Applied Time Series Analysis — Part II

Robert M. Kunst
University of Vienna
and
Institute for Advanced Studies Vienna

November 6, 2011

1 Time series with a trend

1.1 A short history of trends in economics

Evidently, many economic variables show some sort of trending behavior, whether it be the gross domestic product or a stock market index. Typically, for these cases time-series modelling under the assumption of stationarity or forecasting by extrapolation using time-constant means would be highly implausible. Other variables, such as inflation and unemployment rates, show rising or falling trends over longer time spans, but extrapolating these trends may be of doubtful value for modelling.

Because the methodology of time-series analysis is tailored to the handling of stationary variables, it is an obvious suggestion to search for transformations of trending variables that yield time series with a stationary appearance. Often it is argued that prior to the publication of the book by Box& Jenkins economists were routinely fitting functions of time $\tau(t)$ to trending variables X_t and consequently subtracting them, in order to obtain stationary variables $\tilde{X}_t = X_t - \tau(t)$. Note, however, that for example the random-walk model of stock-market prices was developed by Bachelier as early as 1900. If X_t is assumed to be generated by a random walk, then $\tilde{X}_t = X_t - \tau(t)$ cannot be stationary for any function $\tau(t)$. Trend adjustment by subtracting a smooth function of time becomes inadequate. It is, however, fairly obvious that for a random walk

$$X_t = X_{t-1} + \varepsilon_t$$

with white noise ε_t the application of the *first-differences operator* $\Delta = 1 - B$

yields a stationary variable

$$\Delta X_t = X_t - X_{t-1} = \varepsilon_t,$$

in this simple case even white noise. Here, the operator $\Delta = 1 - B$ uses the notation $BX_t = X_{t-1}$ for the lag or *'backshift'* operator. First differences of logarithmic macroeconomic entities have always been considered to study growth rates. Anyway, there is some truth to the fairy tale of the invention of first differences by Box and Jenkins. The introduction of ARIMA modelling has led to a boom in the application of 'differencing', while non-linear functions of time $\tau(t)$ are hardly ever utilized nowadays. We recall that Box&Jenkins recommend comparing correlograms of the original time series X_t to those of the once differenced data ΔX_t and of the twice differenced data $\Delta\Delta X_t = \Delta^2 X_t$. The simplest correlogram indicates the best transformation level, which is then subjected to ARMA modelling.

In the notation of Box&Jenkins, a process is called ARIMA(p, d, q) if it yields a stationary ARMA(p, q) process after differencing it d times. In practice, only the cases $d = 0, 1, 2$ are used, where $d = 0$ indicates a stationary ARMA process. Thus, ARIMA$(p, 0, q)$=ARMA(p, q). The concept has even been generalized to negative and to non-integer d. Since the 1980s, the concept of ARIMA$(p, 1, q)$ processes has been superseded by I(1) processes, and ARIMA$(p, 0, q)$ are now more often named I(0). The exact definitions of I(0)/I(1) show some variation across sources. One intention of the more recent I(d) concept, however, is to generalize the ARIMA definition, such that it allows for the possibility of stationary differences without a finite low-order ARMA representation.

The surge in the application of first differences in economics did meet some resistance. The concept of a random walk or more generally of an ARIMA$(p, 1, q)$ model

$$\Phi(B)\,\Delta X_t = \Theta(B)\,\varepsilon_t$$

implies that a 'shock' at time t modifies the level of the variable X permanently, adopting the popular economic metaphorical interpretation of the errors ε_t as 'shocks'. By contrast, the 'trend-stationary' model

$$\Phi(B)\,(X_t - a - bt) = \Theta(B)\,\varepsilon_t,$$

implies that the effect of such a shock dies out gradually for $t + h$ with $h \to \infty$. In a pure MA model the shock is even completely forgotten after q observations. Many economists refuse to accept that a single one-time shock will affect a variable such as real output *at infinity*. Some authors even equated the validity of the ARIMA$(p, 1, q)$ model with the dominance

of supply shocks and the validity of the trend-adjusted ARMA model with dominance of demand shocks. Another argument against the application of differencing were the uncertain implications on *multivariate* modelling. By construction, differencing tends to destroy the data information on long-run constraints and equilibrium conditions. This problem was then formalized as 'cointegration'.

Whereas BOX&JENKINS recommended to inspect descriptive graphs, such as correlograms, in order to decide on ARIMA$(p, 1, q)$ versus ARIMA$(p, 0, q)$=ARMA(p, q), DICKEY&FULLER (1979) proposed a statistical test for this problem. The DICKEY-FULLER test (later ADF test) quickly developed into a major empirical tool in economics. It became a stylized fact that most trending economic variables are I(1) rather than trend stationary. Unfortunately, however, it also became obvious that many samples of economics data are much too short to admit a reliable decision on this issue. Doubts on the I(1) findings were also raised on grounds of the possibility of structural breaks and of generalizations of the autoregressive model that underlies the ADF test.

Research in the 1990s was dominated by multivariate tests (cointegration tests), by tests hypothesizing structural breaks, and by non- and semiparametric tests. After 2000, the potential of increasing test power by considering panel data created a new focus.

1.2 The test of Dickey and Fuller

In a simple AR(1) model

$$X_t = \phi X_{t-1} + \varepsilon_t, \tag{1}$$

(asymptotic) stationarity is equivalent to $\phi \in (-1, 1)$, while $\phi \geq 1$ implies non-stationary processes. The case $\phi = 1$ characterizes the random walk, i.e. an ARIMA$(0, 1, 0)$ process that can be transformed to stationarity by differencing. A test for the hypotheses

$$\begin{aligned} H_0 &: \phi = 1, \\ H_A &: \phi \in (-1, 1), \end{aligned}$$

appears to be simple. The t statistic

$$DF_0 = \frac{\hat{\phi} - 1}{\sigma\left(\hat{\phi}\right)}$$

may serve as a convenient test statistic. For $\hat{\phi}$ one may substitute any conventional estimator, such as OLS or ML. It is conceivable to define an operable

denominator as the square root of $\sigma_\varepsilon^2 / \sum X_{t-1}^2$. It has been known for a long time—long before the contributions of DICKEY&FULLER—that DF_0 under H_0 is *not asymptotically normally distributed*. This fact also invalidates the popular approximation by the t–distribution. Quantiles for the asymptotic distribution under H_0 were first tabulated in FULLER (1976). Later authors often called this distribution the 'Dickey-Fuller distribution' and represented it in the form

$$DF_0 \sim \frac{\int B(\omega)\, dB(\omega)}{\{\int B(\omega)^2\, d\omega\}^{1/2}}. \tag{2}$$

The symbol $B(\omega)$ denotes the so called *Brownian motion*, in the following *Bm*, a counterpart to the *random walk* in continuous time. Formally, this real process on the interval $[0, 1]$ is defined by the properties

$$\begin{aligned} B(0) &= 0, \\ B(t+h) - B(t) &\sim N(0, h), \\ \mathrm{E}\{B(t+h) - B(t)\} B(t) &= 0. \end{aligned} \tag{3}$$

This definition implies the important property $B(t) \sim N(0, t)$ and also $B(1) \sim N(0, 1)$. Processes that differ by a scaling factor only, such as $\lambda B(t) \sim N(0, \lambda^2 t)$, are often also called Brownian motions. Then, the process according to the above definition (3) will be called *standard Brownian motion*. Ratios like (2) are invariant to scales.

A simple random experiment may serve to convey a first impression of the *Bm*. A scaled random walk defined as

$$X_{t,T} = \frac{1}{\sqrt{T}} \sum_{j=1}^{t} \varepsilon_j, \quad t = 0, \ldots, T,$$

with $N(0, 1)$ random draws ε_j is generated and the trajectory is plotted against $0, 1/T, 2/T, \ldots, 1$ instead of $0, \ldots, T$. Figure 1 shows 20 such trajectories. Letting T increase to ∞ decreases the oscillation to a microscopic dimension and eventually yields *Bm*.

The integrals over $[0, 1]$ such as those appearing in (2) often do not formally exist, as the trajectories of time derivatives of *Bm* oscillate too strongly. They can be seen as limits in the sense of

$$T^{-1} \sum_{t=1}^{T} (X_t - X_{t-1}) X_{t-1} \to \int_0^1 B(\omega)\, dB(\omega),$$

$$T^{-2} \sum_{t=1}^{T} X_t^2 \to \int_0^1 B^2(\omega)\, d\omega.$$

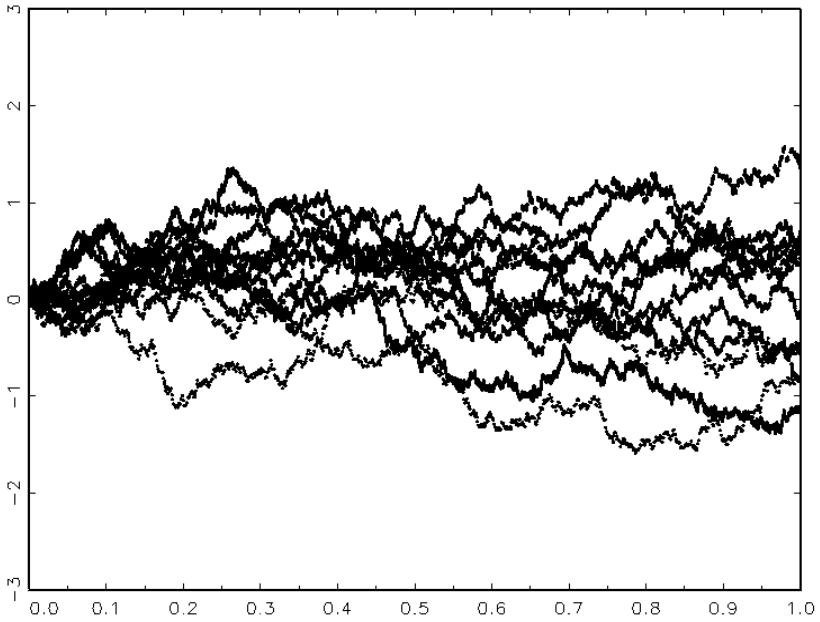

Figure 1: 20 trajectories of the random walk $X_t = X_{t-1} + \varepsilon_t$ with $X_0 = 0$ and $\varepsilon_t \sim iid\, N(0,1)$, for $t = 0, \ldots, 10^5$.

These formulae are important insofar as they can be used to approximate the DF_0 distribution from easily simulated random walks X_t to an arbitrary precision.

Usually, the DF_0 statistic is defined as the t–statistic of φ in the regression

$$\Delta X_t = \varphi X_{t-1} + \varepsilon_t. \tag{4}$$

Note that $\varphi = 0$ if and only if $\phi = 1$, and hence the representations (1) and (4) are equivalent. For this reason, Dickey&Fuller also considered the weighted estimate

$$T\left(\hat{\phi} - 1\right) = T\hat{\varphi}$$

as an alternative test statistic. This concept is less often used nowadays.

1.3 Constants and trends in the Dickey-Fuller test

The pure AR(1) model (1) or (4) is empirically not very interesting. Of more relevance is the model with intercept

$$X_t = \mu + \phi X_{t-1} + \varepsilon_t. \tag{5}$$

Under the hypothesis $\phi = 1$, this model yields a so called *random walk with drift*. Iterative insertion delivers the representation

$$X_t = \mu t + \sum_{s=1}^{t} \varepsilon_s + X_0.$$

Under the alternative $\phi \in (-1, 1)$, the model defines an (at least asymptotically) stationary process with expectation

$$\mathrm{E}X_t = \frac{\mu}{1-\phi}.$$

An obvious suggestion for a test statistic is again the t–statistic

$$DF_\mu = \frac{\hat{\phi} - 1}{\sigma\left(\hat{\phi}\right)}$$

in (5). An important feature, however, is that DF_μ under $H_0 : \phi = 1$ has a different (asymptotic) distribution than DF_0. Today, it is often represented as

$$DF_\mu \sim \frac{\int \bar{B}(\omega)\, d\bar{B}(\omega)}{\{\int \bar{B}(\omega)^2\, d\omega\}^{1/2}},$$

where $\bar{B}(\omega)$ is a centered *Brownian motion* that is obtained by calculating for any given trajectory its sample mean over the interval $(0, 1)$ and subtracting it. Then, the property $B(0) = 0$ changes to $\bar{B}(0) = -\int B(\omega)\, d\omega$. The validity of this DF_μ distribution of the DF_μ statistic requires that $\mu = 0$ holds for the true intercept in (5). Also this distribution was tabulated by Fuller and by Dickey&Fuller.

Figure 2 conveys an impression of the centered *Bm*. The same random draws were used as in Figure 1. A distinctive feature of the trajectories is that, unlike those of the original *Bm*, they do not start from 0 and condense around 0 approximately in the middle of the interval.

Today, the DF_μ test is applied in situations where the researcher has doubts on the stationarity assumption but no long-run trend is visible in the sample or is even ruled out by theoretical considerations. Typical examples are rates of interest or of inflation. If genuine trends are an issue, then model (5) may encompass the null hypothesis of a *random walk with drift* but it does not allow for the logical alternative of a process that becomes stationary after linear de-trending. For this important case we consider the model with a linear trend

$$X_t = \mu + \tau t + \phi X_{t-1} + \varepsilon_t. \tag{6}$$

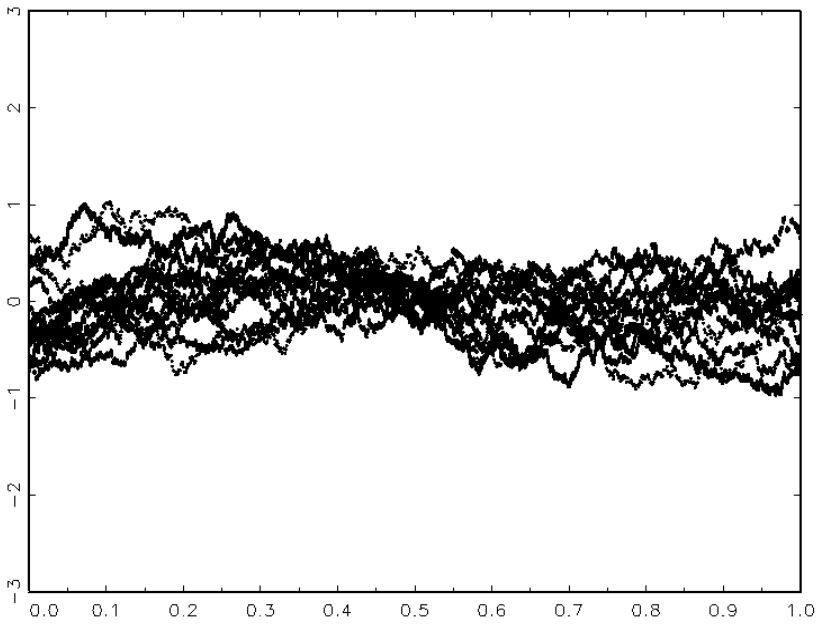

Figure 2: 20 trajectories of the centered random walk $\bar{X}_t = X_t - \frac{1}{T}\sum_{t=0}^{T} X_t$ with $X_t = X_{t-1} + \varepsilon_t$, $X_0 = 0$, and $\varepsilon_t \sim iid\, N(0,1)$, for $t = 0, \ldots, 10^5$.

Two different null hypotheses can be studied. The hypothesis

$$H_{0F} : \tau = 0, \phi = 1$$

implies an F–type test statistic, which again will not have an F distribution under H_{0F}. This test is not so often applied in empirical research. Rather, the t–type statistic on $\phi = 1$ in (6) defines a test for

$$H_{0T} : \phi = 1.$$

This test is called DF_τ test, and it is the most customary *unit-root test*. The asymptotic distribution is again available in the form of tables of significance points. The expression 'unit-root test' refers to the roots (zeros) of the characteristic polynomial of the autoregressive model $\phi(z) = 1 - \phi z$. In the simple AR(1) model, the occurrence of a unit root is equivalent to $\phi = 1$.

A disadvantage of the DF_τ test vis-a-vis the F–test for H_{0F} is that its null hypothesis H_{0T} admits the model

$$X_t = \mu + \tau t + X_{t-1} + \varepsilon_t,$$

a random walk with superimposed quadratic trend, which can be easily demonstrated by iterated substitution (**Exercise**). This model is implausible for most macroeconomic variables and delivers inferior forecasts at longer

horizons. For this reason, it should only be seen as an auxiliary model for the test construction and not as a potential data-generating process if the null hypothesis H_{0T} is accepted. In particular, the estimated values $\hat{\mu}$ and $\hat{\tau}$ should not be used in subsequent modelling.

The distribution of the DF_τ statistic relies on integrals over functions of de-trended Bm that look somewhat like the trajectories in Figure 3. The fitting of linear trends to Bm implies an oscillating shape that is scattered over the entire interval [0,1].

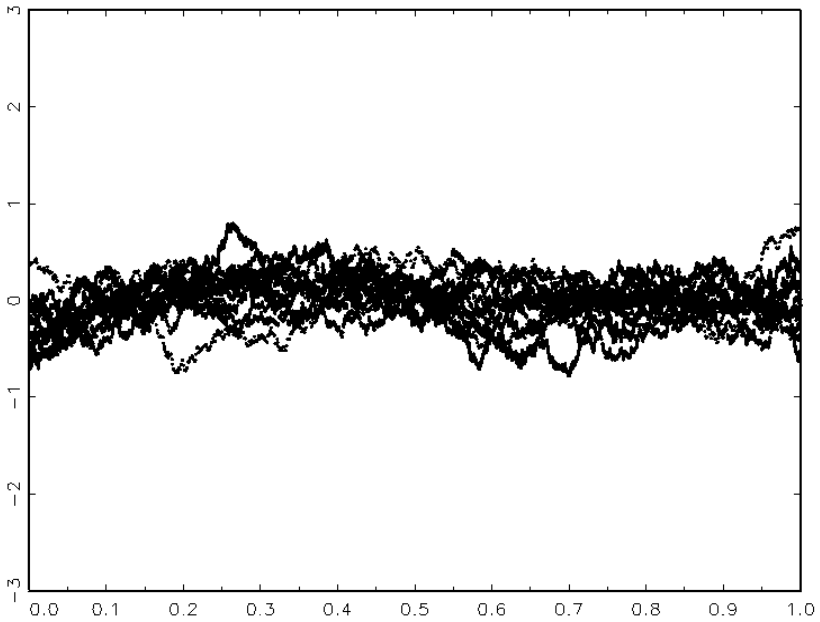

Figure 3: 20 trajectories of the linearly de-trended random walk $\tilde{X}_t$. $\tilde{X}_t$ results as a residual from the regression of X_t on $\mu+\tau t$, where $X_t = X_{t-1}+\varepsilon_t$, $X_0 = 0$, and $\varepsilon_t \sim iid\, N(0,1)$, for $t = 0, \ldots, 10^5$.

Figure 4 shows box-plots of the distributions of the statistics DF_μ and DF_τ under the null hypothesis of a random walk, simulated from 100 observations and 10,000 replications. Obviously, these distributions cluster around -2 and thus clearly differ from normal or t–distributions. Evidence on asymmetry in these distributions, however, is weak, and kurtosis values of 3.32 and 3.48 transgress the normal kurtosis of 3 only slightly. Positive values are rare. The latter observation implies that a random walk, if estimated as a general AR(1) process, usually delivers estimated coefficients that are a bit smaller than 1, and almost never explosive estimates above 1. This is the most extreme manifestation of the so called *Hurwicz bias* that affects

all time-series models. In finite samples, all usual estimators for time-series coefficients are biased, where generally the absolute value of the coefficients is systematically under-estimated.

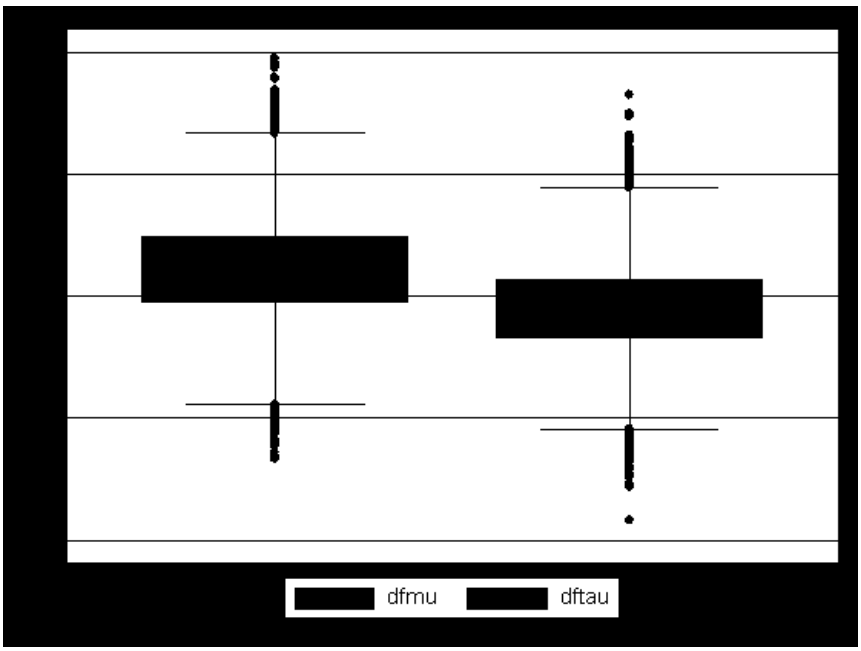

Figure 4: Box plots of the empirical distributions for the statistics DF_μ and DF_τ. Generating model is a random walk with 100 observations. 10,000 replications.

1.4 General autoregressive models in the Dickey-Fuller test

If instead of an AR(1) a general AR(p) model is considered as the data-generating process, then the simple statistics DF_0, DF_μ, DF_τ will not be appropriate for the test decisions, the asymptotic distributions will fail to hold. Two solutions were proposed for this problem. The older and simpler one is due to Dickey&Fuller and is called the ADF test (*augmented Dickey-Fuller*). The younger semi-parametric variant has first been suggested by Phillips& Perron.

The AR(p) process

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t \tag{7}$$

is (asymptotically) stationary if all roots of the polynomial

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p$$

exceed one in modulus. If exactly one root equals 1, then we have

$$\phi(z) = (1 - z)\,\phi_0(z),$$

where all roots of $\phi_0(z)$ are larger than 1 in modulus. It follows that ΔX_t is stationary but not X_t. This property defines a *first-order integrated process* or I(1) process.

A logical consequence is the following generalization of the hypotheses of a unit-root test:

$$H_0 : \phi(1) = 0,$$
$$H_A : \phi(1) \neq 0,$$

where several restrictions on the maintained hypothesis $H_0 \cup H_A$ are required. Multiple roots of 1 must be ruled out as well as polynomial zeros with a modulus of 1 though not equal 1, usually even all 'explosive' roots with modulus less than 1. It is customary to view all variants of the DF tests as *one-sided* tests, thus restricted to stationarity under the alternative.

DICKEY&FULLER proposed the regression model

$$\Delta X_t = \varphi X_{t-1} + \psi_1 \Delta X_{t-1} + \ldots + \psi_{p-1} \Delta X_{t-p+1} + \varepsilon_t \tag{8}$$

as a basis for their test. Under the unit-root null, the t–statistic for $\varphi = 0$ has asymptotically the same distribution as the statistic DF_0 in the AR(1) model. Proving this proposition would require showing the following points:

1. Every AR(p) model (7) can be re-written in the form (8). There is a one-one relation between the parameters $\phi_1, \ldots, \phi_p$ and $\psi_1, \ldots, \psi_{p-1}; \varphi$.

2. $\varphi = 0$ if and only if $\phi(1) = 0$.

3. The regression terms $\Delta X_{t-1}, \ldots, \Delta X_{t-p+1}$ have asymptotically no influence on the distribution of the test statistics DF_0, DF_μ, DF_τ, provided ε_t is white noise.

The first point is a simple algebraic exercise and is recommended as an **exercise**. It is simpler to first take the form (8) to the one of equation (7) and to express $\phi_1, \ldots, \phi_p$ as functions of $\psi_1, \ldots, \psi_{p-1}; \varphi$. The resulting linear equation system is easily inverted.

For the second point, it is worth noting that

$$\phi(z) = 1 - z - \varphi z - \psi_1 (1 - z) z - \ldots - \psi_{p-1} (1 - z) z^{p-1},$$

given that point 1 holds. The properties of the polynomial must be independent of the representation form. The argument uses the multiple lag operator $B^h X_t = X_{t-h}$ that corresponds to z^h, and also $\Delta = 1 - B$, which corresponds to $1 - z$. This yields $\phi(1) = -\varphi$, which proves the second point.

We do not prove the third point here. It remains valid if white noise is replaced by different but similar requirements, but it ceases to hold in the presence of an ARMA process or if the lag order p was under-specified. In these cases, ε_t will be autocorrelated. It was shown, however, that application of the ADF test to ARMA generating processes yields a valid test in an asymptotic setting, contingent on a successive increase in the lag order p that is proportional to the increasing sample size.

In analogy to the original DF test, even the ADF test is applied to models with constant and trend. Again, the variant with trend ADF_τ is the most adequate one in most situations.

An important practical problem is the determination of the lag order for the ADF test. Among others, the following suggestions can be found in the literature:

1. determine p as a function of the sample size;
2. minimize information criteria (often AIC) for AR models in X_t (not in ΔX_t);
3. increase p until the Q statistic for residual autocorrelation becomes insignificant.

Simulations tend to discourage the third suggestion. In doubtful cases, it may be recommended to choose a higher lag order rather than a too small one.

Example. Our role-model example considers data on the Austrian industrial production that is depicted graphically in the first part of these lecture notes. After seasonal adjustment using the moving-average filter $0.25 (1 + B + B^2 + B^3)$, a trend is clearly recognizable in the variable. Unit roots with a value different from 1 are not likely, as these are usually due to seasonal cycles, which have been filtered out. AIC yields a minimum for the autoregressive model with constant and trend at the lag order $p = 6$. For this reason, we apply an ADF test with **five** lags in differences. The test statistic has the value -2.334. A comparison to the tabulated significance

points shows that the null hypothesis cannot be rejected. The variable appears to be $I(1)$. The software EViews provides the critical value at the 10% level as -3.143, Stata has the value -3.141. The tabulated value would be -3.15. EViews and Stata use the technique of *response surfaces* to obtain a closer fit to the actual sample size and to other sample-specific features. This technique, however, does not always imply a convincing improvement of accuracy. □

1.5 The test of Phillips and Perron

The test by Phillips&Perron builds on the following two facts. First, it is true that the asymptotic distribution of the statistics DF_0, DF_μ, DF_τ in the model

$$\Delta X_t = d + \varphi X_{t-1} + u_t$$

(with $d = 0$, $d = \mu$, $d = \mu + \tau t$) is affected by the possible autocorrelation in u_t and does not coincide with the original DF–distributions. Second, however, the discrepancy between the cases for white-noise and for autocorrelated u can be expressed by relatively simple correction factors. These correction factors depend on the ACF of u_t. A suggestion would be to estimate these terms from the correlogram of the estimated residuals $\hat{u}_t$ and to modify the test statistic accordingly. This yields a test statistic that asymptotically follows the distribution that was tabulated by Dickey and Fuller. An alleged advantage of this PP–test is that it remains valid if X_t has been generated by an ARIMA$(p, 1, q)$ model rather than by a pure AR model. Moreover, its calculation does not require the determination of a lag order.

Later researchers, however, found that the PP–test does not necessarily dominate the ADF test for generating general ARIMA models with regard to test power. Furthermore, the test requires in its construction a maximal order (window width), up to which autocorrelations are to be included in the correction factor. The test is often used as a robustness check on the results of the ADF test.

1.6 Tests for stationarity instead of unit roots

Some researchers object to the basic idea that the null hypothesis of the DF tests and of its variants is a non-stationary model and its alternative is a stationary model. By contrast, it may appear that stationarity is a well-defined property that should rather serve as a null hypothesis. This argument is of course not convincing, as the null hypothesis of the DF tests is not general non-stationarity but a special model class such as $I(1)$ or ARIMA$(*, 1, *)$.

If one really wishes to make stationary models (including 'trend-stationary' processes that are representable as sums of trend functions and a stationary part) a null hypothesis, then there are two strategies for constructing such a test. The first one builds on the observation that first differences of stable ARMA processes have a unit root in the MA polynomial of their ARMA representation. For example, first differences of white noise

$$\Delta\varepsilon_t = \varepsilon_t - \varepsilon_{t-1} = (1 - B)\,\varepsilon_t$$

have such an MA unit root. If, on the contrary, the observed variable is $I(1)$, then its first difference is stationary but does not possess such an MA unit root. This fact suggests to estimate an MA model for ΔX_t and to test whether $\Theta(1) = 0$ or $\Theta(1) \neq 0$.

An alternative strategy uses the concept of 'structural' decompositions into *unobserved components, UC*. This paradigm sees every observed X_t as the sum of a random walk and a stationary remainder. If the random-walk portion is not significant, one may classify X_t as stationary. This approach often utilizes non-parametric modelling concepts. A typical UC model has the form

$$\begin{aligned} X_t &= \mu + \tau_t + \gamma t + u_t, \\ \tau_t &= \tau_{t-1} + \eta_t, \end{aligned}$$

where u_t is a stationary remainder process, whereas η_t is white noise with variance σ_η^2. The null hypothesis $H_0 : \sigma_\eta^2 = 0$ corresponds to the trend-stationary case $\tau_t = \tau_{t-1} = \tau$, the alternative $H_A : \sigma_\eta^2 > 0$ to an $I(1)$ process for X_t.

The most popular test in this class is the one proposed by Kwiatkowski& Phillips& Schmidt&Shin (KPSS test). A practical difficulty is caused by findings that imply a contradiction between the KPSS and the ADF test. For example, we may observe that both tests reject, which implies an $I(0)$ variable according to ADF but an $I(1)$ variable according to KPSS. It is more difficult to reconcile these two pieces of information than for two tests with identical null hypotheses, as the interpretation of significance levels does not coincide. Unconventional solutions such as *joint confirmation analysis* are not generally accepted in the literature, as they transgress the familiar framework of classical hypotheses tests.

Example. One of the suggestions found in the relevant literature for the determination of the lag order in the KPSS test is $(5/7)\sqrt{T}$. For the case of industrial production, this value is slightly larger than 9. The KPSS statistic is given by Stata as 0.277. This value exceeds the tabulated 1% significance point of 0.216, such that the test rejects trend stationarity. There is also

a variant of the KPSS statistic that corresponds to the DF_μ test and has stationarity rather than trend stationarity as its null hypothesis. This variant rejects even more strongly: its value is 1.74 and the critical 1% value is 0.74. For first differences, i.e. the growth rate, we obtain a value of 0.169, obviously much less than the 10% point 0.347. In this case, the KPSS tests confirm the impression from the DF tests that industrial production corresponds to an I(1) model.

2 Cointegration

Although various visual devices such as the sample ACF and also hypothesis tests are available to decide on whether or not to difference the data, many researchers will hesitate before considering differences of single variables if the aim is multivariate analysis, the joint modelling of several variables. While a relation such as

$$y_t = \gamma x_t + u_t \tag{9}$$

can be interpreted as the effect of a variable x on a variable y, it appears that the corresponding relation

$$\Delta y_t = \gamma \Delta x_t + u_t \tag{10}$$

expresses a marginal reaction to small changes in x. Indeed, the equation (9) contains information on the joint longer-run dynamics of x and y that is absent from equation (10). Time-series analysis has established an adequate theory for this feature in the 1980s, following some earlier preliminary contributions. The path-breaking work in this regard is due to Engle&Granger (1987).

A basic principle is that already stationary variables ($I(0)$) are not to be differenced in the multivariate model. Integrated variables ($I(1)$) can be *cointegrated* or not. If they are not cointegrated, they are differenced and enter the multivariate models in their differenced form. If they are cointegrated, efficient multivariate modelling requires the specification of *error-correction* models. To denote the feature of cointegration of two or more $I(1)$ variables, the notation $CI(1,1)$ is used by some authors.

A prototypical example for cointegration is a set of two time series: aggregate consumption c and disposable household income yd. For Austria, these time series are depicted in Figure 5. Both variables show clearly recognizable trends, and at least for longer samples the unit-root hypothesis is confirmed. The difference of the logarithmic variables, however, appears to be stationary and is supposed to be so, as many economists assume that the saving ratio

of households is constant in the longer run. Empirically, this cointegration among the two variables is not always confirmed on statistical grounds.

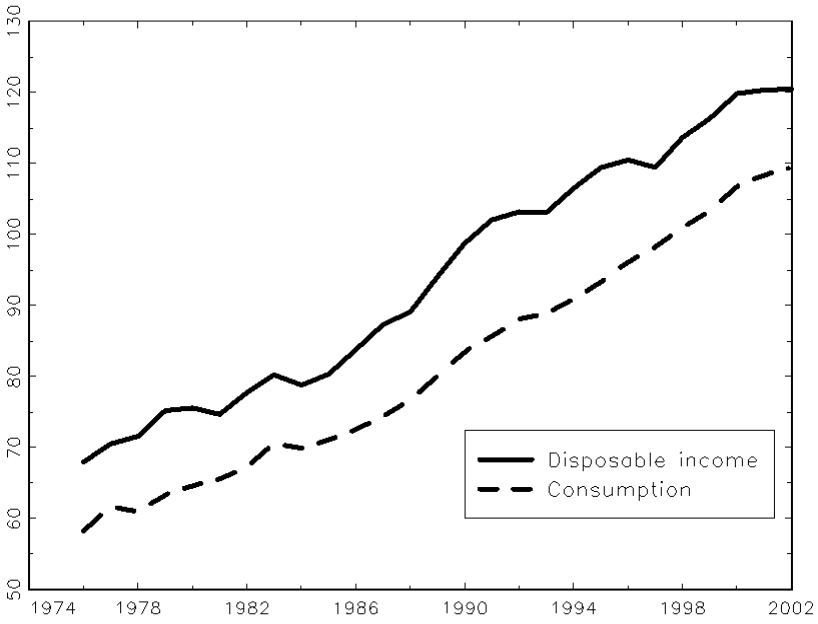

Figure 5: Aggregate private consumption and disposable household income for Austria, at constant prices, in logarithms.

In this example, testing the hypothesis of stationarity in $c - yd$ is easy, for example by an ADF test on this variable. In many other cases, however, the 'cointegrating relation' β in

$$\beta' X_t = \beta_1 X_{1t} + \ldots + \beta_n X_{nt}$$

is not specified by theory but it must be estimated.

The following basic definition of cointegration follows Engle&Granger with slight modifications:

Definition. If two processes X_t, Y_t are each ARIMA$(p, 1, q)$ and if there exists a linear combination $Z_t = \beta_1 X_t + \beta_2 Y_t$ with $\beta_1 \neq 0$ and $\beta_2 \neq 0$ such that Z_t is stationary, then X_t and Y_t are called *cointegrated*, and $(\beta_1, \beta_2)'$ is called the *cointegrating vector*.

It is convenient to restrict focus to ARIMA$(p, 1, q)$ processes. Alternatively, one may consider the concept of general $I(1)$ processes. This requires the general definition:

Definition. A process X_t is called *k-th order integrated*, in symbols $I(k)$, if $\Delta^k X_t$ is stationary but not $\Delta^{k-1} X_t$.

Even this definition can be restricted to ARIMA(p, k, q) processes. The treatment of deterministic components (constants, trends, seasonal dummies) is not unanimous in the literature. The original definition of Engle&Granger excludes all deterministic parts, including constants. Most authors would accept $X_t = \mu + \varepsilon_t$ as $I(0)$, such that also $X_{t-1} = X_t + \mu + \varepsilon_t$, a process with superimposed linear trend becomes $I(1)$. By contrast, it appears to be at odds with the original idea to call X_t and Y_t cointegrated if $Z_t = \beta_1 X_t + \beta_2 Y_t$ is a stationary process with added linear trend. In this case, it makes sense to require that this linear trend disappears in Z_t.

Examples:

1. For any random walk with drift $X_t = a + X_{t-1} + \varepsilon_t$, $Y_t = bX_t + \eta_t$ with a white noise η_t that is independent of ε_t is cointegrated with X_t, as

   $$Z_t = bX_t - Y_t = -\eta_t$$

   is stationary and $(b, -1)$ is a cointegrating vector.

2. $X_t = \varepsilon_t - \varepsilon_{t-1}$ is stationary but not $I(0)$, as

   $$\Delta^{-1} X_t = \sum_{s=1}^{t} X_t = \varepsilon_t - \varepsilon_0$$

   is also stationary and thus formally X does not satisfy the definition. Because $\Delta^{-2} X_t$ is a non-stationary random walk, such processes are classified as $I(-1)$. Note the operator Δ^{-1} that is the inverse operator of first differences Δ. Δ^{-1} corresponds to an accumulation from a starting value to t. It is easily seen that $\Delta\Delta^{-1} X_t = X_t$.

3. Assume X_t and Y_t are generated via the bivariate AR model

   $$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}.$$

   Then, both components are random walks and thus $I(1)$. They are not even cointegrated if ε_t and η_t are correlated with each other.

4. Assume X_t and Y_t are generated from the bivariate AR model

   $$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{bmatrix} 0.9 & 0.1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}.$$

In this case,

$$X_t - Y_t = 0.9X_{t-1} - 0.9Y_{t-1} + \varepsilon_t - \eta_t$$

implies that, for $Z_t = X_t - Y_t$,

$$Z_t = 0.9Z_{t-1} + \varepsilon_t - \eta_t.$$

Hence, Z_t is stationary, and X_t and Y_t are cointegrated with vector $(1, -1)'$. Y_t is a random walk and therefore $I(1)$. X_t is the sum of a random walk and a stationary process Z_t and thus also $I(1)$, the conditions of the definition are fulfilled. It will be seen that the properties of the coefficient matrix are crucial.

For the joint analysis of more than two variables, the above definition does not suffice.

Definition. Given an n–dimensional vector process $X_t = (X_{1t}, \ldots, X_{nt})'$, suppose that $\Delta X_t = (\Delta X_{1t}, \ldots, \Delta X_{nt})'$ is stationary, whereas at least one component X_{jt} is $I(1)$, and that a linear combination with at least one $\beta_k \neq 0$

$$Z_t = \beta_1 X_{1t} + \ldots + \beta_n X_{nt} = (\beta_1, \ldots, \beta_n)' X_t$$

is stationary (or $I(0)$). Then, X_t is called *cointegrated*, in symbols $CI(1, 1)$, and the vector β is a *cointegrating vector*.

Examples:

1. Suppose X_{1t} is a random walk and X_{2t} is a white noise independent from X_1. Then, $(X_{1t}, X_{2t})'$ is cointegrated in the sense of the definition with vector $(0, 1)'$. This rather trivial case is called *self cointegration* and is excluded by some authors. It is inconvenient, however, to demand that *all* components are $I(1)$, as this requirement would exclude too many cases of empirical interest.

2. If X_{1t} is a random walk, and X_{2t} and X_{3t} are defined by adding mutually independent white noises to X_{1t}, then both $X_{2t} - X_{1t}$ and $X_{3t} - X_{1t}$ are stationary. Both $(1, -1, 0)'$ and $(1, 0, -1)'$ are cointegrating vectors. Even all multiples and all linear combinations of these vectors are cointegrating vectors. Cointegrating vectors are not uniquely defined.

3. An economic example from the literature is a collection of interest rates with different terms to maturity. While the individual interest rates are assumed to be $I(1)$, their differences are interpreted as time-constant risk or term premia, and they are stationary. Both assumptions are not always guaranteed in empirical data. Some authors find time-variant, non-stationary risk premia, others doubt that interest rates are $I(1)$.

4. Another example is provided by the three variables GDP, private consumption, and aggregated investment, all in logarithms. These macroeconomic variables are individually $I(1)$, their differences are logarithms of the consumption and investment share in output and thus potentially stationary. Figures 6 and 7 give some graphical evidence for this example.

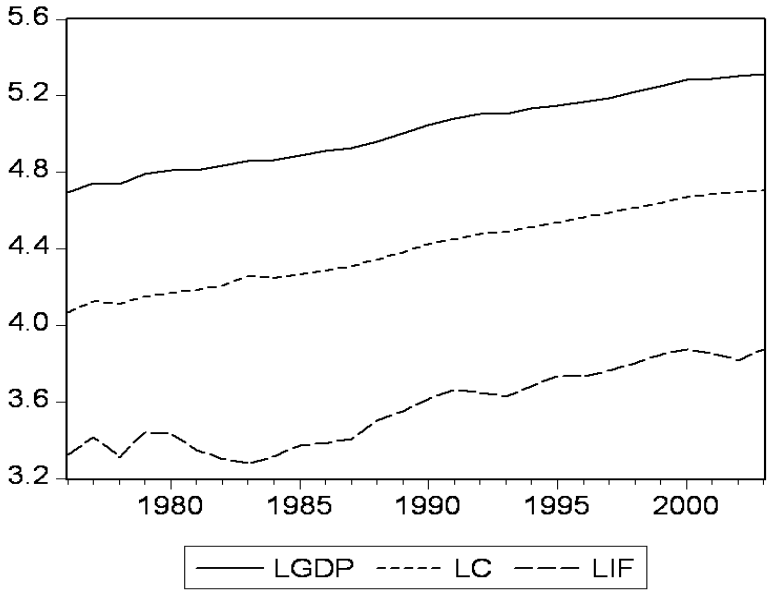

Figure 6: Austrian GDP (solid), private consumption (short dashes), and gross fixed investment (long dashes) series. Logarithmic transformations.

2.1 The cointegrating regression

The historically oldest technique for the estimation of cointegrating relations was proposed by Engle&Granger and relies on the so-called *cointegrating regression*. The idea is to transform the assumed cointegrating relation

$$Z_t = \beta_1 X_{1t} + \ldots + \beta_n X_{nt}$$

after division by β_1—if $\beta_1 = 0$, one chooses a different $\beta_j \neq 0$ and proceeds in an analogous way—into the form

$$\bar{Z}_t = X_{1t} + \bar{\beta}_2 X_{2t} + \ldots \bar{\beta}_n X_{nt}.$$

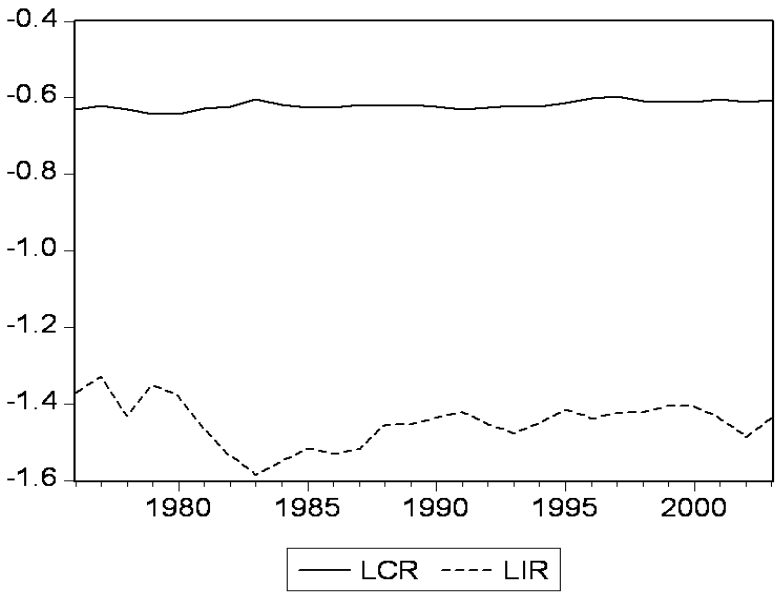

Figure 7: Logarithmic consumption-output ratio (solid) and investment-output ratio (dashes) for Austria.

If Z_t is stationary, then also $\bar{Z}_t = (1/\beta_1)\, Z_t$ will be stationary. Re-arranging yields

$$X_{1t} = -\bar{\beta}_2 X_{2t} - \ldots - \bar{\beta}_n X_{nt} + \bar{Z}_t,$$

which looks like a regression. The residuals are stationary but typically quite different from pure white noise and autocorrelated. However, if all X_{jt} are $I(1)$ variables, it can be shown that nonetheless all OLS estimates of coefficients $\hat{\beta}_j$ converge to the true values, and even 'faster' than in a usual regression with stationary regressors and white-noise errors. Because of this fast convergence, all estimates can be used like 'true' values in further analysis. Note, however, that the standard errors from a naive OLS regression printout and all t–values are invalid. Very large t–values are typical in cointegrating regressions.

For an example, we consider a simple system that consists of consumption and income

$$C_t = \gamma Y_t + \zeta_t,$$
$$Y_t = \delta + Y_{t-1} + \varepsilon_t.$$

Here, income follows a random walk with drift, the process ζ_t is an arbitrary

stationary ARMA–process. The variables are cointegrated, the marginal propensity to consume γ is estimated from the first equation by OLS.

A slightly more sophisticated system serves as a role-model example for the so-called error-correction models:

$$\Delta C_t = \mu_1 + \alpha\,(C_{t-1} - \beta Y_{t-1}) + \gamma \Delta Y_{t-1} + \varepsilon_t, \tag{11}$$

$$\Delta Y_t = \mu_2 + \delta \Delta Y_{t-1} + \eta_t. \tag{12}$$

Here, we may assume that ε_t and η_t are white noise. The negative coefficient α dampens household demand if it was too high in the previous time period as compared with the 'equilibrium relation' $C = \beta Y$, and accelerates it if it had been low before. One can show that such error-correction models occur if and only if C and Y are cointegrated with the vector $(1, -\beta)'$. It was just this proof of equivalence that was central in the article of Engle&Granger. Also this model an be estimated empirically using the technique of cointegrating regression. First, one estimates the regression

$$C_t = \beta_0 + \beta Y_t + \zeta_t$$

with stationary errors ζ_t. Then, $C_{t-1} - \hat{\beta} Y_{t-1}$ is inserted into (11), as if it were the true regressor $C_{t-1} - \beta Y_{t-1}$. This allows to estimate both equations (11) and (12) by OLS, to retrieve the remaining coefficients. In these second-stage regressions, the usual t–values are asymptotically normally distributed, and also some other regression statistics are valid.

According to Engle&Granger, this technique is 'efficient', but it causes strong bias effects in smaller samples. Nowadays, the method, which is usually labelled EG-2 (Engle&Granger *two-step*) is used only rarely in academic work or as a simple benchmark method. Its simplicity, however, guarantees its popularity with many non-academic practitioners. Its advantage is that it allows to treat equations such as (11) in an isolated fashion, without modelling the complete system, i.e. without (12). Practitioners often even consider variants such as

$$\Delta C_t = \mu_1 + \alpha\,(C_{t-1} - \beta Y_{t-1}) + \gamma \Delta Y_t + \varepsilon_t,$$

with simultaneous regressor ΔY_t, which is easily interpreted in economic terms but imposes severe econometric difficulties.

A typical feature of the EG-2 method is that the assumption of existing cointegration is tested in the very same framework of the cointegrating regression (*test for cointegration*). Historically, an early suggestion was to consider the Durbin-Watson (DW) statistic from the regression

$$C_t = \beta_0 + \beta Y_t + \zeta_t.$$

The errors are stationary but not uncorrelated, so it was suggested to diagnose cointegration if the DW–statistic is sufficiently different from zero(!), i.e. the boundary case formally corresponding to a random-walk errors model. Note that this would mean rejecting the *null hypothesis of non-cointegration*. Today, this CRDW test (*cointegrating-regression Durbin-Watson*) is not applied any more due to its unsatisfactory properties. However, the principle of having non-cointegration as the null and cointegration as the alternative is retained in all customary tests.

The next suggestion consisted in checking cointegration by way of ADF tests on the residuals $\hat{\zeta}_t$. In the presence of cointegration, the true ζ_t should be stationary. The statistical properties of residuals $\hat{\zeta}_t$, however, differ from those of the true errors ζ_t, and hence all usual significance points are *invalid*. Using the significance points that are automatically provided by computer software in this *residual ADF test* is a common mistake in empirical applications. Phillips&Ouliaris provide adequate quantiles for this test in the form of tables.

In summary, the procedure for cointegrating modelling consists of the following steps:

1. The integration order of all component variables X_j is determined by ADF or other unit-root tests. Here, standard ADF tables and automatically generated significance points are valid. If the ADF test rejects, then the variable is possibly already stationary ($I(0)$) and cannot play a role in a cointegrating regression. If none of the variables or only one of them is $I(1)$, then no (non-trivial) cointegration can occur.

2. Variables in the acceptance region of unit-root tests are often tested once more for the possibility of $I(2)$. This is done by applying the unit-root tests on first differences. Now, unit roots should be rejected, otherwise the variable may be $I(2)$ or even $I(k)$ with $k > 2$. While $I(2)$ variables may cointegrate with different $I(2)$—and even $I(1)$—variables, the correct handling of such situations imposes additional problems. Genuine $I(2)$ variables are rare in economics, often statistical evidence on $I(2)$ indicates structural breaks etc.

3. The cointegrating regression is conducted. For the regressand, often a variable is chosen, which can be seen as dependent on different X_j on theoretical grounds. In principle, the selection of the regressand should not affect the results, as long as this X_j has a non-zero coefficient in the cointegrating vector.

4. An ADF test is applied to the residuals of the cointegrating regression. Here, standard tables are invalid. If the test rejects, there is evidence in

favor of cointegration. If the test fails to reject, then either the variables do not cointegrate or—if more than two $I(1)$ variables are considered in the model—the variable chosen as the regressand does not occur in the cointegrating vector. In particular to check on the latter possibility, one may repeat the regression for different regressands.

5. If the procedure has confirmed the presence of cointegration, the regression coefficients determine an estimate for the cointegrating vector. The residuals of the cointegrating regression can be used as error-correction variables in other equations.

Example. We investigate annual data for consumption and income in logarithmic form, as shown in Figure 1.

1. The variable c is subjected to another ADF test. An AIC search yields a lag order of $p = 3$, thus two differences of Δc_t are to be used as additional regressors. The corresponding DF_τ statistic has the value -2.16 and is insignificant. For income, $p = 5$ and four lags are recommended. Again, the DF_τ statistic of -2.67 is insignificant. The summary impression confirms the $I(1)$ hypothesis for both variables.

2. DF_μ tests on Δc and Δyd deliver the significant values of -7.26 and -4.75. This excludes the possibility of $I(2)$ that was not plausible *a priori*. Here, it makes sense to use the DF_μ version, as DF_τ would admit linear trends in Δc and hence implausible quadratic trends in c.

3. The cointegrating regression

$$c_t = -0.306 + 1.036 yd_t + \hat{u}_t$$

conforms to the theoretical concept. The elasticity of 1.036 is only slightly larger than unity. The negative constant implies that c is smaller than yd in equilibrium. The DW statistic of 0.63 is sufficiently different from zero, while it of course indicates substantial autocorrelation. The t–statistic of 49.8 for the coefficient 1.036 shows the typical value for cointegrating regression: an extremely large value that reflects the non-standard asymptotic behavior.

4. A DF_0 test on the residuals $\hat{u}$ yields the value -2.93, which should be compared to the tables by Phillips&Ouliaris. Econometric software correctly calculates the test statistic but indicates incorrect significance points. For example, EViews has the 5% point of -1.96, while the table has the correct 5% point of -2.76. Thus, the cointegrating relation is

really significant at the 5% level (non-cointegration is rejected) but the result is closer than it would appear in a naive evaluation.

5. The vector $(1, -1.036)$ can be used as 'the' cointegrating vector in further modelling. It is interesting to note, however, that in the equation corresponding to (11)

$$\Delta c_t = \mu + \alpha\,(c_{t-1} - 1.036 yd_{t-1}) + \gamma \Delta yd_{t-1} + \varepsilon_t$$

the coefficient α attains only a moderate t–value around -1.7, and this does not even improve if further lagged differences of Δc and Δyd are inserted as additional regressors. Here, the traditional asymptotic approximation of the normal distribution for the t–statistic of α is valid. These contradictions and pieces of unclear evidence are better resolved in system estimation, as it is presented in the next section. Because of the small sample size, we cannot expect to attain perfect clarification, however.

2.2 Johansen's system estimation

The literature has offered various suggestions how to improve on the EG-2 procedure. Most of them can be classified into two groups. The first one focuses on improving the estimator of the cointegrating regression. These methods are still based on some kind of cointegrating regression. In principle, a cointegrating relation is not really a regression relation that models one variable to depend on others but an equilibrium relation among several variables. For this reason, the second group of procedures avoids the framework of regression altogether and uses multivariate generalizations of correlation analysis. The most popular procedure in this group was developed by Johansen, and this is also the main technique adopted by the programs RATS and EViews.

Johansen's method views cointegration tests as multivariate generalizations of ADF tests. The ADF test estimates a regression

$$\Delta X_t = \mu + \varphi X_{t-1} + \psi_1 \Delta X_{t-1} + \ldots + \psi_{p-1} \Delta X_{t-p+1} + \varepsilon_t.$$

The t–value of φ serves as the criterion for the null hypothesis $\varphi = 0$ and thus for *unit roots*. If X_t is a vector of n variables, one may estimate the system

$$\Delta X_t = \mu + \Pi X_{t-1} + \Psi_1 \Delta X_{t-1} + \ldots + \Psi_{p-1} \Delta X_{t-p+1} + \varepsilon_t. \qquad (13)$$

This is done efficiently by OLS for each of the n equations (according to Kruskal's Theorem). Here, μ is an n–vector, Π, Ψ_j, $j = 1, \ldots, p-1$ are

$n \times n$–matrices. ε_t is an n–dimensional *white noise*, with $\mathrm{E}\varepsilon_t\varepsilon_t' = \Sigma$: the error terms of individual equations are usually correlated, whereas $\mathrm{E}\varepsilon_t\varepsilon_s' = \mathbf{0}$ must hold for $s \neq t$. No component of ε_t is correlated with any component of ε_s, unless $s = t$.

The proof that every multivariate or vector autoregression (*vector autoregression, VAR*)

$$X_t = \mu + \Phi_1 X_{t-1} + \ldots + \Phi_p X_{t-p} + \varepsilon_t$$

can be re-written in the form (13) follows the univariate case. The relation between parameters $\mu; \Phi_1, \ldots, \Phi_p$ and $\mu; \Pi; \Psi_1, \ldots, \Psi_{p-1}$ is one-one (bijective). Here, we cannot review all of the many interesting details of VAR modeling, such as the multivariate ACF. By the way, multivariate ARMA models are avoided by most researchers for good reasons. They suffer from theoretical identification problems and numerical instability of estimation algorithms.

While the univariate φ can be zero or non-zero only, matrices Π have intermediate forms with rank less than n. It is just these that are in focus for cointegration modeling. One can show that the rank of Π, a number $\mathrm{rk}\Pi$ with $0 \leq \mathrm{rk}\Pi \leq n$, corresponds to the *number* of (linearly independent) cointegrating vectors *in the system*. We note the important special cases:

1. $\mathrm{rk}\Pi = 0$: there is no cointegration. All variables are $I(1)$. We can apply first differences and obtain a simplified VAR system for ΔX.

2. $\mathrm{rk}\Pi = 1$: the cointegrating vector is unique, excepting multiples.

3. $\mathrm{rk}\Pi = n$: all vectors cointegrate and hence all component variables themselves in the sense of *self cointegration*. The vector X contains stationary variables only.

These properties are valid only if some possibilities are excluded, such as seasonal unit roots, higher-order integration, $I(-1)$, explosive behavior. These assumptions are not tested within the procedure.

2.2.1 The determination of the rank

The rank of the matrix Π is determined in sequences of so-called *trace* tests. The trace statistics result from the so-called *canonical correlations* of X_{t-1} and ΔX_t. For all details see the book by Johansen (1995). The distribution of the *trace* statistics J_r under their null hypotheses H_{0r} in the testing

problem

$$H_{0r} : \text{rk}\Pi \leq r$$
$$H_{Ar} : \text{rk}\Pi > r$$

depends on $n - r$ only and is available in the form of quantile tables, for example from JOHANSEN's monograph. Because the JOHANSEN procedure is a multivariate generalization of the DF test, the distributions of all statistics under their null hypotheses are again functions of integrals over functions of Brownian motions.

The researcher typically starts by investigating J_0. If J_0 is insignificant, there is no cointegration and the test sequence stops. If the test rejects H_{00}, consider J_1. If J_1 is insignificant, there exists a unique cointegrating vector. If the test rejects H_{01}, consider J_2, etc.

In analogy to the ADF test, also the JOHANSEN test has several versions that are defined by specifications of deterministic portions, such as constants or trends. These versions have each different null distributions and need specific tables. Fortunately, only two variants are interesting in practice:

1. a vector of constants μ is estimated as in (13), without any further restriction.

2. the vector of constants is subjected to a *no-trend* restriction.

In case #1 will the *system* have linear trends, i.e. at least one component of X_t shows a systematic trend. In case #2, no component variable has systematic trending behavior. While it is possible to discriminate the two variants via hypothesis tests, it makes more sense to determine the choice on grounds of *a priori* information on the behavior of the variables. For example, macroeconomic entities such as output and consumption are obviously trending, while interest or inflation rates should not trend over longer time spans.

2.2.2 The estimation of the coefficients of the system

Understanding the cases #1 and #2 requires some more details on the *estimation* of the system parameters. From linear algebra it is known that every $n \times n$–matrix $\mathbf{M}$ of rank r can be represented in the form $\mathbf{M} = \mathbf{AB}'$, with matrices $\mathbf{A}$ and $\mathbf{B}$ each of dimension $n \times r$ and of *full rank* r. Application of this lemma to the matrix Π of the representation (13) yields the decomposition into factors

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

with $n \times r$–matrices α and β. The decomposition is not unique, equivalent decompositions evolve from linear combinations of columns of α or β.

Suppose now that $\mathrm{rk}\Pi = r$ is known, then (13) can be re-written as

$$\Delta X_t = \mu + \alpha\beta' X_{t-1} + \Psi_1 \Delta X_{t-1} + \ldots + \Psi_{p-1}\Delta X_{t-p+1} + \varepsilon_t. \qquad (14)$$

One can show that the columns of β, i.e. the rows of β', are the *cointegrating vectors*. Even these are, thus, not unique. For this reason, β is called the *cointegrating matrix*. The matrix α describes how the individual variables react to deviations from the equilibrium conditions expressed in β. Often, the elements of the matrix α are called *loading coefficients* or *adjustment coefficients*, or also *speed of adjustment*.

Estimates for β automatically result from the testing procedure. The solution of the canonical correlation problems requires solving an eigenvalue problem, and thus eigenvalues are also required in the calculation of the statistics J_r. The eigenvectors that belong to the non-zero eigenvalues are estimates for the columns of β. After replacing β in (14) by its estimate $\hat{\beta}$, the system can be regarded as a *multivariate regression system*, in which the components of ΔX_t are regressed on a constant and on $\hat{\beta}' X_{t-1}$, $\Delta X_{t-1}, \ldots, \Delta X_{t-p+1}$. These OLS regressions then yields estimates for μ, α, $\Psi_1, \ldots, \Psi_{p-1}$.

If the n–vector μ is proportional to the matrix α, i.e. if there exists a representation $\mu = \alpha\tilde{\mu}$, the system can be written as

$$\Delta X_t = \alpha(\tilde{\mu} + \beta' X_{t-1}) + \Psi_1 \Delta X_{t-1} + \ldots + \Psi_{p-1}\Delta X_{t-p+1} + \varepsilon_t,$$

with r–vector $\tilde{\mu}$. This is just the case #2 of the *no-trend* restriction. This case is based on a modified canonical correlation problem, in which an eigenvector is determined that operates on $(1, X')'$ and not, as in case #1, on X. This implies that one formally investigates cointegration between X and a constant. Of course, this would not make sense literally, as a constant cannot change the property of stationarity and thus of cointegration. Unfortunately, commercial software handbooks often use expressions like "*constant in the cointegrating relation*", a very questionable feature indeed.

Both cases #1 and #2 are multivariate counterparts to univariate ADF_μ. Also the case without a constant can be considered, the counterpart to ADF_0, which is irrelevant in practice. Only in exceptional situations will it make sense to consider counterparts to ADF_τ that use linear trends. Note that these typically imply implausible quadratic trends in the variables. The argument of allowing for trend-stationary processes as an alternative model is not as coercive for the multivariate JOHANSEN procedure as for the univariate DF test, as these are not in focus here. Usually, this possibility is ruled out in pre-tests using ADF statistics on individual variables.

2.2.3 An example

We consider the system with the two variables c and yd. An appropriate program can estimate VAR models with varying lag orders. For $p = 2$, AIC becomes minimal. The observation that both variables are trending rules out case #2. Thus, the model

$$\Delta X_t = \mu + \alpha\beta' X_{t-1} + \Psi_1 \Delta X_{t-1} + \varepsilon_t$$

is estimated. The JOHANSEN test is based on *one* lag in differences, not on two, as it would be recommended automatically by the EViews software. This specification yields the two statistics $J_0 = 8.127$ and $J_1 = 0.027$. A comparison with tables shows that J_0 is *not significant*, the two variables are *not cointegrated*, as H_{00} is not rejected. A presumable reason is the small sample size.

If we tentatively reject H_{00} nevertheless, H_{01} is safely accepted, as J_1 is very small. Imposing (contrary to the test recommendation) $r = 1$ yields the estimate $\hat{\beta}' = (-50.04, 54.35)$, which is close to a multiple of $(1, -1)$. The implied elasticity of 1.086 is slightly larger than one, which would indicate a long-run increase in the propensity to consume. The absolute values of the coefficients in the vector $\hat{\alpha}$ are also quite close to each other.

Normalizing the first entry of the vector β to unity—EViews does this by the normalization condition "B(1,1)=1"—delivers $\hat{\beta} = (1, -1.086)$ and $\hat{\alpha}' = (-0.188, 0.153)$. While the numbers 0.188 and 0.153 are similar, it is to be noted that the first coefficient attains an 'almost significant' t–value of 1.57, while the t–value of 0.81 on the second coefficient is small. These observations confirm that the existence of cointegration is poorly supported on statistical grounds, but that the main reaction to a dynamic disequilibrium is borne by an adjustment of consumption and not by adjusting income. The constants in μ and the elements of Ψ_1 attain significance on the levels of 5% or 10% in most cases. Here, the usual approximation of the normal distribution for the t–statistic is valid.

Tentatively replacing the estimated $\hat{\beta}$ by the theoretical vector $\beta = (1, -1)$ entails a deterioration of significance for the elements of α. The restriction itself is also rejected on statistical grounds. For this test decision, programs utilize the likelihood-ratio statistic $2(\log L_1 - \log L_0)$, which is asymptotically χ^2–distributed with degrees of freedom corresponding to the number of restrictions. This rule yields a $\chi^2(1)$–distribution here. The marginal significance level (the p–value) for the test statistic of 4.52 is just above 3%.□

References

[1] Box, G.E.P., and Jenkins, G.C. (1976). *Time Series Analysis, Forecasting and Control*. Revised edition. Holden-Day.

[2] Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994). *Time Series Analysis, Forecasting and Control.* 3rd edition. Prentice-Hall.

[3] Dickey, D.A., and Fuller, W.A. (1979). 'Distribution of the Estimators for Autoregressive Time Series with a Unit Root'. *Journal of the American Statistical Association* **74**, 427–431.

[4] Engle, R.F., and Granger, C.W.J. (1987) 'Co-integration and Error Correction: Representation, Estimation and Testing'. *Econometrica* **56**, 1333–1354.

[5] Fuller, W.A. (1996). *Introduction to Statistical Time Series*. 2nd Edition, Wiley. (1976: first edition)

[6] Gourieroux, C., and Monfort, A. (1997). *Time series and dynamic models*. Cambridge University Press.

[7] Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.

[8] Johansen, S. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

[9] Kwiatkoswki, D., Phillips, P.C.B., Schmidt, P., and Shin, Y. (1992) 'Testing the null hypothesis of stationarity against the alternative of a unit root'. *Journal of Econometrics* **54**, 159–178.

[10] Phillips, P.C.B., and Perron, P. (1988) 'Testing for a Unit Root in Time Series Regression'. *Biometrika* **75**, 335–346.

[11] Phillips, P.C.B, and Ouliaris, S. (1990) 'Asymptotic Properties of Residual Based Tests for Cointegration'. *Econometrica* **58**, 165–193.