

The analysis of nonstationary time series using regression, correlation and cointegration.

Søren Johansen*

August 20, 2012

Abstract

There are simple well-known conditions for the validity of regression and correlation as statistical tools. We analyse by examples the effect of nonstationarity on inference using these methods and compare them to model based inference using the cointegrated vector autoregressive model. Finally we analyse some monthly data from US on interest rates as an illustration of the methods.

Keywords: Regression, correlation, cointegration, model based inference, likelihood inference

JEL Classification: C32.

1 Introduction

This paper is based on a lecture given at the 56th Session of the International Statistical Institute in Lisbon 2007, and part of the introduction is taken from there.

In his presidential address at the meeting in the Royal Statistical Society November 17, 1925 Udne Yule stated

"It is fairly familiar knowledge that we sometimes obtain between quantities varying with the time (time-variables) quite high correlations to which we cannot attach any physical significance whatever, although under the ordinary test the correlation would be held to be certainly "significant"."

He goes on to show a plot of the proportion of Church of England marriages to all marriages for the years 1866-1911 inclusive, and in the same diagram, the mortality per 1.000 persons for the same years, see Figure 1.

He comments

*University of Copenhagen and CREATES University of Aarhus, Denmark

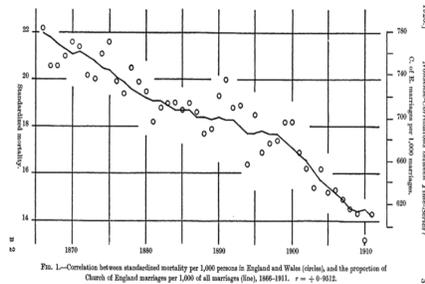


Figure 1: The proportion of Church of England marriages to all marriages for the years 1866-1911 (line), and the mortality per 1.000 persons for the same years (circles), Yule (1926).

"Evidently there is a very high correlation between the two figures for the same year: The correlation coefficient actually works out at +0.9512."

He then points out that

"When we find that a theoretical formula applied to a particular case gives results which common sense judges to be incorrect, it is a generally as well to examine the particular assumptions from which it was deduced and see which of them are inapplicable to the case in point."

In order to describe the probability assumptions behind the "ordinary test" he invents an experiment which consists of writing corresponding numbers of (X_t, Y_t) on cards and defines the distribution of the correlation coefficient as what you get when you draw the cards at random and calculate the correlation coefficient. In this way he simulated the distribution of the empirical correlation coefficient calculated from two independent i.i.d. processes, from two independent random walks, and from two independent cumulated random walks, the latter having a U-shaped distribution, see Figure 2. The interpretation of Figure 2 is that for $T = 10$, there is no evidence that the empirical correlation will be close to the theoretical correlation, which is zero. In fact the results also hold for very large T , as we shall discuss that later.

Thus, 80 years ago Yule pointed to what is wrong with just calculating correlation coefficients without checking the stationarity assumption behind the interpretation, and he suggested calling such correlations "nonsense correlations".

Granger and Newbold (1974) take up the point and note that

"It is very common to see reported in applied econometric literature, time series regression equations with an apparently high degree of fit, as measured by the coefficient of multiple correlation R^2 but with an extremely

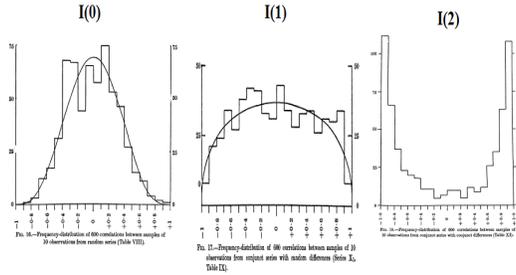


Figure 2: Simulation for $T = 10$ of the distribution of the empirical correlation coefficient for independent i.i.d. processes, $I(0)$, independent random walks, $I(1)$, and independent cumulated random walks, $I(2)$, Yule (1926).

low value for the Durbin-Watson statistic. We find it very curious that whereas virtually every textbook on econometric methodology contains explicit warnings of the dangers of autocorrelated errors this phenomenon crops up so frequently in well-respected applied work."

They show by simulation of ARIMA models that regressions can be quite misleading. The important paper by Phillips (1986) solved the problem of finding the asymptotic distribution of correlation and regression coefficients, when calculated from a class of nonstationary time series. Thus the problem and its solution has been known for a long time but we still find numerous examples of misunderstandings in applied and theoretical work.

The paper by Hoover (2003) discusses Reichenbach's principle of the common cause, that is, "if event X and Y are correlated, then either X causes Y , Y causes X , or X and Y are joint effects of a common cause (one that renders X and Y conditionally probabilistically independent)", see Sober (2001). A counter example to this principle, according to Sober (2001), consists in considering Venetian sea levels and British bread prices. Sober claims they are truly correlated but not causally connected by construction, therefore neither causes the other and there can be no common cause. Hoover points out that the statement "truly correlated" is based on calculating the empirical correlation coefficient, which is clearly a case of a spurious or nonsense correlation, as both series trend with time.

Thus the problem pointed out and analyzed by simulation by Yule in 1926, followed up by Granger and Newbold (1974), and finally solved by Phillips (1986) is still present in applied and theoretical work.

The purpose of this paper is to discuss regression and correlation which are commonly applied statistical techniques, and emphasize the assumptions underlying the analysis in order to point out some instances, where these method cannot be used in a routinely fashion, namely when the variables are nonstationary, either because they contain a deterministic trend or a random walk. We then want to demonstrate that by building a statistical model that allows the variables to nonstationary, using the cointe-

grated vector autoregressive model, we can express our understanding of the variation of the data and apply that to pose question of economic relevance.

2 Two approaches to inference

As an illustrative example we consider two time series X_t and $Y_t, t = 1, \dots, T$, and a substantive theory that X influences Y in a linear fashion formulated as $Y = \beta X$. For given data such a relation does not hold and there is most often no substantive theory for the deviations, and to quote Haavelmo (1943) ‘we need a stochastic formulation to make simplified relations elastic enough for applications’. We therefore introduce the error term ε_t and write the relation as a statistical relation

$$Y_t = \beta X_t + \varepsilon_t, \quad t = 1, \dots, T. \quad (1)$$

We want to estimate the parameter β and evaluate its uncertainty in order to be able to test hypotheses, for instance that $\beta = 0$, which means that in this model there is no influence of X_t on Y_t .

There are two common approaches to deal with inference in linear regression and correlation analysis.

- *The method based approach*

Regression is used to estimate the effect of X on Y by calculating the least squares estimators and the residual error variance using the formulae

$$\hat{\beta} = \frac{\sum_{t=1}^T X_t Y_t}{\sum_{t=1}^T X_t^2}, \quad (2)$$

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T (Y_t - \hat{\beta} X_t)^2. \quad (3)$$

These are then used to conduct asymptotic inference by comparing the t-ratio

$$t_{\beta=\beta_0} = \left(\sum_{t=1}^T X_t^2 \right)^{1/2} \frac{\hat{\beta} - \beta_0}{\hat{\sigma}}, \quad (4)$$

with the quantiles of a standard normal distribution. Regression works well if the estimates $\hat{\beta}$ and $\hat{\sigma}^2$ are close to their theoretical counterparts, β and σ^2 , and if the asymptotic distribution of $t_{\beta=\beta_0}$ is close to the Gaussian distribution. We discuss below some examples, where there is no relation between the empirical regression estimates and their theoretical values.

Correlation is used to describe the linear relation between two observed variables Y and X . We define the *theoretical* correlation coefficient between Y and X as

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(Y)Var(X)}}, \quad (5)$$

and the *empirical* correlation coefficient between two time series Y_t and X_t is calculated as

$$\hat{\rho} = \frac{\sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^T (X_t - \bar{X})^2 \sum_{t=1}^T (Y_t - \bar{Y})^2}}. \quad (6)$$

both (5) and (6) are commonly called correlation, which causes some confusion. We distinguish these concepts here using the qualifications *empirical* and *theoretical*, and we discuss below some examples where the empirical correlation is not related to the theoretical correlation. Thus correlation and regression are summary statistics and the danger of using them is that for some types of data there is no relation between the empirical and the theoretical concepts.

- *The model based approach*

In the model based approach we first formulate a hypothetical mechanism for how the data is generated and then derive the relevant statistical methodology by an analysis of the likelihood function. One such model, which also specifies how X_t is generated, is

$$Y_t = \beta X_t + \varepsilon_{1t}, \quad (7)$$

$$X_t = \xi X_{t-1} + \varepsilon_{2t}, \quad (8)$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})$ are i.i.d. Gaussian with variances σ_1^2 and σ_2^2 and covariance σ_{12} . We then conduct inference using the method of maximum likelihood and likelihood ratio test. These methods, however, require that the assumptions of the model are carefully checked in any particular application in order to show that the model describes the data well, so that the results of asymptotic inference, which are derived under the assumptions of the model, can be applied.

It is well known that linear regression analysis can be derived as the Gaussian maximum likelihood estimator provided that ε_t in (1) are i.i.d. $N(0, \sigma^2)$, and X_t is nonstochastic. Similarly if (X_t, Y_t) are i.i.d. Gaussian with variances σ_1^2, σ_2^2 and covariance σ_{12} , then the theoretical correlation is $\rho = \sigma_{12}/\sigma_1\sigma_2$, and the maximum likelihood estimator of ρ is $\hat{\rho}$ given in (6). Thus there is no clear-cut distinction between the method based approach and the model based approach, but a difference of emphasis, in the sense that regression and correlation are often applied uncritically by "pressing the button on the computer", and the model based method requires more discussion and checking of assumptions. Thus the model based approach we express our understanding of the data by the choice of model. We then apply the model to formulate precise questions and hypotheses, which can possibly be falsified by a statistical analysis of the data. In this way we can actually learn something new about the data.

We discuss below some examples where regression analysis and correlation analysis cannot be used, and hence one has to take properties of the data into account in order to avoid incorrect inference.

3 Regression and Correlation

We specify a set of conditions under which regression and correlation methods work well, and then analyse some examples where the methods do not work.

3.1 Regression

We formulate the statistical assumptions of the regression model (1) as

Algorithm 1 *We assume that*

- $\varepsilon_1, \dots, \varepsilon_T$ are innovations in the sense that they are i.i.d. $(0, \sigma^2)$ and ε_t is independent of X_1, \dots, X_t , $t = 1, \dots, T$
- X_1, \dots, X_T are stochastic (or deterministic) variables for which the normalized sum of squares is convergent to a deterministic limit

$$n_T^{-1} \sum_{t=1}^T X_t^2 \xrightarrow{P} \Sigma > 0,$$

for some sequence $n_T \rightarrow \infty$.

Here \xrightarrow{P} denotes convergence in probability. These assumptions are enough to show that

$$E(n_T^{-1/2} \varepsilon_t X_t | X_1, \dots, X_t) = 0, \tag{9}$$

and

$$n_T^{-1} \sum_{t=1}^T \text{Var}(\varepsilon_t X_t | X_1, \dots, X_t) \xrightarrow{P} \sigma^2 \Sigma. \tag{10}$$

Apart from a technical assumption on the third moment, these relations show that $n_T^{-1/2} \varepsilon_t X_t$ is a so-called martingale difference sequence, and that the sum of its successive conditional variances converges to a deterministic limit. This again implies that one can apply the Central Limit Theorem for martingales, see Hall and Heyde (1980). The theorem shows, in this particular case, that

$$n_T^{-1/2} \sum_{t=1}^T \varepsilon_t X_t \xrightarrow{d} N(0, \sigma^2 \Sigma), \tag{11}$$

where \xrightarrow{d} means convergence in distribution.

From (2) and (3) we find that

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{t=1}^T Y_t X_t}{\sum_{t=1}^T X_t^2} = \frac{\sum_{t=1}^T (\beta X_t + \varepsilon_t) X_t}{\sum_{t=1}^T X_t^2} \\ &= \beta + n_T^{-1/2} \frac{n_T^{-1/2} \sum_{t=1}^T \varepsilon_t X_t}{n_T^{-1} \sum_{t=1}^T X_t^2}, \\ \hat{\sigma}^2 &= T^{-1} \left[\sum_{t=1}^T \varepsilon_t^2 - \frac{(n_T^{-1/2} \sum_{t=1}^T \varepsilon_t X_t)^2}{n_T^{-1} \sum_{t=1}^T X_t^2} \right].\end{aligned}$$

The result (11) then implies that

$$\hat{\beta} \xrightarrow{P} \beta, \tag{12}$$

$$\hat{\sigma} \xrightarrow{P} \sigma^2, \tag{13}$$

$$n_T^{1/2} (\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 \Sigma^{-1}), \tag{14}$$

$$\begin{aligned}t_{\beta=\beta_0} &= \left(\sum_{t=1}^T X_t^2 \right)^{1/2} \frac{(\hat{\beta} - \beta_0)}{\hat{\sigma}} \\ &\xrightarrow{d} N(0, 1).\end{aligned} \tag{15}$$

The first two results state that the estimators are close to the theoretical values, that is, the estimators are consistent, and the third that $\hat{\beta}$ is asymptotically normally distributed. The last result is used to conduct asymptotic inference and test the hypothesis that $\beta = \beta_0$, by comparing a t -ratio with the quantiles of the normal distribution. In this sense the regression method works well when the above Assumption 1 is satisfied.

3.2 Correlation

We formulate the condition that guarantees that the theoretical correlation can be measured by the empirical correlation.

Algorithm 2 *We assume that (Y_t, X_t) is a stationary (and ergodic) time series with finite second moments.*

It follows from the Law of Large Numbers, see for example Stock and Watson (2003, p. 578), that if Assumption 2 is satisfied, then

$$\hat{\rho} \xrightarrow{P} \rho. \tag{16}$$

Thus in order for the calculation of an empirical correlation to make sense as an approximation to the theoretical correlation, it is important to check Assumption 2.

3.3 Examples

The first example shows that we have to choose different normalizations depending on which regressor variable we have.

Example 1. (Regression) If $X_t = 1$ we have $\sum_{t=1}^T X_t^2 = T$ and we choose $n_T = T$, and if $X_t = t$, then $\sum_{t=1}^T X_t^2 = \sum_{t=1}^T t^2 \approx \frac{1}{3}T^3$, and we choose $n_T = T^3$ so that $n_T^{-1} \sum_{t=1}^T X_t^2 = T^{-3} \sum_{t=1}^T t^2 \rightarrow \frac{1}{3}$. If X_t is an ergodic process with $E(X_t^2) < \infty$, then the Law of Large Numbers for ergodic processes shows that $T^{-1} \sum_{t=1}^T X_t^2 \xrightarrow{P} E(X_t^2)$. Hence we use the normalization $n_T = T$ in this case. This, however, is not enough to apply the regression method because we also need ε_t to be independent of the regressor, see Assumption 1.

Consider for instance the model defined in (7) and (8) for $|\xi| < 1$, which defines an ergodic process X_t . Then $T^{-1} \sum_{t=1}^T X_t^2 \xrightarrow{P} \text{Var}(X_t) = \sigma_2^2/(1 - \xi^2)$, but note that (9) fails because

$$\begin{aligned} E(\varepsilon_{1t}X_t|X_1, \dots, X_t) &= X_t E(\varepsilon_{1t}|\varepsilon_{2t}) = \sigma_{12}\sigma_2^{-2}X_t\varepsilon_{2t} \\ &= \sigma_{12}\sigma_2^{-2}X_t(X_t - \xi X_{t-1}) \neq 0, \end{aligned}$$

when ε_{1t} is not independent of the regressor, and we cannot apply the asymptotic theory unless $\sigma_{12} = 0$. Thus even for stationary processes an autocorrelated regressor variable is enough to invalidate the simple regression.

If, however, we take the model based approach we can analyse model (7) and (8) as follows. We first find, assuming Gaussian errors, the conditional mean of Y_t given X_1, \dots, X_t , where $\omega = \sigma_{12}\sigma_2^{-2}$

$$\begin{aligned} E(Y_t|X_1, \dots, X_t) &= \beta X_t + E(\varepsilon_{1t}|X_1, \dots, X_t) \\ &= \beta X_t + \omega(X_t - \xi X_{t-1}). \end{aligned}$$

This means we can replace (7) and (8) by the equations

$$Y_t = \beta X_t + \omega(X_t - \xi X_{t-1}) + \varepsilon_t^{1,2} \quad (17)$$

$$X_t = \xi X_{t-1} + \varepsilon_{2t}. \quad (18)$$

Because the error terms $\varepsilon_t^{1,2} = \varepsilon_{1t} - \omega\varepsilon_{2t}$ and ε_{2t} are independent, we can analyse the equations separately and estimate ξ by regressing X_t on X_{t-1} , and determine $\beta + \omega$ and $-\omega\xi$ by regression of Y_t on X_t and X_{t-1} , and that allows one to derive consistent asymptotically Gaussian estimators for the parameter of interest β . Thus by analysing the model we can determine the relevant regression analysis. ■

Example 2 (Correlation) Let again the data be generated by (7) and (8) for $|\xi| < 1$. Then X_t, Y_t is an ergodic process and the empirical correlation, $\hat{\rho}$, will converge towards

the theoretical correlation ρ given by

$$\begin{aligned} & \frac{Cov(\beta X_t + \varepsilon_{1t}, X_t)}{\sqrt{Var(\beta X_t + \varepsilon_{1t})Var(X_t)}} \\ &= \frac{\beta\sigma_2^2 + \sigma_{12}(1 - \xi^2)}{\sqrt{[\beta^2\sigma_2^2 + \sigma_1^2(1 - \xi^2) + 2\beta\sigma_{12}(1 - \xi^2)]\sigma_2^2}}, \end{aligned}$$

using the results that $Var(X_t) = \sigma_2^2/(1 - \xi^2)$ and $Cov(X_t, \varepsilon_{1t}) = \sigma_{12}$.

If X_t instead is generated by

$$X_t = \gamma t + \varepsilon_{2t},$$

then

$$Y_t = \beta\gamma t + \beta\varepsilon_{2t} + \varepsilon_{1t}$$

and correlation analysis does not work. We find $E(X_t) = \gamma t$ and $E(Y_t) = \beta\gamma t$, so that the theoretical correlation is

$$\begin{aligned} \rho &= \frac{E(Y_t - \beta\gamma t)(X_t - \gamma t)}{\sqrt{E(Y_t - \beta\gamma t)^2 E(X_t - \gamma t)^2}} \\ &= \frac{E((\varepsilon_{1t} + \beta\varepsilon_{2t})\varepsilon_{2t})}{\sqrt{E(\varepsilon_{1t} + \beta\varepsilon_{2t})^2 E(\varepsilon_{2t}^2)}} \\ &= \frac{\sigma_{12} + \beta\sigma_2^2}{\sqrt{(\beta^2\sigma_2^2 + 2\beta\sigma_{12} + \sigma_1^2)\sigma_2^2}}, \end{aligned}$$

that is, the correlation between the stochastic error term of Y_t and X_t .

The empirical correlation, however, measures something quite different. It contains the averages $\bar{X} = \gamma\bar{t} + \bar{\varepsilon}_2$, where $\bar{t} = T^{-1}\sum_{t=1}^T t = (T+1)/2$, so that $X_t - \bar{X} = \gamma(t - \bar{t}) + \varepsilon_{2t} - \bar{\varepsilon}_2$ and $Y_t - \bar{Y} = \beta(X_t - \bar{X}) + \varepsilon_{1t} - \bar{\varepsilon}_1 = \beta\gamma(t - \bar{t}) + \beta(\varepsilon_{2t} - \bar{\varepsilon}_2) + \varepsilon_{1t} - \bar{\varepsilon}_1$ are dominated by the linear trend and we have

$$\hat{\rho} \xrightarrow{P} \frac{\beta}{|\beta|} = \pm 1,$$

if $\beta \neq 0$. Thus, if the regressor is trending with a linear trend, there is no relation between the empirical correlation, which is often very close to ± 1 , and the theoretical correlation which measures a correlation between the error terms see the example in Figure ???. The mistake made is of course that \bar{X} and \bar{Y} do not measure the expectation of X_t and Y_t .

The model based approach leads to estimating $(\beta\gamma, \gamma)$ from a regression of (Y_t, X_t) on t and that gives consistent asymptotically Gaussian estimators of the parameters of interest without using or misusing any measure of correlation.

A good check of the relevance of the empirical correlation is simply to calculate it recursively, that is, define $\hat{\rho}_t$ based on data up to time t , and then plot it and check if it is reasonably constant in t , see Figure 3.

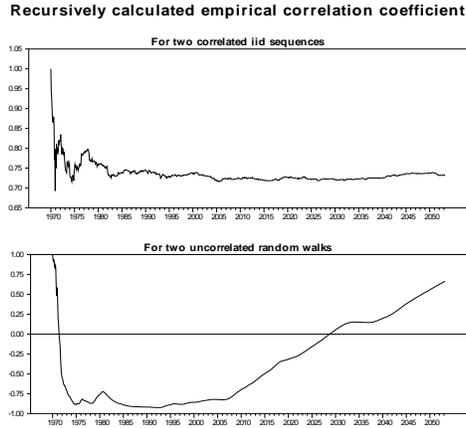


Figure 3: The recursively calculated correlation coefficient. Note how the value stabilizes for the two i.i.d. sequences at the theoretical value $1/\sqrt{2} = 0.71$, whereas for the two uncorrelated random walks the value does not settle down.

Next we give an example where one cannot normalize $\sum_{t=1}^T X_t^2$ so that the limit exists as a deterministic limit, and hence that simple regression analysis may fail. ■

Example 3. (Random walk regressor) A very special situation occurs in example (7) and (8) if $\xi = 1$, so that X_t is stochastic and nonstationary in the sense that,

$$X_t = \sum_{i=1}^t \varepsilon_{2i} + X_0.$$

In this case $E(X_t|X_0) = X_0$ and $Var(X_t|X_0) = \sigma_2^2 t$ which increases to infinity, and something completely different happens. Let us first find out how to normalize $E(\sum_{t=1}^T X_t^2|X_0)$, because such a normalization could be a good candidate for the normalization of $\sum_{t=1}^T X_t^2$. We find

$$\begin{aligned} E\left(\sum_{t=1}^T X_t^2|X_0\right) &= \sum_{t=1}^T E(X_t^2|X_0) = \sigma_2^2 \sum_{t=1}^T t \\ &= \frac{1}{2}\sigma_2^2 T(T+1). \end{aligned}$$

Thus a good choice seems to be $n_T = T^2$, which at least makes sure that the mean converges when normalized by T^2 .

Unfortunately $T^{-2} \sum_{t=1}^T X_t^2$ does not converge to a deterministic limit but to a stochastic variable. The detailed theory of this is quite complicated because it involves Brownian motion, see Davidson and MacKinnon (2004, p 616).

Brownian motion is a continuous stochastic process defined on the unit interval for which $B(0) = 0$, $B(u)$ is distributed as $N(0, u)$ and for $0 \leq u_1 < u_2 < u_3 \leq 1$ we have that $B(u_2) - B(u_1)$ is independent of $B(u_3) - B(u_2)$. The main reason for this to be interesting in the present context, is that we can approximate Brownian motion by random walks, because on the interval, $0 \leq u \leq 1$,

$$T^{-1/2} \sum_{1 \leq i \leq Tu} \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \sigma_1 B_1(u) \\ \sigma_2 B_2(u) \end{pmatrix}, \quad (19)$$

Thus a Brownian motion can be thought of as a random walk with a very large number of steps, and that is how its properties are studied using stochastic simulation. The two Brownian motions in (19) are correlated with correlation $\rho = \sigma_{12}/\sigma_1\sigma_2$.

Two fundamental results about Brownian motion are

$$\begin{aligned} T^{-2} \sum_{t=1}^T X_t^2 &\xrightarrow{d} \sigma_2^2 \int_0^1 B_2(u)^2 du, \\ T^{-1} \sum_{t=1}^T X_t \varepsilon_{1t} &\xrightarrow{d} \sigma_2 \sigma_1 \int_0^1 B_2(u) (dB_1). \end{aligned}$$

These limits are stochastic variables, and for our purpose the main result is that the product moments should be normalized by T^2 and T respectively to get convergence. It follows that Assumption 1 is not satisfied because the limit of $T^{-2} \sum_{t=1}^T X_t^2$ is stochastic, and we cannot count on the results (12) to (16) to be correct.

If we run a regression anyway, we can calculate the t-ratio and find its limit

$$\begin{aligned} & \left(\sum_{t=1}^T X_t^2 \right)^{1/2} (\hat{\beta} - \beta) \\ &= \frac{T^{-1} \sum_{t=1}^T \varepsilon_{1t} X_t}{\sqrt{T^{-2} \sum_{t=1}^T X_t^2}} \xrightarrow{d} \frac{\sigma_1 \int_0^1 B_2(u) dB_1(u)}{\sqrt{\int_0^1 B_2(u)^2 du}}. \end{aligned} \quad (20)$$

If ε_{1t} and ε_{2t} are independent, one can show that the limit distribution (20) is $N(0, \sigma_1^2)$, and therefore (12) and (15) hold anyway, whereas (14) is different, because we get instead a so-called mixed Gaussian distribution of the limit of $T(\hat{\beta} - \beta)$. So despite the fact $\hat{\beta}$ is not asymptotically normally distributed one can still test hypotheses on β using the usual t-ratio, but the independence of ε_{1t} and ε_{2t} is crucial for this last result. A simulation is show in Figure 4. It is seen that for $\rho = 0$, where there is independence between the regressor and the error term in the regression, the distribution of the t-ratio is very close to Gaussian, but the distribution of $T(\hat{\beta} - \beta)$ is centered around zero, but far from Gaussian. Thus it is important to normalize $T(\hat{\beta} - \beta)$ by an estimate of the asymptotic conditional standard error, or square root of the information, which gives the t -ratio, which is asymptotically Gaussian, and not by its variance which will not give an asymptotically Gaussian variable, see the right hand panel in Figure 4.

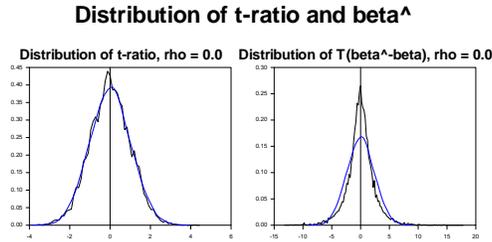


Figure 4: The plots show simulations of the t-ratio, (15) or (20), and $T(\hat{\beta} - \beta)$, (14), in the regression of $Y_t = \beta X_t + \varepsilon_{1t}$, when X_t is a random walk, $\Delta X_t = \varepsilon_{2t}$, see Example 3, and ε_{1t} is independent of ε_{2t} . Each plot contains a Gaussian density for comparison. It is seen that the t-ratio has approximately a Gaussian distribution and that the estimator normalized by T has a distribution with longer tails than the Gaussian. The densities are based upon 10.000 simulations of $T = 100$ observations.

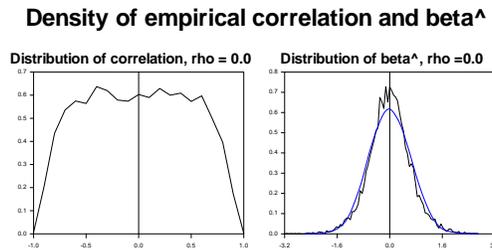


Figure 5: The left panel shows the distribution of the empirical correlation between two independent random walks, S_{1t} and S_{2t} . The right panel shows the similar results for the empirical regression coefficient of S_{1t} on S_{2t} . The results are the same for higher values of T , thus there is no tendency to converge to $\rho = 0$. The densities are based upon 10.000 simulations of $T = 100$ observations.

The result in (20) shows that applying a simple regression analysis, without checking Assumption 1, can be seriously misleading, and we next want to show how we can solve the problem of inference by analysing the model, that generated the data.

If $\xi = 1$, then $\Delta X_t = \varepsilon_{2t}$, and we find the equations, see (17) and (18)

$$\begin{aligned} Y_t &= \beta X_t + \omega \Delta X_t + \varepsilon_t^{1,2}, \\ \Delta X_t &= \varepsilon_{2t}. \end{aligned} \tag{21}$$

Here the errors are independent and

$$\begin{aligned} \text{Var}(\varepsilon_t^{1,2}) &= \text{Var}(\varepsilon_{1t} | \varepsilon_{2t}) \\ &= \sigma_1^2 - \sigma_{12}^2 \sigma_2^{-2} = \sigma_{1|2}^2. \end{aligned}$$

Equation for Y_t is analysed by regression of Y_t on X_t and ΔX_t to find an asymptotically Gaussian estimator for β . This simple modification of the regression problem solves the inference problem. We still get an expression like (20)

$$\left(\sum_{t=1}^T X_t^2 \right)^{1/2} (\hat{\beta} - \beta) \xrightarrow{d} \frac{\sigma_{1|2} \int_0^1 B_2(u) dB_{1|2}(u)}{\sqrt{\int_0^1 B_2(u)^2 du}}, \tag{22}$$

where $B_{1|2}(u) = B_1(u) - \rho B_2(u)$ is independent of B_2 , so the limit is mixed Gaussian and inference can be conducted using the usual t-ratio and comparing it to the quantiles of the Gaussian distribution.

The correlation analysis of Y_t and X_t leads to a theoretical correlation (conditional on X_0)

$$\begin{aligned} \rho_t &= \frac{\text{Cov}(\beta X_t + \varepsilon_{1t}, X_t | X_0)}{\sqrt{\text{Var}(\beta X_t + \varepsilon_{1t} | X_0) \text{Var}(X_t | X_0)}} \\ &= \frac{\beta \sigma_2^2 t + \sigma_{12}}{\sqrt{[\beta^2 \sigma_2^2 t + \sigma_1^2 + 2\beta \sigma_{12}] \sigma_2^2 t}}, \end{aligned}$$

which converges to $\rightarrow \beta|\beta|^{-1} = \pm 1$ if $\beta \neq 0$. Thus for large t we find a value ± 1 depending on the sign of β .

The empirical correlation coefficient has the same limit as $\hat{\rho}$

$$\frac{\beta \sum_{t=1}^T (X_t - \bar{X})^2 + \sum_{t=1}^T (\varepsilon_{1t} - \bar{\varepsilon}_1)(X_t - \bar{X})}{\sqrt{\sum_{t=1}^T (\beta(X_t - \bar{X}) + \varepsilon_{1t})^2 \sum_{t=1}^T (X_t - \bar{X})^2}}$$

which converges in probability $\xrightarrow{P} \beta|\beta|^{-1} = \pm 1$, if $\beta \neq 0$, so that it estimates the limit of the theoretical correlation for $T \rightarrow \infty$.

This model with $\xi = 1$ is an example of two nonstationary variables with a stationary linear combination, that is, a model for cointegration. ■

Example 4. (Spurious correlation and regression)
 Assume (X_t, Y_t) are generated by the equations

$$\begin{aligned}\Delta Y_t &= \varepsilon_{1t}, \\ \Delta X_t &= \varepsilon_{2t},\end{aligned}$$

where we assume that $\sigma_{12} = 0$, so X_t and Y_t are independent of each other. The theoretical correlation is, conditioning on initial values,

$$\begin{aligned}\rho_t &= \frac{\text{Cov}(Y_t, X_t | Y_0, X_0)}{\sqrt{\text{Var}(Y_t | Y_0) \text{Var}(X_t | X_0)}} = \frac{t\sigma_{12}}{\sqrt{t\sigma_1^2 t\sigma_2^2}} \\ &= \frac{\sigma_{12}}{\sigma_1\sigma_2} = 0.\end{aligned}$$

If we calculate the empirical correlation, (6), all product moments should be normalized by T^{-2} and we find, using the notation $\bar{B} = \int_0^1 B(u)du$, the limit in distribution of $\hat{\rho}$:

$$\frac{\int_0^1 (B_2(u) - \bar{B}_2)(B_1(u) - \bar{B}_1)du}{\sqrt{\int_0^1 (B_1(u) - \bar{B}_1)^2 du \int_0^1 (B_2(u) - \bar{B}_2)^2 du}}.$$

Thus $\hat{\rho}$ does not converge to zero or any other value but is stochastic even for infinitely many observations. This is the phenomenon observed by Yule (1926). A regression of Y_t on X_t gives similarly

$$\hat{\beta} \xrightarrow{d} \frac{\int_0^1 B_2(u)B_1(u)du}{\int_0^1 B_2(u)^2 du},$$

where the stochastic limit is totally unrelated to any theoretical measure of the effect of X_t on Y_t . Thus by calculation of a correlation or a regression coefficient one may infer an effect of X_t on Y_t , when absolutely no effect is present because they are independent, see Figure 5.

If the independent random walks contain a trend, we model them as

$$\Delta Y_t = \varepsilon_{1t} + \mu_1, \tag{23}$$

$$Y_t = \sum_{i=1}^t \varepsilon_{1i} + \mu_1 t + Y_0,$$

$$\Delta X_t = \varepsilon_{2t} + \mu_2, \tag{24}$$

$$X_t = \sum_{i=1}^t \varepsilon_{2i} + \mu_2 t + X_0,$$

where we again assume $\sigma_{12} = 0$. In this case, the trend is dominating the random walk,

and we find that for instance

$$\begin{aligned} & T^{-1}(X_t - \bar{X}) \\ &= T^{-1} \sum_{i=1}^t \varepsilon_{2i} - T^{-1} \sum_{t=1}^T [T^{-1} \sum_{i=1}^t \varepsilon_{2i}] \\ &+ \mu_2 \left(\frac{t}{T} - \frac{T+1}{2T} \right) \xrightarrow{P} \mu_2(u - 1/2), \end{aligned}$$

for $t/T \rightarrow u$, because $T^{-1} \sum_{i=1}^t \varepsilon_{2i} \xrightarrow{P} 0$. It follows that because $\sum_{t=1}^T (t - \bar{t})^2 \approx T^3/3$ we get

$$\hat{\rho} \xrightarrow{P} \frac{\mu_2 \mu_1}{|\mu_2 \mu_1|} = \pm 1,$$

if $\mu_1 \mu_2 \neq 0$. Thus, despite the fact that Y_t and X_t are stochastically independent, an empirical correlation suggests something quite different.

The regression coefficient satisfies similarly

$$\hat{\beta} \xrightarrow{P} \frac{\mu_1}{\mu_2},$$

which is the ratio of the slopes of the trends, which makes some sense, but an analysis of the data, using model (23) and (24), would find a linear trend in each variable and estimates of μ_1 and μ_2 which would contain more information.

It is therefore very easy to calculate an empirical correlation between two variables that are completely uncorrelated, but which each depend on the same third variable, like here a time trend. It is important in the calculation of correlations to replace $E(X_t)$ and $E(Y_t)$ by reasonable estimates, not use averages. ■

4 The cointegrated vector autoregressive model

Cointegration was introduced in econometrics by Granger (1981) because many macro variables show nonstationarity of the random walk type, but also clear co-movement. We present in this section the cointegrated vector autoregressive model (CVAR) and apply it in the next section to analyse some interest data. The purpose is to demonstrate that by modelling the data, taking into account the nonstationarity, we can actually learn something new about the economic relations underlying the variation of the data.

The paper by Engle and Granger (1987) contains the first statistical analysis of cointegration using regression methods, and Phillips (1991) modified the regression approach to allow for valid inference. The analysis of cointegration and model based inference in the vector autoregressive framework was initiated by Johansen (1988). The technique of cointegration is described in most text book on times series econometrics and many computer programs are available, see for instance *Cats for Rats*, (Dennis *et al.* 2005), which was used for the calculations in Section 5. For a systematic account of

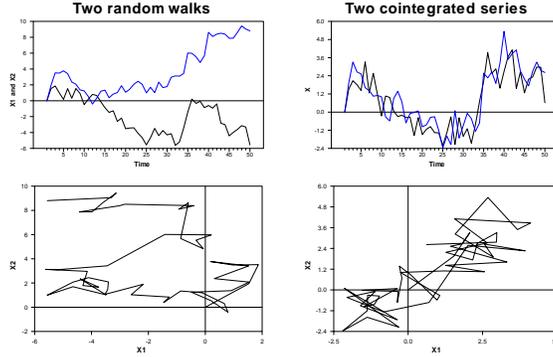


Figure 6: Plots of integrated series generated by equations (25) and (26). To the left are two random walks ($\eta = \tau = 0$). To the right are two cointegrated nonstationary processes ($\gamma = 1, \tau = -1/2, \eta = 1/2$). Note how they follow each other in the upper panel and move around the line $Y - X = 0$ in the lower panel.

the theory, see Johansen (1996), and for applications the monograph by Juselius (2006) is recommended. A recent survey with more details is given in Johansen (2006).

Below we give a simple example of such a model and discuss briefly the statistical analysis of the model.

4.1 An example of a model for cointegration

We consider two variables X_t and Y_t which are generated by the equations for $t = 1, \dots, T$

$$\Delta Y_t = \tau(Y_{t-1} - \gamma X_{t-1}) + \varepsilon_{1t}, \quad (25)$$

$$\Delta X_t = \eta(Y_{t-1} - \gamma X_{t-1}) + \varepsilon_{2t}. \quad (26)$$

The special choices of $\tau = -1$, $\eta = 0$, and $\gamma = \beta$ give the model (7) and (8) with a redefinition of the error term. Each equation is linear in past variables, but note that the levels Y_{t-1} and X_{t-1} enter only through the same linear combination $U_{t-1} = Y_{t-1} - \gamma X_{t-1}$ in both equations. We call U_{t-1} the disequilibrium error and think of the relation $Y = \gamma X$ as an equilibrium relation, to which the variables react with adjustment coefficients τ and η respectively.

It is seen that the equation for $U_t = Y_t - \gamma X_t$ is

$$\Delta U_t = (\tau - \gamma\eta)U_{t-1} + \varepsilon_{1t} - \gamma\varepsilon_{2t},$$

so that U_t is an autoregressive process with one lag, which is stationary if $|1 + \tau - \gamma\eta| < 1$. By eliminating U_{t-1} from (25) and (26) we get

$$\eta\Delta Y_t - \tau\Delta X_t = \eta\varepsilon_{1t} - \tau\varepsilon_{2t},$$

which, by summation, shows that

$$\eta Y_t - \tau X_t = \sum_{i=1}^t (\eta \varepsilon_{1i} - \tau \varepsilon_{2i}) + \eta Y_0 - \tau X_0 = S_t,$$

where S_t , is a random walk and hence nonstationary.

The solution of the equations can be expressed as

$$(X_t, Y_t)' = (\eta\gamma - \tau)^{-1}(S_t - \eta U_t, \gamma S_t - \tau U_t)'$$

which equals

$$\begin{pmatrix} 1 \\ \gamma \end{pmatrix} \begin{pmatrix} \eta \\ -\tau \end{pmatrix}' \sum_{i=1}^t \varepsilon_i - \begin{pmatrix} \eta \\ \tau \end{pmatrix} U_t. \quad (27)$$

This is a special case of the general formula below, see (29).

That is, the model produces nonstationary variables, each of which is composed of a stationary and a random walk. The linear combination $(1, -\gamma)$ eliminates the common random walk (common trend) and makes the linear combination stationary.

This is expressed by saying that (Y_t, X_t) is nonstationary but cointegrated with cointegrating vector $(1, -\gamma)$ and common stochastic trend S_t , see Granger (1981).

Note that the variables are both modelled and treated similarly, unlike in a regression of Y_t on X_t . Thus for instance if $Y_t - \gamma X_t$ is stationary then so is $\gamma^{-1} Y_t - X_t$, so we can normalize on either one of them, provided the coefficient is nonzero. A cointegration relation is a relation between variables not a causal relation.

4.2 The general vector autoregressive model and its solution

The vector autoregressive model with two lags and a constant term for a p -dimensional process X_t is given by the equations,

$$\mathcal{H}_r : \Delta X_t = \alpha \beta' X_{t-1} + \Gamma \Delta X_{t-1} + \mu + \varepsilon_t, \quad (28)$$

where ε_t i.i.d. $N_p(0, \Omega)$, $t = 1, \dots, T$, and (α, β) are $p \times r$ matrices. Note that we need the values X_{-1} and X_0 as initial values in order to be able to determine the process recursively.

We define the polynomial

$$\phi(z) = \det((1 - z)I_p - \alpha \beta' z - \Gamma(1 - z)z).$$

In order to avoid explosive processes we assume that the roots of $\phi(z) = 0$ satisfy either $|z| > 1$ or $z = 1$. Under a further regularity condition, the solution is nonstationary with stationary differences and given by

$$X_t = C \sum_{i=1}^t \varepsilon_i + C \mu t + \sum_{n=0}^{\infty} C_i^* (\varepsilon_{t-i} + \mu). \quad (29)$$

The matrix C satisfies $\beta' C = 0$ and $C\alpha = 0$, and C_i^* are functions of α, β , and Γ . Note that $\beta' X_t = \beta' \sum_{n=0}^{\infty} C_i^*(\varepsilon_{t-i} + \mu)$ is stationary because $\beta' C = 0$, and note that the trend disappears if $\mu = \alpha\rho'$ because $C\alpha = 0$, and in this case there is no linear trend in the solution only a level $\sum_{n=0}^{\infty} C_i^* \mu$, and $E(\beta' X_t) + \rho' = 0$. The representation (27) is a special case of (29) for a bivariate system where $\beta = (1, -\gamma)'$ and $\alpha = (\tau, \eta)$, so that $\alpha'_{\perp} = (\eta, -\tau)$. The model defined in (7) and (8) is a special case of (28) for $\alpha_1 = -1$, $\alpha_2 = 0$ and $\beta' = (1, -\gamma)$.

Thus we have seen that

- X_t is nonstationary with linear trend, $C\mu t$, and ΔX_t is stationary.
- $\beta' X_t = \sum_{n=0}^{\infty} \beta' C_i^*(\varepsilon_{t-i} + \mu)$ is stationary, so that X_t is cointegrated with r cointegrating relations β and disequilibrium error $\beta' X_t - E(\beta' X_t)$.
- X_t has $p - r$ common stochastic trends, $\alpha'_{\perp} \sum_{i=1}^t \varepsilon_i$, where α_{\perp} is $p \times (p - r)$ of full rank and $\alpha' \alpha_{\perp} = 0$.

4.3 Statistical inference in the cointegrated VAR model

It is important to emphasize that before inference can be made in this model the assumptions of the model should be carefully checked. Thus we have to fit a lag length so that the residuals are close to being i.i.d. We therefore plot the residuals and their autocorrelation function, see Figure 8 for an example. The Gaussian assumption is not so important for the analysis, but the assumption that the error term is i.i.d. is crucial for the application of the result from the asymptotic theory below.

Thus briefly summarize, we can conduct inference as follows

- First determine the lag length needed to describe the data and check the assumptions behind the model, in particular the independence of the residuals.
- Find the cointegration rank and estimate and interpret the cointegrating relation.
- Simplify the model by testing coefficients to zero.

4.4 The test for cointegrating rank

The rank of α and β is the number of cointegrating relations and it is important either to check ones knowledge of the rank, or estimate it from the data. The statistical formulation starts by considering the unrestricted vector autoregressive model

$$\mathcal{H}_p : \Delta X_t = \Pi X_{t-1} + \Gamma \Delta X_{t-1} + \mu + \varepsilon_t, \quad (30)$$

where ε_t i.i.d. $N(0, \Omega)$ and Π, Γ, μ , and Ω are unrestricted. If we denote

$$\varepsilon_t(\Pi, \Gamma, \mu) = \Delta X_t - \Pi X_{t-1} - \Gamma \Delta X_{t-1} - \mu,$$

then the conditional Gaussian log likelihood function, $\log L(\mathcal{H}_p)$, given the initial values X_{-1} and X_0 , is apart from a constant, given by

$$\begin{aligned}
& -\frac{T}{2}[\log \det(\Omega)] \\
& + \text{tr}\{\Omega^{-1}T^{-1} \sum_{t=1}^T \varepsilon_t(\Pi, \Gamma, \mu)\varepsilon_t(\Pi, \Gamma, \mu)'\}.
\end{aligned} \tag{31}$$

Note that likelihood functions are usually based upon the joint density of the data. This is not possible for nonstationary variables, like random walks, as there is no joint density. We therefore condition on X_0 and X_{-1} , and consider the conditional density of X_1, \dots, X_T given X_0 and X_{-1} . It follows that, conditional on initial values, the (conditional) maximum likelihood estimators of $(\Pi, \Gamma, \mu, \Omega)$ in (30) can be found by multivariate regression of ΔX_t on X_{t-1} , ΔX_{t-1} , and a constant. The maximized likelihood function, $L_{\max}(\mathcal{H}_p)$, can be found from (31) by inserting the maximum likelihood estimators $(\hat{\Pi}, \hat{\Gamma}, \hat{\mu}, \hat{\Omega})$.

The hypothesis of r cointegrating relations is formulated as in model (28)

$$\Pi = \alpha\beta',$$

where α and β are $p \times r$ matrices. It turns that maximum likelihood estimators can be calculated explicitly by an eigenvalue problem, even though this is a nonlinear maximization problem, see Anderson (1951) for reduced rank regression or Johansen (1996) for the application to cointegration. This gives estimates $(\check{\alpha}, \check{\beta}, \check{\Gamma}, \check{\mu}, \check{\Omega})$ and the maximized value, $L_{\max}(\mathcal{H}_r)$, calculated from (31). From this we calculate the likelihood ratio test

$$-2 \log LR(\Pi = \alpha\beta') = -2 \log \frac{L_{\max}(\mathcal{H}_r)}{L_{\max}(\mathcal{H}_p)}.$$

The asymptotic distribution of this statistic is a functional of Brownian motion, which generalizes the so-called Dickey-Fuller test, see Dickey and Fuller (1981), for testing a unit root in a univariate autoregressive model. The asymptotic distribution does not depend on parameters, but depends on the type of deterministic terms and different tables are provided by simulations, because the distributions are analytically quite intractable, see Johansen (1996, Ch. 15). It should be noted that the asymptotic distribution is not a χ^2 distribution as one often finds when applying likelihood methods, see Table 1 for an example.

After the rank is determined, we often normalize on one of the variables to avoid the indeterminacy in the choice of coefficients. When that is done, one can find the asymptotic distribution of the remaining parameter estimators in order to be able to test hypotheses on these, using either likelihood ratio tests statistics or t -test statistics. Thus the only nonstandard test is the test for rank, and all subsequent likelihood ratio tests in the model are asymptotically distributed as $\chi^2(f)$, where f is the number of restrictions being tested.

4.5 Asymptotic distribution of the coefficients of the cointegrating relation

Unlike usual regression, as described in Section 3, the estimators of the parameters in the cointegrating relation are not asymptotically Gaussian. Nevertheless one can estimate scale factors, $\hat{\tau}_i$, so that

$$t_{\beta_i=\beta_{i0}} = \hat{\tau}_i^{-1}(\hat{\beta}_i - \beta_{i0}) \xrightarrow{d} N(0, 1). \quad (32)$$

Thus one can use these t-ratios for testing hypotheses on individual coefficients, for instance that they are zero. In general one can also test any linear (or nonlinear) hypothesis on the cointegrating parameters using a likelihood ratio test, which is asymptotically distributed as $\chi^2(f)$, where f is the number of restrictions tested.

A simple example of maximum likelihood estimation is given in model (21), where the scale factor can be chosen as $\hat{\tau} = \hat{\sigma}_{11.2}^{1/2}(\sum_{t=1}^T x_t^2)^{-1/2}$, and the limit is Gaussian because $B_{1|2}$ is independent of B_2 in (22).

4.6 Regression analysis of cointegrating relations

The cointegration coefficients can also be estimated by regression, provided we know the value of r , but inference is difficult in the sense that running a regression of X_{1t} on X_{2t}, \dots, X_{pt} will give consistent estimators of the cointegrating coefficients, but the corresponding t -ratios will not converge to the normal distribution, and one cannot find scale factors so that (32) holds. This was illustrated in Example 3 above, where the equations for $\xi = 1$, become

$$\begin{aligned} Y_t &= \beta X_t + \varepsilon_{1t}, \\ \Delta X_t &= \varepsilon_{2t}. \end{aligned}$$

This is an example of two cointegrated series, where the usual t-test leads to a strange limit distribution if $\sigma_{12} \neq 0$, see (20). The problem of how to modify the regression approach by finding a nonparametric estimator of the so-called long-run variance, $C\Omega C'$, was solved by Phillips (1991).

If, however, X_t contains a trend, then the analysis is different because a regression will in fact give valid inference because $\sum_{t=1}^T X_t^2(\hat{\beta} - \beta)^2$ is given by

$$\begin{aligned} & \frac{(\sum_{t=1}^T \varepsilon_{1t} X_t)^2}{\sum_{t=1}^T X_t^2} \\ &= \frac{(\sum_{t=1}^T \varepsilon_{1t} (\sum_{i=1}^t \varepsilon_{2i} + \mu t))^2}{\sum_{t=1}^T (\sum_{i=1}^t \varepsilon_{2i} + \mu t)^2} \\ &\approx \frac{(\sum_{t=1}^T \varepsilon_{1t} t)^2}{\sum_{t=1}^T t^2} \end{aligned}$$

which converges in distribution to $\chi^2(1)$. The reason for the change of result is that the trend dominates the random walk asymptotically in this case.

5 An illustrative example

As an illustration of the cointegration methodology we next give an analysis of US monthly interest rates in the period 1987:1 to 2006:1. This is the period when Greenspan was the chairperson of the Federal Reserve System and avoids the recent turmoil of the financial crisis. The data is taken from IMF's financial database and consists of the following four interest rates: the federal funds rate i_{ff} , the 6 month treasury bills rate, the 3 year and 10 year bond rates, denoted i_{6m} , i_{3y} , i_{10y} respectively. A more thorough analysis of these data is given in Giese (2008) who also give the references to earlier analyses.

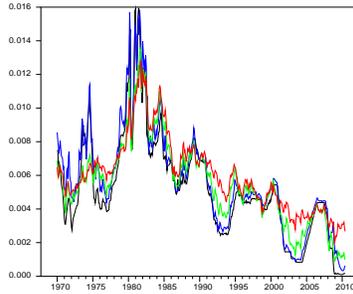


Figure 7: The graph shows the four interest rates i_{ff} , i_{6m} , i_{3y} , and i_{10y} from 1987:1 to 2006:1. Note how they move together and that they do not appear to be stationary.

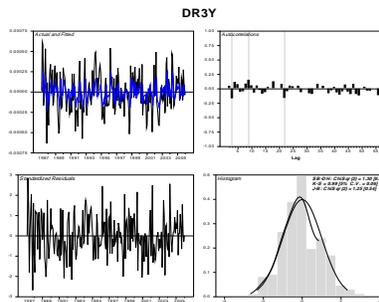


Figure 8: These plots show how well the VAR(2) fits the data for one of the interests i_{3y} . The upper left hand panel shows the data in differences and the fitted values. Below are the standardized residuals. To the right is the autocorrelation function for the estimated residuals and their histogram.

The first task is to find a reasonable statistical description of the data and we try with the VAR model. After some investigation we end up with model with two lags and an unrestricted constant, see Figure 8.

As an important diagnostic tool we calculate the roots of the companion where a unit root will indicate the type of nonstationarity we can handle with these models. We find there are two roots close to one, $0.979 \pm 0.009i$ and the rest are smaller than 0.798, indicating two unit roots.

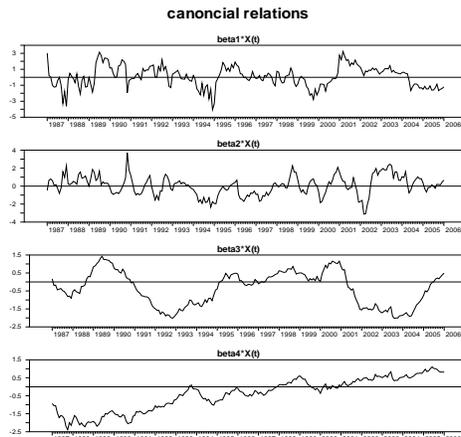


Figure 9: A plot of the canonical relations. It is seen that the first two which are the estimated cointegrating relations appear stationary an the last two are more nonstationary

The formal test for rank of the matrix Π can be found in Table 1.

r	$\hat{\lambda}$	Trace	c_{95}	p -val
0	0.23	95.71	53.94	0.00
1	0.11	35.09	35.07	0.05
2	0.02	7.45	20.16	0.86
3	0.00	1.84	9.14	0.80

Table 1: The trace test for rank shows that $r = 1$ or 2 are acceptable at a 5% level

The tests for rank $r = 0, \dots, 3$, indicate that $r = 1$ can be accepted at the 5% level, and at the 10% level we find $r = 2$, with a p -value of 0.861. The choice of $r = 2$ is supported by the finding of two roots close to one and will be maintained in the following. In the tests we assume a restricted constant so that $\mu = \alpha\rho'$, in order to avoid a linear trend in the data and maintain the possibility that interest rate spreads are stationary around a level.

The two estimated cointegrating relations are given in Table 2, see also Figure 9, where the first two panels show the estimated cointegrating relations. In order to

	β'_1	β'_2	α'_1	α'_2
i_{ff}	1	0	-0.206 [-4.124]	-0.070 [-2.422]
i_{6m}	-1.157 [-26.821]	1	0.840 [1.127]	-0.066 [-1.536]
i_{3y}	0.204 [3.968]	-1.987 [-17.760]	0.002 [0.024]	0.063 [1.310]
i_{10y}	0	1.004 [7.492]	0.068 [0.917]	0.030 [0.706]
1	0.00 [.429]	-0.00 [-1.232]		

Table 2: The parameters α and β estimated unrestrictedly. The t -ratios are given in brackets.

estimate the uncertainty of the coefficients we have imposed a zero restriction and a normalization on each. The natural hypothesis that the spreads are stationary cannot be satisfied because there are three spreads among four variables and we have only two stationary relations. Thus this simple cointegration analysis questions the usual expectation hypothesis that would predict the stationary of the spreads. What can we find instead?

A formal test that the cointegration relations are homogeneous, and therefore only depend on the spreads, can be tested as the hypothesis that $\beta = H\phi$ or $H'_\perp\beta = 0$, where

$$H'_\perp = (1, 1, 1, 1).$$

Considering the estimates of the second relation in Table 2, it is obvious that it describes the change of slope $(i_{6m} - i_{3y}) - (i_{3y} - i_{10y})$ which is almost like a curvature, and in fact the first relation is approximately $(i_{ff} - i_{6m}) - 0.2(i_{6m} - i_{3y})$ which also represents a change of slope. We can test that as the hypothesis that $\beta_1 = H_1\phi_1$ and $\beta_2 = H_2\phi_2$, where

$$H_1 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ -2 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We find from Table 3 the stationary relations $(i_{6m} - i_{3y}) - (i_{3y} - i_{10y})$, and $(i_{ff} - i_{6m}) - 0.082(i_{6m} - i_{3y})$, so it seems from the data that the term structure is described more by stationary changes of the slope and less by stationary slopes.

	β'_1	β'_2	α'_1	α'_2
i_{ff}	0	1	-0.052 [-1.714]	-0.172 [-3.588]
i_{6m}	1	-1.082 [-25.228]	-0.073 [-1.642]	0.100 [1.413]
i_{3y}	-2	0.082 [1.908]	0.059 [1.175]	0.027 [0.339]
i_{10y}	1	0	0.020 [0.449]	0.085 [1.193]
1	0.00 [-4.492]	-0.00 [9.592]		

Table 3: The restricted estimates of α and β . The test for accepting these is $-2\log LR(\beta \text{ restricted}) = 6.34$ corresponding to a p -value of 0.096 in the $\chi^2(3)$ distribution. The t -ratios are given in brackets.

6 Conclusion

We have discussed spurious regression and correlation, which has been well recognized for more than 80 years and have applied this to contrast two approaches. The regression or correlation based approach and the model based approach to statistical inference.

It is argued that it is a good idea to distinguish between the *empirical* and the *theoretical* correlation and regression coefficients. We need a limit theorem to relate the empirical value to the theoretical value, and this limit theorem may not hold for nonstationary variables.

We illustrate by examples that empirical correlation and regression coefficients may be spurious, in the sense that the conclusions drawn from them cannot be considered conclusions about the theoretical concepts. There are two reasons why regression and correlation may be spurious. One is that using \bar{X} , and the empirical variance as estimates of the mean and variance of X_t is that these need not be constant if the process is nonstationary.

The solution to the spurious correlation or regression problem in practice, is to model the data and check the model carefully before proceeding with the statistical analysis and this is where cointegration analysis comes in, as a possibility to model the nonstationary variation of the data. The vector autoregressive model is a natural framework for analysing some types of nonstationary processes and cointegration turns out to be a natural concept to describe long-run relations in economics and other fields.

By modelling the data carefully, taking into account the nonstationarity, we can actually learn something new about the economic relations underlying the variation of the data by posing sharp questions and analyse the statistical model.

7 Acknowledgement

The author gratefully acknowledges support from Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Founda-

tion.

8 References

1. Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22, 327–351.
2. Davidson, R., & MacKinnon, J. G. (2004). *Econometric Theory and Methods*. Oxford University Press.
3. Dennis, J., Johansen, S., & Juselius, K. (2005). *CATS for RATS: Manual to Cointegration Analysis of Time Series*, Estima, Illinois.
4. Dickey D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49, 1057–1072.
5. Engle R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55, 251–276.
6. Giese, J. V. (2008). Level, slope, curvature: characterizing the yield curve in a cointegrated VAR model, *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 2, 2008–28.
7. Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 121-130.
8. Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics* 2, 111–120.
9. Haavelmo, T. (1943). Statistical implications of a system of simultaneous equations. *Econometrica* 11, 1-12.
10. Hall, P., & Heyde, C. C. (1980). *Martingale Limit Theory and its Application*. Academic Press.
11. Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.
12. Johansen, S. (1996) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.
13. Johansen, S. (2006). Cointegration: a survey. In Mills T. C. and Patterson K. (eds) *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*. Palgrave Macmillan. 16, 17–34.

14. Juselius, K. (2006). *The Cointegrated VAR Model: Econometric Methodology and Macroeconomic Applications*. Oxford University Press.
15. Phillips, P. C. B. (1986). Understanding Spurious Regressions in Econometrics. *Journal of Econometrics* 33, 311-340.
16. Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
17. Sober, E. (2001). Venetian sea levels, British bread prices and the principle of the common cause. *British Journal for the Philosophy of Sciences* 52, 331–346.
18. Stock, J. H., & Watson, M. W. (2003). *Introduction to Econometrics*. Addison Wesley.
19. 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.