [CrossRef]
- Johansen, Søren. 1992. A representation of vector autoregressive processes integrated of order 2. *Econometric Theory* 8: 188–202. [CrossRef]
- Johansen, Søren. 1995a. Identifying restrictions of linear equations with applications to simultaneous equations and cointegration. *Journal of Econometrics* 69: 111–132.
- Johansen, Søren. 1995b. A statistical analysis of cointegration for I(2) variables. *Econometric Theory* 11: 25–59. [CrossRef]
- Johansen, Søren. 1996. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd ed. Oxford: Oxford University.
- Johansen, Søren. 1997. Likelihood analysis of the I(2) model. *Scandinavian Journal of Statistics* 24: 433–62. [CrossRef]
- Johansen, Søren. 2008. A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* 24: 651–76. [CrossRef]
- Johansen, Søren. 2019. Cointegration and adjustment in the CVAR( $\infty$ ) representation of some partially observed CVAR(1) models. *Econometrics* 7: 2. [CrossRef]
- Johansen, Søren, and Bent Nielsen. 2009. An analysis of the indicator saturation estimator as a robust regression estimator. In *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. Edited by Jennifer Castle and Neil Shephard. Oxford: Oxford University Press, pp. 1–36.
- Johansen, Søren, and Bent Nielsen. 2016a. Analysis of the forward search using some new results for martingales and empirical processes. *Bernoulli* 22: 1131–83. [CrossRef]
- Johansen, Søren, and Bent Nielsen. 2016b. Asymptotic theory of outlier detection algorithms for linear time series regression models. *Scandinavian Journal of Statistics* 43: 321–48. [CrossRef]
- Johansen, Søren, and Bent Nielsen. 2019. Boundedness of M-estimators for linear regression in time series. *Econometric Theory* 35: 653–83. [CrossRef]
- Johansen, Søren, and Jack Karush. 1966. On the semimartingale convergence theorem. *The Annals of Mathematical Statistics* 37: 690–94. [CrossRef]
- Johansen, Søren, and Katarina 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. [CrossRef]
- Johansen, Søren, and Morten Ørregaard Nielsen. 2010. Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* 158: 51–66.
- Johansen, Søren, and Morten Ørregaard Nielsen. 2012. Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica* 80: 2667–732. [CrossRef]
- Johansen, Søren, and Morten Ørregaard Nielsen. 2018. The cointegrated vector autoregressive model with general deterministic terms. *Journal of Econometrics* 202: 214–29. [CrossRef]
- Johansen, Søren, and Fred L. Ramsey. 1979. A bang-bang representation for 3x3 embeddable stochastic matrices. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 47: 107–18. [CrossRef]
- Johansen, Søren, and Susanne Keiding. 1981. A family of models for the elimination of substrate in the liver. *Journal of Theoretical Biology* 89: 549–56. [CrossRef]
- Johansen, Søren, Henrik Hansen, and Stefano Fachin. 2005. A Simulation Study of some Functionals of Random Walk. Technical Report, University of Copenhagen. Available online: <http://web.math.ku.dk/~sjo/> (accessed on 3 April 2022).
- Juselius, Katarina, ed. 1994. *Econometric Modelling of Long-run Relations and Common Trends: Theory and Applications*. Copenhagen: University of Copenhagen, Institute of Economics, vols. I–IV.
- Keiding, Susanne, Søren Johansen, Inge Midtbøll, Alan Rabøl, and Leif Christiansen. 1979. Ethanol elimination kinetics in human liver and pig liver in vivo. *American Journal of Physiology-Endocrinology and Metabolism* 237: E316. [CrossRef]
- Nielsen, Heino Bohn, and Anders Rahbek. 2007. The likelihood ratio test for cointegration ranks in the I(2) model. *Econometric Theory* 23: 615–37. [CrossRef]
- Onatski, Alexei, and Chen Wang. 2018. Alternative asymptotics for cointegration tests in large VARs. *Econometrica* 86: 1465–78. [CrossRef]
- Parke, William R. 1999. What is Fractional Integration? *The Review of Economics and Statistics* 81: 632–38. [CrossRef]
- Phillips, Peter C. B. 1986. Understanding spurious regressions in econometrics. *Journal of Econometrics* 33: 311–40. [CrossRef]
- Phillips, Peter C. B. 1995. Fully modified least squares and vector autoregression. *Econometrica* 63: 1023–78. [CrossRef]
- Santos, Carlos, David F. Hendry, and Søren Johansen. 2008. Automatic selection of indicators in a fully saturated regression. *Computational Statistics* 23: 317–35. correction 337–39. [CrossRef]
- Velu, Raja P., Gregory C. Reinsel, and Dean W. Wichern. 1986. Reduced rank models for multiple time series. *Biometrika* 73: 105–18. [CrossRef]
- Welsh, Dominic J. A. 1976. *Matroid Theory*. London: Academic Press.
- Yule, G. Udny. 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. [CrossRef]